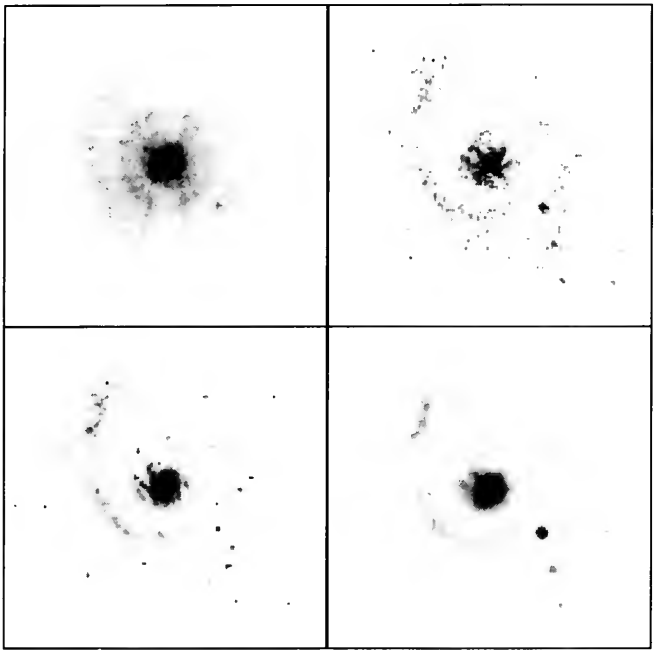
THE RESTORATION OF HST IMAGES AND SPECTRA–II



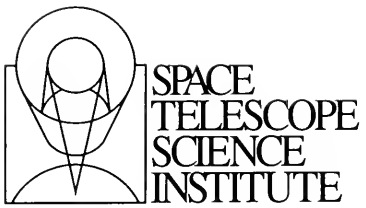
Proceedings of a Workshop held at
The Space Telescope Science Institute
Baltimore, Maryland USA
18-19 November 1993

Robert J. Hanisch and Richard L. White, editors

SPACE TELESCOPE SCIENCE INSTITUTE
Operated for NASA by AURA

NASA

Cover: A two-channel restoration of a simulated Planetary Camera image of a QSO with underlying spiral galaxy; the QSO point source is located at the nucleus of the galaxy as is 100 times as bright as the underlying peak brightness. Top left: test image, convolved with HST PSF and noise added. Bottom left: true, noise free image of QSO and galaxy. Top right: result standard Richardson-Lucy restoration. Bottom right: two-channel restoration. For details see Hook & Lucy, pp. 86–94.



The Restoration of HST Images and Spectra — II

Proceedings of a Workshop held at The Space Telescope Science Institute Baltimore, Maryland USA 18–19 November 1993

Edited by

Robert J. Hanisch and Richard L. White

Advanced Systems Group
Science Computing and Research Support Division
Space Telescope Science Institute
3700 San Martin Drive
Baltimore, MD 21218

The Space Telescope Science Institute is operated by the Association of Universities for Research in Astronomy, Inc., for the National Aeronautics and Space Administration.

Table of Contents

Preface

This volume contains the collected papers presented at the Second Workshop on the Restoration of Images and Spectra from the Hubble Space Telescope, which was held at the Space Telescope Science Institute, Baltimore, Maryland, on 18 and 19 November 1993. The Workshop was attended by 126 registered participants. Fifty-five papers were presented in a combination of oral and poster sessions.

The Workshop was sponsored by the Image Restoration Project of the Space Telescope Science Institute. Although work on image restoration began soon after the spherical aberration problem of the HST was discovered in the summer of 1990, a dedicated effort in the area of image restoration did not get underway until the fall of 1992. Since that time, however, a small group led by Bob Hanisch and Rick White has been doing research and implementing software aimed at improving the quality of HST images. The members of this group included Ivo Busko, Jinger Mo, Nailong Wu, and Nancy Hamilton. Collaborators in the project included Don Lindler, John Krist, and Dave Baxter. Our work was also supported (through several small research contracts) by Aggelos Katsaggelos, Bobby Hunt, David Redding, and their co-workers. The Project was run out of the Advanced Systems Group of the Science Computing and Research Support Division of ST ScI.

The first workshop on HST image restoration (August 1990) presented a large variety of algorithms and approaches, but relatively little had been done to adapt these approaches to the peculiarities of HST images and spectra. The attitudes were both optimistic and, perhaps, a bit naive. HST image restoration turned out to be extremely complicated, with both spatially and temporally variable point spread functions and a variety of other instrumental effects that made high quality restorations difficult. Over the past eighteen months we have learned how to deal with many of these problems, and how to modify algorithms and generate better quality PSFs. Although HST image restoration is far from the point where an arbitrary image can be tossed into a "black box" deconvolution machine, good quality restoration can indeed be obtained in a large number of cases. The papers in the current volume demonstrate this quite convincingly. Perhaps the most stringent requirement on image restoration has been to retain photometric linearity. Through studies of the convergence properties of algorithms such as the Richardson-Lucy method and various Maximum Entropy approaches, we now have a good understanding of how to proceed with an image restoration so that photometric integrity is maintained. Perhaps the most profound effect of the efforts on image restoration has been to open up a new method of dealing with optical imaging data that had previously been utilized only in a segment of the astronomical community (e.g., radio astronomy).

The wonderfully successful HST Servicing Mission in December 1993 has led to HST recovering virtually all of its original design goals for optical performance. The first science images to be returned to the ground are simply awesome. The improved optical performance of the telescope allows HST to recapture the sensitivity that was irretrievably lost in the aberrated images and to tackle the observational problems that require high dynamic range. It is our expectation, however, that the algorithms and PSF modeling expertise we have developed to deal with pre-servicing mission images will continue to be useful on WFPC-II and FOC+COSTAR data. It should be possible to increase dynamic range in crowded fields, where the diffraction spikes of stars overlap to create an apparent background haze, and allow studies of the galaxies underlying quasars that are otherwise obstructed by features of the unaberrated PSF. In some cases super-resolution may be possible, extending HST's capacity for seeing detail to beyond the diffraction limit.

With any printing process it is possible to lose some of the contrast and definition of gray-scale figures. As there are many half-tones in this volume, and since subtle features may be lost in the reproduction process, we wish to point out that the papers will be available via the Internet for the indefinite future. The Proceedings is available via ST ScI's home page on the World-Wide Web (URL `http://stsci.edu/top.html`) under the *Documentation* item. You may retrieve any paper and its associated figures in PostScript. Not all authors were able to provide us with original PostScript format figures, but we scanned the figures we were given to produce a totally digital version of the Proceedings.

We would like to extend thanks to those who helped to make the Image Restoration Project and this second Workshop possible. The Project was supported by the Hubble Space Telescope Project Office at the NASA Goddard Space Flight Center. We would especially like to thank Dr. Stanley Sobieski, ST ScI's Technical Officer at the time, for his support and encouragement. Other NASA officials who were especially supportive of this work include Dr. Robert Stachnik of NASA Headquarters and Dr. Jan Hollis of NASA GSFC. Within ST ScI we thank Dr. Ron Allen for his support and guidance, and thank John Krist, the developer of the Tiny TIM PSF modeling software, which has made it possible to restore many images for which observed PSFs were not otherwise available. John's work is based on the TIM PSF modeling code developed by Chris Burrows and Hashima Hasan, who also contributed to this project through their monitoring of the time variability of the PSF. Pierre Bely also provided useful inputs in this regard, especially in the area of spacecraft pointing effects. Finally, the Workshop itself would not have been possible had it not been for the dedicated efforts of Angie Clarke, with assistance from Britt Sjoeberg of the ST-ECF.

Robert J. Hanisch
Richard L. White
Space Telescope Science Institute
March 1994

List of Attendees

Hans-Martin Adorf
ST-ECF
Karl-Schwarzschild-Str. 2
D-85748 Garching
Germany
hmadorf@eso.org

M. Marsha Allen
The Johns Hopkins University
Department of Physics and Astronomy
34th and Charles Streets
Baltimore, MD 21218
allen@pha.jhu.edu

Ron Allen
Space Telescope Science Institute
3700 San Martin Drive
Baltimore, MD 21218
rjallen@stsci.edu

Bill Baity
University of California at San Diego
Center for Astrophysics and Space Sciences
La Jolla, CA 92093-0111
baity@cass05.ucsd.edu

David Baxter
Space Telescope Science Institute
3700 San Martin Drive
Baltimore, MD 21218
baxter@avion.stsci.edu

Pierre Bely
Space Telescope Science Institute
3700 San Martin Drive
Baltimore, MD 21218
bely@avion.stsci.edu

Piero Benvenuti
ST-ECF
Karl-Schwarzschild-Str. 2
D-85748 Garching
Germany
pbenvenu@eso.org

John Biretta
Space Telescope Science Institute
3700 San Martin Drive
Baltimore, MD 21218
biretta@stsci.edu

Mark Birkinshaw
Center for Astrophysics
60 Garden Street
Cambridge, MA 02138-1596
mbl@cfa.harvard.edu

Bill Blass
University of Tennessee
Deparment of Physics and Astronomy
Knoxville, TN 37966
blass@martha.utk.edu

Gary Bower
Space Telescope Science Institute
3700 San Martin Drive
Baltimore, MD 21218
bower@avion.stsci.edu

Charles Bowers
NASA/Goddard Space Flight Center
Code 683.3
Greenbelt, MD 20771
bowers@stars.gsfc.nasa.gov

Daniel Briggs
National Radio Astronomy Observatories
P.O. Box 0
Socorro, NM 87801
dbriggs@aoc.nrao.edu

Bernard Bundschuh
University Siegen
Institut fuer Nachrichtenverarb.
D-57068 Siegen
Germany
astrorub@ruba.rz.ruhr-uni-bochum.de

Jesus Burgos-Martin
Space Telescope Science Institute
3700 San Martin Drive
Baltimore, MD 21218
burgos@stsci.edu

Chris Burrows
Space Telescope Science Institute
3700 San Martin Drive
Baltimore, MD 21218
burrows@avion.stsci.edu

Ivo Busko
Astrophysics Division
INPE
C.P. 515
12201-970 Sao Jose dos Campos, SP
Brazil
busko@das.inpe.br

Uri Carsenty
DLR, NE-PE
D-82243 Wessling
Germany
carsenty@28842.dnet.nasa.gov

Stefano Casertano
The Johns Hopkins University
Department of Physics and Astronomy
34th and Charles Streets
Baltimore, MD 21218
stefano@mds.pha.jhu.edu

Adeline Caulet
ST-ECF
Karl-Schwarzschild-Str. 2
D-85748 Garching
Germany
acaulet@eso.org

Pete Challis
Center for Astrophysics
60 Garden Street
Cambridge, MA 02138-1596
pchallis@cfa.harvard.edu

Gordon Chin
NASA/Goddard Space Flight Center
Greenbelt, MD 20771
u2gnc@lepvax.gsfc.nasa.gov

Julian Christou
Starfire Optical Range
PL/LIG
Kirtland Air Force Base
3550 Aberdeen Ave. SE
Albuquerque, NM 87117
christoj@essvax.plk.af.mil

Angie Clarke
Space Telescope Science Institute
3700 San Martin Drive
Baltimore, MD 21218
aclarke@stsci.edu

Mike Cobb
Naval Research Laboratory
4555 Overlook Avenue, SW
Washington, DC 20375-5000
cobb@exploit.nrl.navy.mil

James Coggins
Computer Science Department
University of North Carolina
Chapel Hill, NC 27599-3175
coggins@cs.unc.edu

Tim Cornwell
National Radio Astronomy Observatories
P.O. Box 0
Socorro, NM 87801
tcornwel@nrao.edu

Dennis Crabtree
Dominion Astrophysical Observatory
Canadian Astronomy Data Center
5071 W. Saanich Rd.
Victoria, B.C. V8X 4M6
Canada
crabtree@dao.nrc.ca

Cindy Cunningham
Space Astrophysics Laboratory, ISTS
4850 Keele St.
North York, Ont. M3J 3K1
Canada
ccc@nereid.sal.ists.ca

Eric Deutsch
University of Washington
Department of Astronomy
Seattle, WA 98195
deutsch@astro.washington.edu

David Devine
University of Colorado
Boulder, CO 80309
devine@janos.colorado.edu

Athanassios Diplas
University of California at San Diego
Center for Astrophysics and Space Sciences
La Jolla, CA 92093-0111
diplas@cassfos02.ucsd.edu

Ken Ebisawa
ASCA GOF
NASA/Goddard Space Flight Center
Greenbelt, MD 20771
ebisawa@bluejay.gsfc.nasa.gov

Mohammad Faisal
Washington University
One Brookings Drive
St. Louis, MO 63130
faisal@saturn.wustl.edu

Luciana Federici
Osservatorio Astronomico
Via Zamboni 33
I-40126 Bologna
Italy
luciana@alma02.bo.astro.it

Laura Ferrarese
Space Telescope Science Institute
3700 San Martin Drive
Baltimore, MD 21218
ferrarese@stsci.edu

Duncan Forbes
Lick Observatory
Santa Cruz, CA 95064
forbes@lick.ucsc.edu

Robert Fosbury
ST-ECF
Karl-Schwarzschild-Str. 2
D-85748 Garching
Germany
rfosbury@eso.org

Didier Fraix-Burnet
Observatoire Midi-Pyrenees
14, Avenue E. Belin
F-31400 Toulouse
France
fraix@srvdec.obs-mip.fr

Wolfram Freudling
ST-ECF
Karl-Schwarzschild-Str. 2
D-85748 Garching
Germany
wfreudli@eso.org

Laura Fullton
University of North Carolina
Department of Physics and Astronomy
CB #3255, Phillips Hall
Chapel Hill, NC 27599-3255
laura@gluttony.astro.unc.edu

Flavio Fusi Pecci
Observatory of Bologna
via Zamboni 33
I-40126 Bologna
Italy
flavio@alma02.cineca.it

Ron Gilliland
Space Telescope Science Institute
3700 San Martin Drive
Baltimore, MD 21218
gillil@stsci.edu

Bob Gonsalves
Tufts University
Electro-Optics Technology Center
4 Colby Street
Medford, MA 02155
szaccheo@osiris.ee.tufts.edu

Eric Gotthelf
NASA/Goddard Space Flight Center
Greenbelt, MD 20771
gotthelf@lheavx.gsfc.nasa.gov

Perry Greenfield
Space Telescope Science Institute
3700 San Martin Drive
Baltimore, MD 21218
greenfield@stsci.edu

Richard Griffiths
The Johns Hopkins University
Department of Physics and Astronomy
34th and Charles Streets
Baltimore, MD 21218
griffith@mds.pha.jhu.edu

Nancy Hamilton
Space Telescope Science Institute / NSA
3700 San Martin Drive
Baltimore, MD 21218
nancyh@stsci.edu

Robert Hanisch
Space Telescope Science Institute
3700 San Martin Drive
Baltimore, MD 21218
hanisch@stsci.edu

Eduardo Hardy
Laval University
Department of Physics
Quebec, PQ G1K 7P4
Canada
hardy@phy.ulaval.ca

Richard Harms
Applied Research Corporation
8201 Corporate Drive
Landover, MD 20785
harms@fosvax.arclch.com

Hashima Hasan
Space Telescope Science Institute
3700 San Martin Drive
Baltimore, MD 21218
hasan@avion.stsci.edu

D. John Hillier
University of Pittsburgh
3941 O'Hara Street
Pittsburgh, PA 15260
hillier@vms.cis.pitt.edu

Phil Hodge
Space Telescope Science Institute
3700 San Martin Drive
Baltimore, MD 21218
hodge@stsci.edu

Richard Hook
ST-ECF
Karl-Schwarzschild-Str. 2
D-85748 Garching
Germany
rhook@eso.org

Scott Horner
Pennsylvania State University
Department of Astronomy and Astrophysics
525 Davey Laboratory
University Park, PA 16802
horner@astro.psu.edu

Bobby Hunt
Electrical and Computer Engineering Department
University of Arizona
Building 104
Tucson, AZ 85721
hunt@hermes.ece.arizona.edu

Giacinto Iannicola
Osservatorio Astronomico di Roma
Viale del Parco Mellini 84
I-00136 Rome
Italy
iannicola@astrmp.astro.it

Peter Jansson
E.I. duPont de Nemours and Co. Inc.
P.O. Box 80357
Wilmington, DE 19880-0357
janssopa@esvax.dnet.dupont.com

Stuart Jefferies
Bartol Research Institute
University of Delaware
c/o National Solar Observatory
950 N. Cherry Ave.
Tucson, AZ 85719
stuartj@noao.edu

Margarita Karovska
Center for Astrophysics
60 Garden Street
Cambridge, MA 02138-1596
karovska@cfa.harvard.edu

Aggelos Katsaggelos
Northwestern University
2131 Sheridan Rd.
Evanston, IL 60208
aggk@delta.eecs.nwu.edu

Bill Keel
University of Alabama
Department of Physics and Astronomy
Box 870324
Tuscaloosa, AL 35487-0324
keel@okra.astr.ua.edu

Ivan King
University of California
Astronomy Department
Berkeley, CA 94720
king@bkyast.berkeley.edu

Ajay Kochhar
Applied Research Corporation
8201 Corporate Drive
Landover, MD 20785
kochhar@arcfos1.arclch.com

John Krist
Space Telescope Science Institute
3700 San Martin Drive
Baltimore, MD 21218
krist@stsci.edu

Lukas Labhardt
University of Basel
Astronomical Institute
Venusstrasse 7
CH-4102 Binningen
Switzerland
labhardt@urz.unibas.ch

Peter Linde
Lund Observatory
Box 43
S-22100 Lund
Sweden
peter@astro.lu.se

Don Lindler
Advanced Computer Concepts, Inc.
11518 Gainsborough Road
Potomac, MD 20854
hrslindler@hrs.gsfc.nasa.gov

Jorge Llacer
Univerity of California
Lawrence Berkeley Laboratory
Building 46A
Berkeley, CA 94530
llacer@aida.lbl.gov

Leon Lucy
ST-ECF
Karl-Schwarzschild-Str. 2
D-85748 Garching
Germany
llucy@mc1.hq.eso.org

Fabio Maggio
CRS4
Via N. Sauro, 10
I-09123 Cagliari
Italy
maggio@crs4.it

Stephen Mahan
University of Tennessee
401 Nielsen Physics Building
Knoxville, TN 37996-1200
mahan@utkux3.utk.edu

Hugo Marraco
CONAE
Av. Dorrego 4010
1425 Buenos Aires
Argentina
hmarraco@orbis.conae.gov.ar

Javier Mateos
Universidad de Granada
Department of Computer Science and A.I.
E-18071 Granada
Spain
jmd@robinson.ugr.es

Gerhardt Meurer
Space Telescope Science Institute
3700 San Martin Drive
Baltimore, MD 21218
meurer@stsci.edu

Kenneth Mighell
Columbia University
Department of Astronomy
538 West 120th Street
New York, NY 10027
mighell@figaro.phys.columbia.edu

Jinger Mo
Space Telescope Science Institute
3700 San Martin Drive
Baltimore, MD 21218
jinger@stsci.edu

Rafael Molina
Universidad de Granada
Facultad de Ciencias
Department of Computer Science and A.I.
E-18071 Granada
Spain
rms@robinson.ugr.es

Stephen Morris
Dominion Astrophysical Observatory
Canadian Astronomy Data Center
5071 W. Saanich Rd.
Victoria, B.C. V8X 4M6
Canada
morris@dao.nrc.ca

Fionn Murtagh
ST-ECF
Karl-Schwarzschild-Str. 2
D-85748 Garching
Germany
fmurtagh@eso.org

Steve Mutz
Arizona State University
Department of Physics and Astronomy
Tempe, AZ 85287
sbm@omega.la.asu.edu

Lyman Neuschaefer
The Johns Hopkins University
Department of Physics and Astronomy
34th and Charles Streets
Baltimore, MD 21218
lwn@mds.pha.jhu.edu

Pete Nisenson
Center for Astrophysics
60 Garden Street
Cambridge, MA 02138-1596
nisenson@cfa.harvard.edu

Jorge Nunez
Universidad de Barcelona
Department d'Astronomia i Meteorologia
Av. Diagonal 647
E-08028 Barcelona
Spain
jorge@fajnm1.ub.es

Bill Oliver
Armed Forces Institute of Pathology
14th NW and Alaska Ave.
Washington, DC 20306-6000
oliver@ipas.afip.mil

Earl O'Neil, Jr.
Kitt Peak National Observatory
P.O. Box 26732
Tucson, AZ 85726-6732
oneil@vela.tuc.noao.edu

Eric Pantin
Services d'Astrophysique CE Saclay
CEA
F-91191 Gif-sur-Yvette
France
pantin@sapvxg.saclay.cea.fr

Richard Paxman
Environmental Research Institute of Michigan
P.O. Box 134001
Ann Arbor, MI 48114-4001
paxman@erim.org

Steve Penton
University of Colorado
Center for Astrophysics and Space Astronomy
Campus Box 389
Boulder, CO 80389
spenton@zwicky.colorado.edu

Kevin Perry
Department of Electrical Engineering
Broun Hall
Auburn University, AL 36849-5201
kperry@eng.auburn.edu

Andrew Phillips
Lick Observatory
University of California at Santa Cruz
Santa Cruz, CA 95064
phillips@lick.ucsc.edu

E. R. Pike
King's College
The Strand
London WC2R 2LS
United Kingdom
erp@ipg.ph.kcl.ac.uk

Chris Pritchet
University of Victoria
Department of Physics and Astronomy
P.O. Box 3055
Victoria, BC
Canada
pritchet@clam.phys.uvic.ca

Rick Puetter
University of California at San Diego
Center for Astrophysics and Space Sciences
La Jolla, CA 92093-0111
rick@cassir.ucsd.edu

Jason Pun
Harvard-Smithsonial Center for Astrophysics
60 Garden Street
Cambridge, MA 02138
pun@cfa.harvard.edu

P. Raadhakrishnan
NASA/Goddard Space Flight Center
Greenbelt, MD 20771
rads@hartley.gsfc.nasa.gov

Kavan Ratnatunga
The Johns Hopkins University
Department of Physics and Astronomy
34th and Charles Streets
Baltimore, MD 21218
kavan@mds.pha.jhu.edu

David Redding
Jet Propulsion Laboratory
MS 306-463
4800 Oak Grove Dr.
Pasadena, CA 91109
dave@huey.jpl.nasa.gov

Stan Reeves
Department of Electrical Engineering
Broun Hall
Auburn University, AL 36849-5201
sjreeves@eng.auburn.edu

Susan Ridgway
University of Hawaii
Institute for Astronomy
2680 Woodlawn Drive
Honolulu, HI 96822
ridgway@hubble.ifa.hawaii.edu

David Robinson
University of Cambridge
Insitute of Astronomy
Madingley Road
Cambridge CB3 0HA
United Kingdon
drtr@mail.ast.cam.ac.uk

Mark Robison
University of Minnesota
13305 34th Avenue, North
Plymouth, MN 58441
mrobison@shemesh.spa.umn.edu

Sylvie Roques
Observatoire Midi-Pyrenees
14 Avenue E. Belin
F-31400 Toulouse
France
roques@srvdec.obs-mip.fr

Paul Schmidtke
Arizona State University
Department of Physics and Astronomy
Tempe, AZ 85287-1504
schmidtke@scorpius.la.asu.edu

Timothy Schulz
Michigan Technological University
Department of Electrical Engineering
1400 Townsend Drive
Houghton, MI 49931-1295
schulz@mtu.edu

Ethan Schreier
Space Telescope Science Institute
3700 San Martin Drive
Baltimore, MD 21218
schreier@stsci.edu

Sue Simkin
Michigan State University
Department of Physics and Astronomy
East Lansing, MI 48824-4530
simkin@grus.pa.msu.edu

Britt Sjoeberg
ST-ECF
Karl-Schwarzschild-Str. 2
D-85748 Garching
Germany
bsjoeber@eso.org

Ralph Snel
Lund Observatory
Box 43
S-22100 Lund
Sweden
ralph@astro.lu.se

Don Snyder
Washington University
Department of Electrical Engineering
Box 1127
St. Louis, MO 63130
dls@saturn.wustl.edu

Jean-Luc Starck
Observatoire de la Cote d'Azur
B.P. 229
F-06304 Nice Cedex 4
France
starck@corelli.obs-nice.fr

Peter Stetson
Dominion Astrophysical Observatory
507 West Saanich Road
Victoria, BC V8X 4M6
Canada
stetson@dao.nrc.ca

Alan Stockton
University of Hawaii
Institute for Astronomy
2680 Woodlawn Drive
Honolulu, HI 96822
stockton@galileo.ifa.hawaii.edu

Alex Storrs
Space Telescope Science Institute
3700 San Martin Drive
Baltimore, MD 21218
storrs@stsci.edu

Larry Trafton
University of Texas
DL 7262895
1208-A Elm Street
Austin, TX 78703
lmt@astro.as.utexas.edu

Neil Trentham
University of Hawaii
Institute for Astronomy
Woodlawn Drive
Honolulu, HI 96822
nat@galileo.ifa.hawaii.edu

David Turnshek
University of Pittsburgh
3941 O'Hara Street
Pittsburgh, PA 15260
turnshek@vms.cis.pitt.edu

Tony Tyson
AT&T Bell Laboratories
ID335
600 Mountain Ave.
Murray Hill, NJ 07974
tyson@physics.att.com

Olaf Vancura
Smithsonian Astrophysical Observatory
60 Garden Street
Cambridge, MA 02138
vancura@cfa.harvard.edu

Emanuel Vassiliadis
Space Telescope Science Institute
3700 San Martin Drive
Baltimore, MD 21218
vass@stsci.edu

Jeremy Walsh
ST-ECF
Karl-Schwarzschild-Str. 2
D-85748 Garching
Germany
jwalsh@eso.org

Richard L. White
Space Telescope Science Institute
3700 San Martin Drive
Baltimore, MD 21218
rlw@stsci.edu

Eric Wyckoff
The Johns Hopkins University
Department of Physics and Astronomy
34th and Charles Streets
Baltimore, MD 21218
wyckoff@mds.pha.jhu.edu

Nailong Wu
Space Telescope Science Institute
3700 San Martin Drive
Baltimore, MD 21218
nailong@stsci.edu

T. Scott Zaccheo
Tufts University
Department of Electrical Engineering
College Ave.
Medford, MA 02155
szaccheo@osiris.ee.tufts.edu

Algorithms

The Restoration of HST Images and Spectra II
Space Telescope Science Institute, 1994
R. J. Hanisch and R. L. White, eds.

Adaptive Regularized Restoration Algorithms Applied to HST Images

Aggelos K. Katsaggelos, Moon Gi Kang, and Mark R. Banham

Dept. of Electrical Engineering and Computer Science, Northwestern University, Evanston, IL 60201-3118

Abstract. This paper analyzes the performance of two set theoretic-based iterative image restoration algorithms for Hubble Space Telescope (HST) degraded images. The iterative adaptive constrained least squares and frequency adaptive constrained least squares algorithms are optimized here for HST data, and applied to several simulated and real degraded HST images. Evaluations of both the flux linearity and resolution enhancement of these algorithms are presented and compared to results obtained by the Richardson-Lucy algorithm (Lucy 1974). These results indicate that the iterative algorithms investigated here are quite suitable for HST data, and provide excellent results in terms of all evaluation criteria tested.

1. Introduction

There have been a variety of techniques applied to the restoration of HST images over the last several years (White 1991, Weir 1991). Most of these approaches have concentrated on algorithms already developed and well accepted by those in the astronomical community. At the same time, however, the signal and image processing community was developing a number of digital image restoration techniques which were motivated largely by non-astronomical problems (Andrews and Hunt 1977). For a recent review and classification of image restoration algorithms, see Katsaggelos (1991). For the most part, the algorithms introduced by image processing scientists tend to be very general and applicable to a variety of imaging problems. In examining the HST degraded images, it is apparent that many of the image restoration techniques, and in particular the iterative techniques, developed in the realm of signal and image processing would be well suited for application to this special problem.

In this paper two iterative image restoration techniques are applied to the HST data restoration problem. These are the generalized iterative adaptive algorithm (Kang and Katsaggelos 1993) and the frequency domain iterative adaptive algorithm (Kang and Katsaggelos 1992). Both are based on a set-theoretic approach to image restoration. The critical factor in choosing these algorithms is that, unlike many regularized restoration techniques, these algorithms need no prior information about the signal and no knowledge of the noise variances present. These are also relatively fast algorithms, and have been developed with convergence speed as a consideration. The generalized iterative adaptive algorithm allows for spatial adaptivity and is applicable to the case of spatially variant degradation.

The performance of these algorithms is very good with the HST data that we have tested. The criteria which are most important to the astronomical community have been utilized in compiling the results here. It can be seen from these results that there is much potential for use of these iterative constrained least squares algorithms from the viewpoint of not only quality of results, but also convergence speed. In the next section, we describe the fundamentals of set theoretic image restoration, and describe the iterative techniques that we have developed through this approach.

2. Iterative Image Restoration

For the algorithms discussed here, the restoration approaches have been formulated to handle the problem of an image degraded by a spatially invariant or spatially variant blur operator with additive Gaussian noise. This description represents the degradation present in the HST WFPC images. The image degradation process can be modeled according to

$$y = Dx + n \,, \tag{1}$$

where a lexicographical ordering of the observed digital image, y, the original image, x, and the additive noise, n, is used. D represents the degradation operator of the imaging system. The image restoration problem calls for obtaining an estimate of x given y, D, and some characterization of the noise process, n.

Iterative approaches may be used to solve this inverse problem very effectively. The primary advantages of iterative techniques are (Schafer et al. 1981, Katsaggelos 1989): (i) there is no need to explicitly implement the inverse of an operator; (ii) knowledge about the solution may be directly incorporated into the restoration process; (iii) the process may be monitored as it progresses; (iv) the effect of noise may be controlled with certain constraints; and (v) parameters determining the solution can be updated as the iteration progresses.

The iterative techniques applied here to HST data are developed through a set theoretic approach. In this approach, prior constraints on the solution are imposed by (Katsaggelos et al. 1985, Katsaggelos 1989, Katsaggelos et al. 1991)

$$Q_x = \{x \mid \|Cx\|^2 \leq E^2\} \,, \tag{2}$$

and

$$Q_{x|y} = \{x \mid \|y - Dx\|^2 \leq \epsilon^2\} \,, \tag{3}$$

where the solution belongs to both ellipsoids described by Eqs. (2) and (3). In Eq. (2), the operator C represents a highpass filter which bounds the high frequency energy of the restored image. If the bounds ϵ^2 and E^2 are known, and the intersection of Q_x and $Q_{x|y}$ is not empty, the solution may be found by solving

$$(D^T D + \alpha C^T C)x = D^T y \,, \tag{4}$$

where α, the regularization parameter, is equal to $(\epsilon/E)^2$, which controls the trade-off between fidelity to the data, and smoothness of the solution.

2.1. Generalized Iteration Adaptive Algorithm

The first algorithm we tested for HST data was the generalized iterative adaptive algorithm which simultaneously restores the image and determines a single regularization parameter based on the restored image at each iteration. This algorithm does not depend on the initial conditions. In other words, the smoothing functional to be minimized can be shown to have a unique minimizer. An accurate estimate of the noise variance is also obtained at convergence. The approach we have proposed (Kang and Katsaggelos 1993) extracts the properties of the original image from the partially restored image at each iteration step.

In order to solve for an optimal restored image and regularization parameter at each iteration, we use

$$M(\alpha(x), x) = \|y - Dx\|^2_{A(x)} + \alpha(x)\|Cx\|^2_{B(x)} \,, \tag{5}$$

as the functional to be minimized. The weighting matrices $A(x)$ and $B(x)$ are included to make the restoration algorithm spatially adaptive. The necessary condition for a minimum is that the gradient of $M(\alpha(x), x)$ with respect to x be equal to zero. This results in

$$\begin{aligned} [D^T(A(x) + A(x)^T)D &+ \alpha_w(x)C^T(B(x) + B(x)^T)C]x + \|Cx\|^2_{B(x)}\nabla_x\alpha_w(x) \\ &= D^T(A(x) + A(x)^T)y \,. \end{aligned} \tag{6}$$

When the noise and the high pass filtered image are stationary, the weighting matrices $A(x)$ and $B(x)$ become symmetric, and thus the equation can be rewritten as

$$[D^T A(x)D + \alpha_w(x)C^T B(x)C]x + \|Cx\|^2_{B(x)}\nabla_x\alpha_w(x) = D^T A(x)y \ , \tag{7}$$

and it becomes

$$[D^T A(x)D + \alpha_w(x)C^T B(x)C]x = D^T A(x)y \ . \tag{8}$$

since $\nabla_x\alpha_w(x) = 0$ with a proper choice of $\alpha_w(x)$. When the noise and the high pass filtered image are uncorrelated with themselves, even though they are nonstationary, $A(x)$ and $B(x)$ become diagonal, and therefore we obtain the solution for Eq. (7). When the noise and the high pass filtered image are white, then the weighting matrices become constant identity matrices multiplied by the constant variances. The highly nonlinear term $\|Cx\|^2_{B(x)}\nabla_x\alpha_w(x)$ can be removed by the proper choice of the regularization functional, based on the global convexity of the smoothing functional. Since Eq. (8) is nonlinear, we can not solve for x in a direct way, but we can use an iterative technique.

The regularization parameter is defined here as a function of the original image (but in practice becomes a function of an estimate of the original image). The form of the smoothing functional to be minimized is of great importance since it preserves convexity and exhibits only a global minimizer. After investigating the desirable properties for the regularization functional to satisfy, the following two forms have been shown to provide optimal solutions (Kang and Katsaggelos 1993)

$$\alpha(x) = \frac{\|y - Dx\|^2_{A(x)}}{(1/\gamma_1) - \|Cx\|^2_{B(x)}} \ , \tag{9}$$

where $\gamma_1 = 1/(2\|y\|^2)$, and

$$\alpha(x) = \gamma_2\|Cx\|^2_{B(x)} - 1 + \sqrt{(1 - \gamma_2\|Cx\|^2_{B(x)})^2 + 2\gamma_2\|y - Dx\|^2_{A(x)}} \ , \tag{10}$$

where $\gamma_2 = 3/(4\|y\|^2)$ controls convergence and convexity.

Given an optimal choice for $\alpha(x)$, we can solve Eq. (8) by the method of successive approximations with (Schafer et al. 1981, Katsaggelos et al. 1985, Katsaggelos 1989, Katsaggelos et al. 1991):

$$x_{k+1} = x_k + \left[D^T A(x)y - (D^T A(x)D + \alpha(x_k)C^T B(x)C)x_k\right] \ . \tag{11}$$

Using either choice of regularization functional (Eq. (9) or Eq. (10)), the iterative algorithm does not depend on the initial condition despite the nonlinearity of the iteration. This is due to the convexity of the functional and the convergence criteria satisfied by the globally optimal iteration. The algorithm is applicable to any type of degradation D and stabilizing matrix C (both of which may be spatially varying). So, there is no requirement that D and C be block circulant matrices. It is also important to note that no knowledge of the noise variance, or of the bound which determines the ellipsoid that expresses the smoothness of the image, is assumed.

2.2. Frequency Domain Iterative Adaptive Algorithm

The second iterative algorithm examined here is a nonlinear frequency domain algorithm in which the regularization parameter is frequency dependent, and updated at each iteration step. In this case, the algorithm considers that D is a block circulant matrix representing a spatially invariant blur. Also, the set-theoretic formulation is constructed in a weighted space, such that

$$Q_x = \{x \mid \|Cx\|^2_R \leq E^2_R\} \ , \tag{12}$$

and

$$Q_{x|y} = \{x \mid \|y - Dx\|^2_P \leq \epsilon^2_P\} \ , \tag{13}$$

where P and R are both block circulant weighting matrices. These matrices are chosen to maximize the speed of convergence at every frequency component, and to compensate for the near-singular frequency components of the iteration. The solution which belongs to the intersection of the ellipsoids given by Eqs. (12) and (13) is given by

$$(D^T P^T P D + \lambda C^T R^T R C)x = D^T P^T P y \; . \tag{14}$$

We define $P^T P = B$, $R = PQ$ and $\lambda Q^T Q = A$. The block circulant matrix B is the spatial domain representation of the relaxation parameter which will be called $\beta(\underline{l})$ in the frequency domain, and A is the block circulant spatial domain representation of the regularization parameter which is shown next in the frequency domain as $\alpha_k(\underline{l})$. Again, successive approximation may be used to solve Eq. (14). Since all of the matrices in this equation are block-circulant, the iteration may be written in the discrete frequency domain as

$$X_0(\underline{l}) = \beta(\underline{l}) D^*(\underline{l}) Y(\underline{l}) \; , \tag{15}$$

$$X_{k+1}(\underline{l}) = X_k(\underline{l}) + \beta(\underline{l}) \left[D^*(\underline{l}) Y(\underline{l}) - (|D(\underline{l})|^2 + \alpha_k(\underline{l}) |C(\underline{l})|^2) X_k(\underline{l}) \right] \; , \tag{16}$$

where $\underline{l}$ represents a single 2-D frequency component. In this case,

$$\alpha_k(\underline{l}) = \frac{\sum_{\underline{m}} |Y(\underline{m}) - D(\underline{m}) X_k(\underline{m})|^2}{\sum_{\underline{n}} |C(\underline{n}) X_k(\underline{n})|^2 + \delta_k(\underline{l})} \; . \tag{17}$$

Here, the $\delta_k(\underline{l})$ term we use is defined by

$$\delta_k^{used}(\underline{l}) = \delta_k^{conv}(\underline{l}) + \gamma(\delta_k^{opt} - \delta_k^{conv}(\underline{l})) \; , \tag{18}$$

where $0 < \gamma < 1$, and

$$\delta_k^{opt} = N^2 \max_{\underline{l}} (|X_k(\underline{l})|^2 |C(\underline{l})|^2) - \sum_{\underline{m}} |C(\underline{m}) X_k(\underline{m})|^2 \; , \tag{19}$$

and

$$\delta_k^{conv}(\underline{l}) = \frac{-2|D(\underline{l})|^2 \sum_{\underline{m}} |C(\underline{m}) X_k(\underline{m})|^2 - \sum_{\underline{n}} |Y(\underline{n}) - D(\underline{n}) X_k(\underline{n})|^2 |C(\underline{l})|^2}{2|D(\underline{l})|^2}$$

$$+ \frac{|C(\underline{l})| \sqrt{|C(\underline{l})|^2 (\sum_{\underline{n}} |Y(\underline{n}) - D(\underline{n}) X_k(\underline{n})|^2)^2 + 8|D(\underline{l})|^2 |C(\underline{l}) X_k(\underline{l})|^2 \sum_{\underline{n}} |Y(\underline{n}) - D(\underline{n}) X_k(\underline{n})|^2}}{2|D(\underline{l})|^2} \tag{20}$$

The optimized parameter β is given at each frequency component by (Strand 1974)

$$\begin{aligned} \beta(\underline{l}) &= F(|D(\underline{l})|^2) \\ &= 31.5 - 315|D(\underline{l})|^2 + 1443.75|D(\underline{l})|^4 - 3465|D(\underline{l})|^6 + 4504.5|D(\underline{l})|^8 \\ &\quad -3003|D(\underline{l})|^{10} + 804.375|D(\underline{l})|^{12} \; . \end{aligned} \tag{21}$$

According to this iteration, since β and δ_k are frequency dependent the convergence of the iteration can be accelerated, making this an attractive algorithm where speed is a concern.

2.3. Optimization for HST Data

In the course of testing our algorithms, we have made appropriate consideration of the optimization of these algorithms for HST data. In particular, we have included a positivity constraint at each step of the iteration. This imposes the condition that there should be no negative flux in our image source. The algorithms were developed with additive Gaussian noise as the observation noise in the model. So, these algorithms are well equipped to deal with the read-out noise problem of the WFPC.

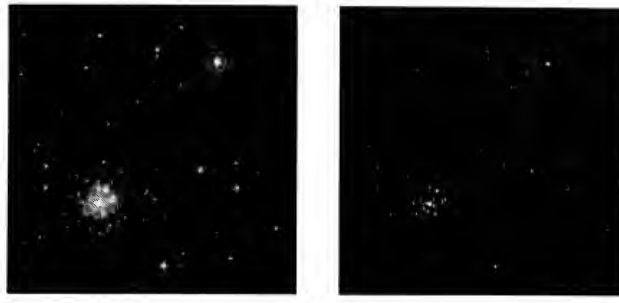


Figure 1. Blurred image "sim1.fit" and Richardson-Lucy restoration of "sim1.fit".

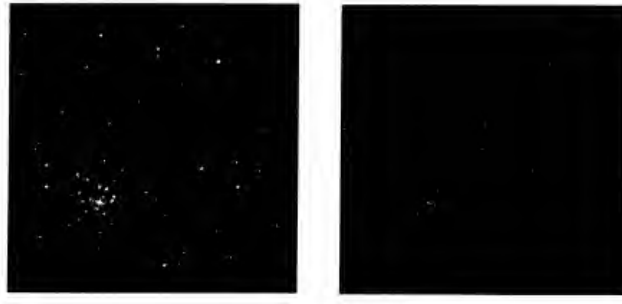


Figure 2. Generalized iterative adaptive restoration of "sim1.fit" and frequency adaptive restoration of "sim1.fit".

3. Results

All of the results presented here use the test data prepared by ST ScI and obtained from the directory `software/stsdas/restore` at the stsci.edu Internet site. Using the image "sim1.fit", which represents a star cluster with a globular cluster-like luminosity function, with the point spread function (PSF) "mpsf12.fit", we generated restored images using the two iterative algorithms presented in this paper. The image "sim1.fit" simulates a monochromatic observation from the Wide Field Camera with a space invariant blur. For comparison purposes, we utilized an implementation of the modified Richardson-Lucy algorithm developed at ST ScI. This algorithm has modifications which address the constant readout noise problem. Fig. 1 shows the original blurred image and the Richardson-Lucy restoration. Fig. 2 shows the restored result using the generalized iterative adaptive algorithm and the frequency adaptive algorithm. For all cases where we have used the generalized iterative adaptive algorithm in this paper, we have taken the weighting matrices in Eq. (11) to be equal to identity, although a spatially adaptive approach may easily be employed as well.

For all of the iterative algorithms, we set the maximum number of iterations to be 300. In addition, each algorithm used another termination criterion. For the Richardson-Lucy algorithm, the convergence criteria was that the value of χ^2 between the data and the blurred model was equal

to unity per degree of freedom (Lucy 1974). For the other two algorithms, a convergence criteria based on the $L2$ norm of the residual was used. In this case, we used a measure of the normalized error at each iteration, defined as $\|x_k - x_{k-1}\|^2 / \|x_{k-1}\|^2 < 10^{-6}$. The Richardson-Lucy algorithm reached 300 iterations before the parameter χ^2 was equal to 1, but the value of χ^2 was changing very slowly at this point, such that further iterations resulted in little change in the restored image. The generalized iterative adaptive method took 292 iterations while the frequency adaptive method required only 39 iterations. The form of the iterations are somewhat different for the Richardson-Lucy algorithm and the two algorithms presented here, so a direct comparison in terms of iteration counts is not completely straightforward. However, each iteration of the three algorithms requires approximately the same amount of time, so it can be seen that the frequency adaptive algorithm provides a very fast solution. The mean square error (MSE) was measured for each of these results after normalizing the restored images to the maximum value of the observed image in order to account for any linear scaling present in the different algorithms. The Richardson-Lucy algorithm had an MSE of 1083.69. The generalized iterative adaptive algorithm had an MSE of 1177.41, and the frequency adaptive algorithm had an MSE of 26.27. The MSEs were measured in terms of the 470 stars in the truth list.

For images having a spatially varying PSF, it is possible to apply the generalized iterative adaptive algorithm as well. We have applied the spatially varying implementation of this algorithm to the synthetic image "sim3.fit" which represents a star field with a spatially varying blur. For this image the PSF changed between 25 different PSFs at various positions in the image. The varying PSF is easily represented by a full degradation matrix in Eq. (1)., as opposed to a block circulant. For the restored image, the MSE was equal to 3316.3.

These simulated star field images provide the best data for testing the flux linearity of the restoration algorithms. We have evaluated the flux linearity of these three results according to the following method. The linearity was measured by taking the residual (measured as: original (truth) image - restored image) for each star in the truth image. The results are displayed by ordering the 470 stars in this image by increasing magnitude. So, the faintest stars are at the left, and the brightest stars are at the right. The flux linearity of the blurred image is shown in Fig. 3. The graph for the Richardson-Lucy implementation is seen in Fig. 4, the spatial iterative adaptive algorithm's graph is seen in Fig. 5, and the frequency adaptive algorithm's is seen in Fig. 6. The frequency adaptive algorithm produces the most linear curve for this test, with a very noticeable reduction in the number of outlying stars having a large error. The flux linearity results for spatially varying PSF case ("sim3.fit") are shown in Figs. 7 and 8.

For testing the resolution enhancement properties of our algorithm, we generated a synthetic image containing points of equal intensity separated by progressively increasing distances. These points represent simulated neighboring stars. This image was then blurred with the Gaussian point spread function having a variance of 9, and a support of 50x50 pixels. Using this PSF allows us to test the resolution enhancement for binary pairs of stars which are poorly resolved in the blurred image. The minimum separation between a pair of points in the synthetic image was one pixel, and we measured the resolution of each pair according to the modified Rayleigh criterion (Wu, in Hanisch 1993). Table 1 shows the numerical values of the measured resolution criterion for the blurred image, the Richardson-Lucy algorithm, the generalized iterative adaptive algorithm, and the frequency domain adaptive algorithms. All stars in our simulated image were chosen to be of intensity value (100). The distance separating each pair of stars (in pixels) is given in the first column. Neighboring pairs were all separated by a large distance, so that the PSF never covered more than one pair at a time. The measure expressed here is $R_1 = (2I_2)/(I_1 + I_3)$ where I_1 and I_3 are the peak intensities of the stars in the pair, and I_2 is the intensity at the middle point between them.

These resolution tests show that the frequency adaptive algorithm performed quite well in terms of peak resolution enhancement. It performed better than the Richardson-Lucy algorithm for most pairs. Although the generalized iterative adaptive algorithm and frequency adaptive algorithms did not perform better than the Richardson-Lucy algorithm for all pairs, the amount of resolution enhancement provided by them was significantly better than that of Richardson-Lucy for a large

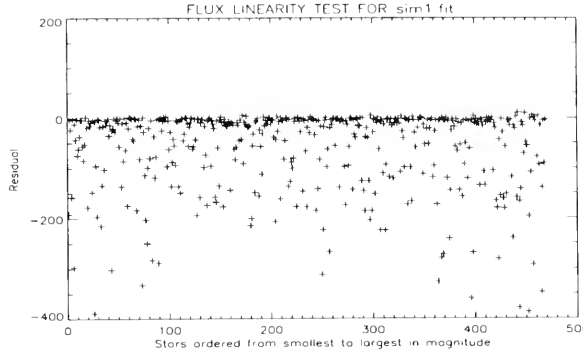


Figure 3. Simulated degraded image: "sim1.fit".

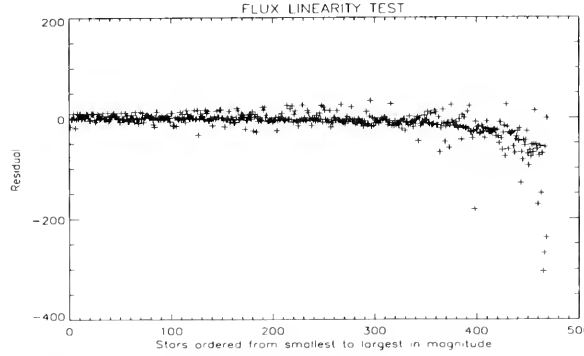


Figure 4. Richardson-Lucy restoration of "sim1.fit".

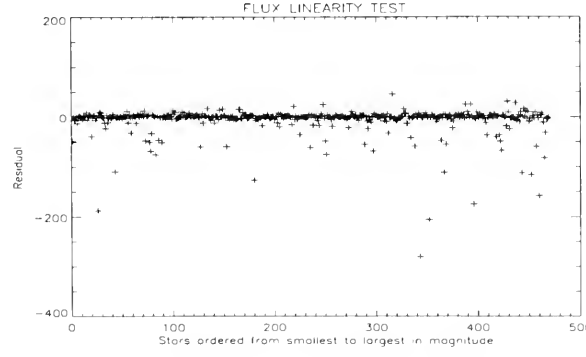


Figure 5. Generalized iterative adaptive restoration of "sim1.fit".

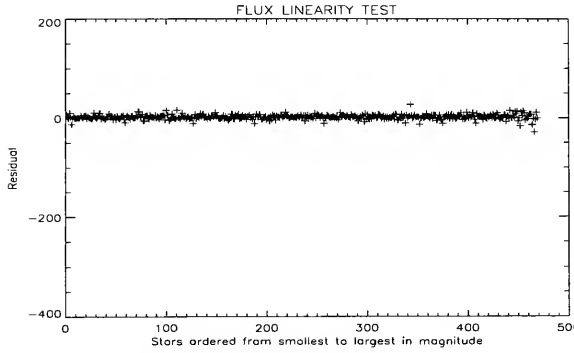


Figure 6. Frequency adaptive restoration of "sim1.fit".

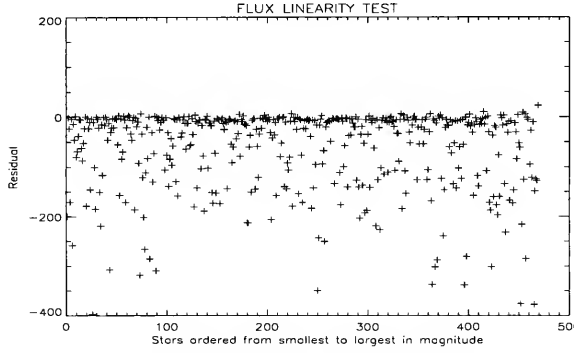


Figure 7. Spatially varying simulated degraded image: "sim3.fit".

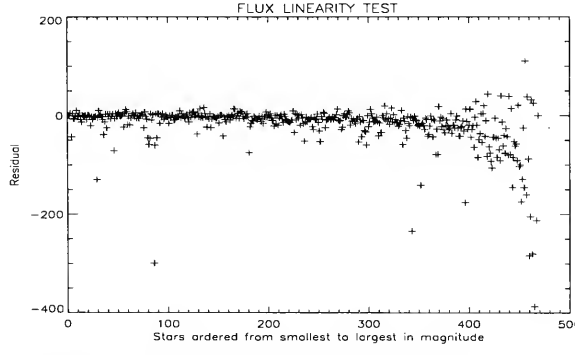


Figure 8. Generalized spatially varying restoration of "sim3.fit".

distance	R_1			
	Blurred	Lucy	Iter. Adapt	Freq. Adap.
1	1.0490925312	1.2807079554	1.0966295004	1.0897836685
3	1.1297003031	1.4553285837	1.2098157406	1.0970517397
5	1.0652104616	0.7866657376	0.8871659040	0.3229246140
7	0.8098430634	0.2116829008	0.2591046989	0.0000000000
9	0.5236188769	0.0355268493	0.0000000180	0.0000000000
11	0.3096068203	0.0043879338	0.0000000357	0.0000000000
13	0.1779404581	0.0006407523	0.0000000000	0.0000000000
15	0.1075951308	0.0002951132	0.0000000082	0.0239557754
17	0.0745507777	0.0009177358	0.0000000283	0.0328580774
19	0.0608400740	0.0056707887	0.0296860356	0.0246167779
21	0.0558008738	0.0136818457	0.0476930998	0.0694369525
23	0.0541570261	0.0185489822	0.0536854640	0.0000000000
25	0.0536803119	0.0189928524	0.0486392528	0.0000000000
27	0.0535573885	0.0157092437	0.0206987783	0.0005924603

Table 1. Resolution Tests.

number of the pairs. It should be noted that the Richardson-Lucy algorithm requires more a priori knowledge than the other two iterative algorithms, again making a direct comparison somewhat difficult. The characteristic of the last two iterative algorithms which causes the large number of zero values in columns 4 and 5 stems from the fact that these algorithms exhibit a ringing which is not present in the Richardson-Lucy algorithm. Because of the positivity constraint, a number of the center values between pairs of stars are actually negative values which have been clipped to have a value of zero, resulting in $R_1 = 0$. These results are, however, much sharper than the Richardson-Lucy results.

Based on our analysis, we have found that the frequency adaptive iterative algorithm provides an optimal choice for restoring HST data. We have applied this algorithm to some of the real Hubble data as well. One of the real images we have tested with this algorithm is the "j413_crr.fit" image of Jupiter, seen in Fig. 9. The restoration of this image using the frequency adaptive algorithm is shown in Fig. 10.

4. Discussion

From the results presented here, it is apparent that the set-theoretic based iterative restoration algorithms provide very good tools for the HST data restoration problem. We have shown that these algorithms are not dependent on any a priori knowledge of the signal and noise variances present. For the generalized iterative adaptive choice of the general regularization parameter, §2.1., there is also no dependency on the initial conditions of the iteration. In addition, the algorithms are relatively fast. This is especially true of the frequency adaptive algorithm which generally converges in a small number iterations. Given these considerations, the algorithms discussed in this paper present a very good alternative to some of the standard approaches being applied to HST data currently.

References

Andrews, H. C., & Hunt, B. R. 1977, *Digital Image Restoration*, Prentice-Hall, Englewood Cliffs

Hanisch, R. J., ed. 1993, *Restoration Newsletter*, 1, Space Telescope Science Institute, Baltimore

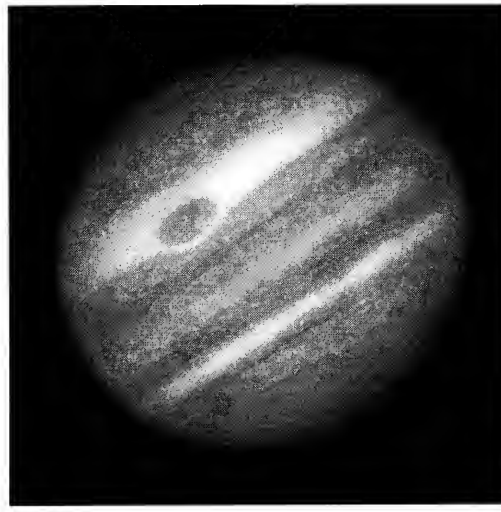


Figure 9. Blurred image "j413_crr.fit".

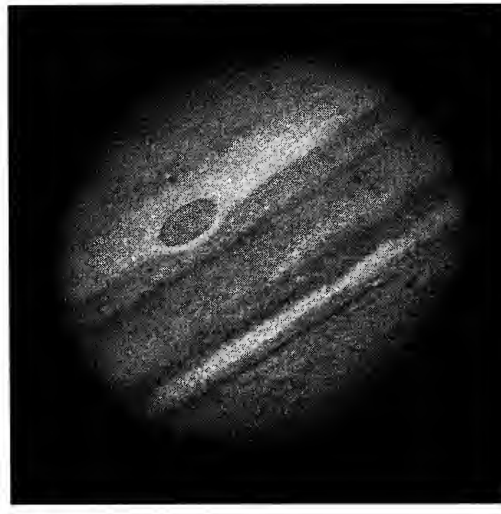


Figure 10. Frequency adaptive restoration of "j413_crr.fit".

Kang, M. G., & Katsaggelos, A. K. 1992, Proc. SPIE Conf. Visual Comm. and Image Proc., 1414

Kang, M. G., & Katsaggelos, A. K. 1993, Proc. SPIE Conf. Visual Comm. and Image Proc., 1364

Katsaggelos, A. K., Biemond, J., Mersereau, R. M., & Schafer, R. W. 1985, Proc. IEEE ICASSP, 700

Katsaggelos, A. K. 1989, Optical Engr., 28, 735

Katsaggelos, A. K., ed. 1991, *Digital Image Restoration*, Springer-Verlag, New York

Katsaggelos, A. K., Biemond, J., Schafer, R. W., & Mersereau, R. M. 1991, IEEE Trans. Signal Proc., 39, 914

Lucy, L. B. 1974, AJ, 79, 745

Schafer, R. W., Mersereau, R. M., & Richards, M. A. 1981, Proc. IEEE, 69, 432

Strand, O. N. 1974, SIAM J. Numerical Analysis, 11, 798

Weir, N. 1991, in *Proc. 3rd ESO/ST-ECF Data Analysis Workshop*, P. J. Grosbøl & R. H. Warmels, eds., European Southern Observatory, Garching, 115

White, R. L. 1991, in *Proc. 25th Conf. on Information Sciences and Systems*, 655

The Restoration of HST Images and Spectra II
Space Telescope Science Institute, 1994
R. J. Hanisch and R. L. White, eds.

Pixon-Based Image Restoration

R. C. Puetter and R. K. Piña
Center for Astrophysics and Space Science, UCSD, 9500 Gilman Drive, La Jolla, CA 92093-0111

Abstract. This paper presents the theory of the pixon, the fundamental unit of picture information, and its application to Bayesian image reconstruction. Examples of the applications of these methods to artificial and real data are presented. These examples demonstrate that pixon-based methods produce results superior to both pure Goodness-of-Fit (i.e. Maximum Likelihood) methods and the best examples of Maximum Entropy methods.

1. Introduction

The act of measurement of physical quantities inevitably introduces artifacts. These artifacts can be associated with the statistical limits of the measurement, e.g., counting statistics, as well as characteristics of the measurement device, e.g., finite resolution. Modern approaches for deducing the underlying, uncorrupted physical quantities from the recorded data often turn to Bayesian estimation in which the measurement process is statistically modeled. For astronomical imaging, the relationship between the data, $D(\vec{x})$, the underlying image, $I(\vec{x})$, and the noise, $N(\vec{x})$, is given by

$$D(\vec{x}) \;=\; \int dV_y \, H(\vec{x}-\vec{y})I(\vec{y}) + N(\vec{x}) \,. \qquad (1)$$

where $H(\vec{x})$ is the Point Spread Function (PSF) and the integration is over the volume in y-space.

The most successful methods for inverting Eq. (1) are non-linear in their approach. The simplest define a figure-of-merit, or Goodness-of-Fit (GOF) criterion, for the image and then use multi-dimensional optimization methods to maximize this function. Such methods include the familiar Least-Squares method as well as the Richardson-Lucy method (e.g., Lucy 1974). While these pure GOF techniques typically give superior inversions to Eq. (1) than do Fourier methods, they still have many undesirable properties. A common problem is over-resolution in which the algorithm attempts to fit the noise and introduces features which are unnecessary to fit the data. Lucy-Richardson reconstructions, for example, are typically stopped after an arbitrary number of iterations in an attempt to overcome this difficulty. This leaves the unpleasant (and difficult) task of determining which features are "real" and which are not. More sophisticated approaches such as Maximum Entropy (e.g., Skilling 1989) place additional constraints upon the solution based on prior expectations. These additional constraints greatly improve the quality of the solution by regularizing the problem and controlling over-resolution.

Recently, we have introduced a new image reconstruction method based on the *pixon* (Piña and Puetter 1993, Puetter and Piña 1993a). This method greatly expands upon ME methods by introducing a prior with a variable local scale. In fact, our Uniform Pixon Basis (UPB) method (Piña and Puetter 1993a) results in a "Super–Maximum Entropy" reconstruction in which entropy is maximized exactly. In our use of variable correlation length scales, pixon-based methods are similar in some respects to the multi-channel methods of Weir (1991, 1993a) or the Pyramidal Maximum Entropy techniques of Bontekoe, Koper, and Kester (1994). Unlike these techniques, however, pixon-based methods explicitly determine the appropriate local scale based on various

criteria. Our most recent Fractal Pixon Basis (FPB) method (Puetter and Piña 1993a, 1994) selects the local correlation length based on the local structural scale of the image and represents the highest performance image reconstruction method we are aware of to date.

2. The Pixon and Bayesian Estimation

To statistically model the measurement process, a number of issues must be addressed. First, there is the physics describing the measurement process. This, however, is assumed to be set down in Eq. (1). Second, a number of details related to the measurement process itself, must be properly incorporated. These include the manner in which the data was collected, e.g., as a rectangular grid of counts as is the case for astronomical imaging with a solid state detector, as well as the characteristics of the noise, etc. Finally, a number of decisions must be made regarding how Eq. (1) is to be inverted mathematically. Common assumptions here are that the image can be represented by a grid of numbers and that the integral represented in Eq. (1) can be approximated by a discrete sum over this grid. Each of these aspects of the problem, i.e., the physics, the particulars of the measurement, and the mathematical assumptions are part of the model, M, which links the image and the data.

The Bayesian approach to inverting Eq. (1) is to use Bayes' Theorem to develop a formula for the most probable value of $I(\vec{x})$. This begins by factoring the joint probability distribution of the data, image, and model triplet, D, I and M, i.e., $p(D, I, M)$:

$$\begin{aligned} p(D,I,M) &= p(D|I,M)p(I,M) = p(D|I,M)p(I|M)p(M) \\ &= p(I,M|D)p(D) = p(I|D,M)p(D|M)p(M) \end{aligned} \tag{2}$$

where $p(X|Y)$ is the probability of X given that Y is known. [Bayes' Theorem states that $p(X,Y) = p(X|Y)p(Y) = p(Y|X)p(X)$.] Equating the various terms in Eq. (2), we find

$$p(I|D,M) = \frac{p(D|I,M)p(I|M)}{p(D|M)}\,, \tag{3}$$

$$p(I,M|D) = \frac{p(D|I,M)p(I,M)}{p(D)} \propto p(D|I,M)p(I,M)\,. \tag{4}$$

From Eq. (4) we can find the M.A.P. (Maximum *A Posteriori*) image using

$$p(I|D) = \int dM\; p(I,M|D) = \int dM \frac{p(D|I,M)p(I,M)}{p(D)} \propto p(D|I,M_o)p(I,M_o) \tag{5}$$

where M_o is the model from the M.A.P. image/model pair. (The proportionality in Eq. (5) should really be an "approximately-proportional" symbol since we have dropped some terms [e.g. $p(D)$] and we have assumed that the peak in the probability distribution is representative of a typical sample from the posterior probability distribution. While this is usually the case, this is not absolutely guaranteed.)

Eq. (3) is the typical starting point of Bayesian image reconstruction in which one wishes to determine the M.A.P. image, i.e. the image which maximizes $p(I|D,M)$. (The M.A.P. image, of course, is only one of several choices for the "best image". Another sensible choice might be the average image, $\langle I \rangle = \int_{D,M} dM dD\; I\; p(I|D,M)$.)

Eqs. (4) and (5) are our preferred prescription for Bayesian image reconstruction in which both the image and the model are varied simultaneously to obtain the best combined image/model

solution, e.g., the M.A.P. image/model pair. Eq. (4) might be used directly to find the optimal image, i.e., the image paired with the model in the M.A.P. image/model pair. Alternatively, the M.A.P. image can be found using Eq. (5).

The significance of the terms in the above equations are well known. The first term, $p(D|I,M)$, is a goodness-of-fit (GOF) quantity. A typical choice for $p(D|I,M)$ is to use $p(D|I,M) = \exp(-\chi^2/2)$, where χ^2 is the chi-square of the residuals. This form for the GOF statistic allows the best unbiased estimate for the image to be determined. However, it produces a value for χ^2 which is uncharacteristically small (i.e., $\chi^2 = 0$). Another approach that we have explored is to use $p(D|I,M) = p_{\chi^2}(\chi^2)$ where $p_{\chi^2}(\chi^2)$ is the actual chi-squared distribution. This choice for the GOF criterion recognizes that the most probable value for χ^2 is not zero, but depends on the number of random variables. This GOF criterion also seems to produce quite good image reconstructions and to avoid problems associated with "over-fitting" the data.

The terms $p(I|M)$ and $p(I,M)$ are "priors". Since they do not depend on the data they can be decided *a priori*. The first of these, i.e., $p(I|M)$, is normally termed the *image prior* and expresses the *a priori* probability of an image given the model. The second, i.e., $p(I,M)$, we have termed the *image/model prior*, and expresses the *a priori* probability of both I and M. In GOF image reconstruction the prior is ignored or is effectively set equal to unity, i.e., there is no prior bias concerning the image or the model. In Maximum Entropy (ME) image reconstruction, the image prior is based upon "phase space volume" or counting arguments and the prior is expressed as $p(I|M) = \exp(\alpha S)$, where S is the entropy of the image and α is an adjustable constant that is used to weight the relative importance of the GOF and image prior.

The final quantity, $p(D|M)$, is termed the "Evidence" for the model. Actually, we normally refer to $p(M|D)$ as the Evidence, but $p(D|M)$ is proportional to this quantity under the common assumption that $p(D) = \text{const}$ and $p(M) = \text{const}$ (there generally is not an *a priori* manner for choosing between valid data sets nor between "sensible" models.) Choosing models on the basis of the Evidence allows one to "peak-up" on more favorable solutions.

3. The Pixon Concept

Eqs. (4) and (5) indicate that two quantities, the GOF criterion, $p(D|I,M)$, and the image/model prior, $p(I,M)$, are of key importance in obtaining the most probable (i.e., M.A.P.) values for the image/model. In this section we shall concentrate on the development of a pixon-based prior.

3.1. A New Pixon-Based Prior

Eq. (4) will form the basis of our pixon-based methods; our goal is to determine the M.A.P. image/model pair. Like ME methods we shall base our image/model prior on counting arguments. Unlike standard ME methods, however, we will allow certain aspects of the model to vary simultaneously with the image. By allowing both the image and model to vary simultaneously, we are optimizing our solution over a considerably larger solution space than methods which hold the model constant. Previous workers (Gull 1989, Sibisi 1989, Skilling 1991, MacKay 1992a,b) have already demonstrated the merits of varying the model and have shown the efficacy of selecting between models by maximizing the Evidence for the model.

Our pixon-based methods attempt optimize the quality of the image reconstruction by varying the portion of the model that deals with the way in which the image is mathematically. To show how the selection of image representation (we shall use the term "basis") affects the quality of the reconstruction, let us first consider the abstract nature of an image and how a generalized image/model prior might be constructed. To do this we shall follow closely the development of Piña and Puetter (1993) and Puetter and Piña (1993a). They pointed out that in an abstract sense, an image is a collection of distinguishable events which occur in distinct cells. Hence the value for the image/model prior can be determined from simple counting arguments. If there are N_i events

in cell i, and a total of n cells, then the prior probability of that particular image is:

$$p\left(\{N_i\}, n, N\right) \quad = \quad = \frac{N!}{n^N \prod_{cells,i} N_i!} = p(I, M)\,, \tag{6}$$

where $\{N_i\}$ is the set of all numbers of events in cells i, N is the total number of events, i.e., $N = \sum N_i$, and the image is now considered to be made up of these events. [In practical terms, an event is a photon count in a photon counting detector or the number of counts in units of the standard deviation of the noise in non-photon counting systems. Furthermore, these "events" are, in fact, units of information, i.e., the knowledge that something (an event) has occurred is the minimal unit of information — see Piña and Puetter (1993) for a discussion. This sense of the term "information" is somewhat different than the classical definition which is in terms of the logarithm of the number of states, i.e., Shannon information.] Also note, that the cells used in Eq. (6) are quite general. In their definition we have not specified a size, shape, or position for the cells. The cell concept simply serves to localize some collections of events.

Since the goal of our reconstruction is to determine the M.A.P. image/model pair, we must maximize the product of the image/model prior given in Eq. (6) and the GOF term. Most people have a well developed intuition regarding how to maximize the GOF term, i.e., the residuals, $R(\vec{x}) = D(\vec{x}) - \int dV_y\, H(\vec{x} - \vec{y}) I(\vec{y})$, must be comparable to the noise (strictly speaking, they should be exactly equal to the noise). Intuition concerning priors is usually less well developed. Eq. (6), however, points out the *a priori* desirable properties of the model for the image. These are that the model should contain the fewest number of cells with each containing the largest number of events consistent with maintaining an adequate GOF. We shall call these generalized cells pixons. The pixon name recognizes the pixel (or cell) heritage, and the "–on" suffix recognizes the fundamental nature of the pixon in that the pixons represent an optimal set of cells. Ideally, an image's pixons represent the smallest number of cells (of arbitrary shape, position, etc.) required to fit the data, and represent the minimum Degrees of Freedom (DOFs) necessary to specify the image. If properly selected, this set is irreducible to a smaller set. Hence pixons are the fundamental units of information in the image. Using a pixon basis for the image is the fulfillment of Occam's Razor formalized in Bayesian terms — it forces the use of the simplest model consistent with the data.

3.2. Fuzzy Pixons: A Practical Pixon-Basis Choice

The simple counting arguments presented in the section above point out the crucial features of the pixon basis: there should be the fewest number of pixons consistent with fitting the data within the accuracy allowed by the noise. In our attempts to derive suitable, practical pixon bases for image reconstruction, we adopted techniques which are similar to those adopted by other authors, i.e., a correlation length method (c.f. Weir 1991, 1993a). This approach controls the number of DOFs by reducing the independence of different parts of the image through explicit spatial correlation. This also causes the resulting degrees of freedom (or pixons) to be "fuzzy", i.e., to be localized but without hard boundaries. This still allows the use of the pixon prior of Eq. (6), although it does introduce a few computational complexities and mental hurdles for the intuition of the uninitiated. Nonetheless, the practical and performance merits of this approach seem to warrant these modest burdens.

Explicitly, then, our procedure for reducing the DOFs in the image reconstruction is to define the image in terms of a pseudo-image, $I_{\rm pseudo}(\vec{x})$, convolved with a local correlation length, $\delta(\vec{x})$:

$$I(\vec{x}) \quad = \quad \int_{V_y} dV_y\, K_{\rm pixon}\left(\frac{\vec{y} - \vec{x}}{\delta(\vec{x})}\right) I_{\rm pseudo}(\vec{y})\,, \tag{7}$$

where $K_{\rm pixon}$ is a pixon shape function and $\int_{V_y} dV_y\, K_{\rm pixon}(\vec{y}\ /\delta) = 1$. Hence our procedure is somewhat akin to representing the image in terms of non-orthogonal wavelets. As a matter of

practicality, the pseudo-image is defined on a pseudo-grid which typically has a resolution as fine or finer than the data pixel grid. The image is then also defined on a grid with the resolution of the pseudo-grid. Because of the local correlation in Eq. (7), however, the number of DOFs in the image can be greatly reduced from the number of pixels in the pseudo-grid. For example, if the local correlation length at position $\vec{x}$ is 10 pseudo-pixels then each 100 pixels (10 by 10 pixels) represent a single DOF at this location. Reduction in the DOFs greatly improves the formal value of the image/model prior and removes many of the problems commonly seen in competing methods, such as signal correlated residuals and the production of spurious sources — see below.

3.3. An Iterative Procedure for Pixon-Based Reconstruction

The iterative scheme we typically employ for calculating the M.A.P. image/model pair starts with an initial guess for the model, i.e., the spatial correlation lengths. A common starting point is to assume that the scale lengths are all equal to 1 pseudo-pixel. This is equivalent to starting out with the standard ME solution for the image. In other words, for the first estimate for the image the fuzzy pixon prior is essentially the ME prior and the GOF criterion can be chosen to be the standard chi-squared value of the residuals. In practice, however, we typically use a simple GOF solution and ignore the ME prior. This is considerably faster in practice and results in a very good first guess. The next step estimates the new local scales, holding the image fixed. This is done by maximizing $p(M|D) = \int dI \frac{p(I,M|D)p(I,M)}{p(D)} \propto p(D|I_o, M)p(I_o, M)$, i.e., finding the M.A.P. model given the fixed data and current image estimate, I_o. [Note the parallel and complementary nature of the M.A.P. model estimate relative to the M.A.P. image estimate of Eq. (5).] In our current implementations, this M.A.P. model is determined in only an approximate manner. We simply note, for example, that the prior term, $p(I_o, M)$, will insist on the largest possible correlation lengths consistent with the GOF, while the GOF term is indifferent to very small correlation lengths since they should always produce acceptable fits. Our procedure is thus simply to find the largest local correlation lengths that provide an acceptable fit. Once the local scales have been determined, a new image is calculated, etc., and the entire procedure iterated until convergence is obtained.

A bit of intuition into this procedure reveals the fundamental reasons for this method's success. Effectively, this is a fractal technique. Here we use the term "fractal" in a limited sense. Independent of the exact method for obtaining the local correlation lengths (e.g., the brute force method or an iterative method as described above), our procedure explicitly seeks the local smoothing scale at which the GOF ceases to be sensitive, i.e., we are looking for the smallest scale for which there is evidence in the data. The use of the pixon prior ensures that our procedure takes the largest, or least informative, scale consistent with fitting the data. To be explicit about our use of the term fractal, we do not mean to suggest that self-similarity plays a central role in our methods, but rather that just as in many fractal concepts, our procedure analyzes how a geometric quantity varies as the local scale is varied. In this case we ask how $p(I, M|D)$ varies as the local scale is varied, just as the definition of fractal dimension asks how does the measure (length, area, etc.) of a geometric object vary as the local scale is changed. For this reason, we have named this entire class of methods Fractal-Pixon methods and the pixon representation of the image the Fractal-Pixon Basis (FPB).

4. Sample Image Restorations

To demonstrate the advances represented by our methods, we present in this section reconstructions from both real and "mock" data sets. We have done this for several reasons. In the mock data case, the image reconstruction conditions are essentially perfect — the noise, PSF, and true answer are known *a priori* with arbitrary precision. This leaves no uncertainty in how well each algorithm has performed. However, in imaging situations we rarely encounter such benign conditions. For this reason, we have also included a real data test case in which the noise and PSF characteristics are experimentally determined. Unfortunately, the true answer is also imperfectly known, making validation of the technique more difficult.

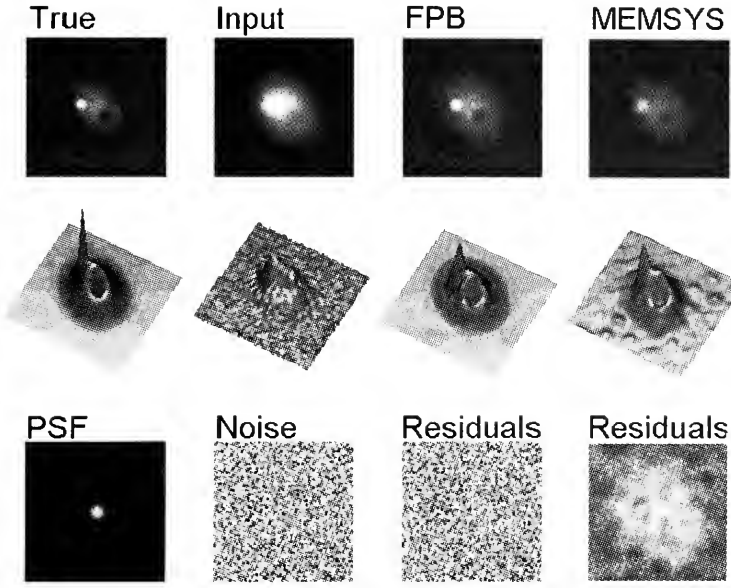


Figure 1. FPB/MEMSYS 5 Comparison for the mock data set described in the text. Reproduced from Puetter and Piña (1994).

In order to make the comparisons as fair as possible, we compare our reconstructions to those performed by other professionals well versed in the competing techniques. This avoids issues of whether the competing reconstructions are the best possible. For the real data test, we have chosen IRAS (Infrared Astronomical Satellite) 60μm survey scans of the interacting galaxy pair M51 (the "Whirlpool"). This data was used for an international image reconstruction contest at the 1990 MaxEnt Workshop (see Bontekoe 1991), which was attended by leaders in the field of image reconstruction. Hence our reconstruction of M51 will be compared to the best state-of-the-art reconstructions *circa* 1990. From comparisons like the M51 contest, experts generally agree that ME produces results superior to GOF methods (e.g., Least-Squares and Richardson-Lucy). For this reason, we shall concentrate on comparing our reconstructions to ME reconstructions (although we shall present a Richardson-Lucy reconstruction of the M51 data set as well, allowing the reader to judge the validity of these claims). The ME code we shall make our comparisons to is MEMSYS, a powerful set of ME algorithms developed by Gull and Skilling (1991). The MEMSYS algorithms probably represent the best commercial software package available for image reconstruction. The mock data reconstruction example compares our fractal pixon methods with MEMSYS 5, the most current version of the MEMSYS algorithms (see Gull and Skilling 1991). The M51 example compares our results to those of MEMSYS 3, the current version of MEMSYS at the time.

4.1. Example 1: A Mock Data Set Reconstruction

Fig. 1 presents FPB and MEMSYS 5 reconstructions of a mock data set. The MEMSYS 5 reconstructions were performed by Nick Weir of Caltech, a recognized MEMSYS expert, and were supplemented with his multi-channel correlation method which has been shown to enhance the quality of MEMSYS reconstructions (Weir 1991, 1993a). The true, noise-free, unblurred image presented in the top row is constructed from a broad, low-level elliptical Gaussian (a two-dimensional

Gaussian with different FWHMs in perpendicular directions), and two additional narrow, radially symmetric Gaussians. One of these narrow Gaussians is added as a peak on top of the low-level Gaussian. The other is subtracted to make a hole. To produce the input image, the true image was convolved with a Gaussian PSF of FWHM=6 pixels, then combined with a Gaussian noise realization. The resulting input image is displayed in the top row. The signal-to-noise ratio on the narrow Gaussian spike is roughly 30. The signal-to-noise on the peak of the low level Gaussian is about 20. The signal-to-noise at the bottom of the Gaussian hole is 12.

As can be seen, the FPB reconstruction is superior to the multi-channel MEMSYS result. The FPB reconstruction is free of the low-level spurious sources evident in the MEMSYS 5 reconstruction. These false sources are due to the presence of unconstrained degrees of freedom in the MEMSYS 5 reconstruction and are superimposed over the entire image, not just in the low signal to noise portions of the image where they are most evident. Furthermore, the FPB reconstruction's residuals show no spatially correlated structure, while the MEMSYS 5 reconstruction systematically underestimates the signal, resulting in biased photometry.

4.2. Example 2: 60 Micron IRAS Survey Scans of M51

We have also reconstructed an image from 60μm IRAS survey scans of the interacting galaxy pair M51. This data was selected for several reasons. First, M51 is a well studied object at optical, IR, and radio wavelengths. Hence "reality" for this galaxy is relatively well known. Second, as mentioned before, this particular data set was chosen as the basis of an image reconstruction contest. Consequently, there have been a number of serious attempts at performing image reconstruction on this data set by specialists in the field. Finally, the IRAS data for this object is particularly strenuous for image reconstruction methods. This is because all the interesting structure is on "sub-pixel scales" (IRAS employed relatively large, discrete detectors — 1.'5 by 4.'75 at 60μm) and the position of M51 in the sky caused all scan directions to be nearly parallel. This means that reconstructions in the cross-scan direction (i.e. the 4.'75 direction along the detector length) should be significantly more difficult than in the scan direction. In addition, the point source response of the 15 IRAS 60μm detectors (pixel angular response) is known only to roughly 10% accuracy, and finally, the data is irregularly sampled.

Our FPB reconstruction appears in Fig. 2 along with Richardson-Lucy and Maximum Correlation Method (MCM) reconstructions (Rice 1993) and a MEMSYS 3 reconstruction (Bontekoe et al. 1991) — see Gull and Skilling (1991) for a description of the MEMSYS algorithms. The winning entry to the MaxEnt 90 image reconstruction contest was produced by Nick Weir of Caltech and is not presented here since quantitative information concerning this solution has not been published; however, see Bontekoe (1991) for a gray-scale picture of this reconstruction. Nonetheless, Weir's solution is qualitatively similar to Bontekoe's solution (Weir 1993b). Both were made with MEMSYS 3. Weir's solution, however, used a single correlation length channel in the reconstruction. This constrained the minimum correlation length of features in the reconstruction, preventing break-up of the image on smaller size scales. This is probably what resulted in the "winning edge" for Weir's reconstruction in the MaxEnt 90 contest (Weir 1993b).

As can be seen from Fig. 2, our FPB-based reconstruction is superior to those produced by other methods. The Richardson-Lucy and MCM reconstructions fail to significantly reduce image spread in the cross-scan direction, i.e., the rectangular signature of the 1.'5 by 4.'75 detectors is still clearly evident, and fail to reconstruct even gross features such as the "hole" (black region) in the emission north of the nucleus; this hole is clearly evident in optical images of M51. The MEMSYS 3 reconstruction by Bontekoe is significantly better. This image clearly recovers the emission hole and resolves the north-east and south-west arms of the galaxy into discrete sources. Nonetheless, the level of detail present in the FPB reconstruction is clearly absent, e.g., the weak source centered in the emission hole (again, this feature corresponds to a known optical source).

To assess the significance of the faint sources present in our FPB reconstruction, in Fig. 3 we present our reconstruction overlaid with the 5 GHz radio contours of van der Hulst et al. (1988). The radio contours are expected to have significant, although imperfect, correlation with the far infrared

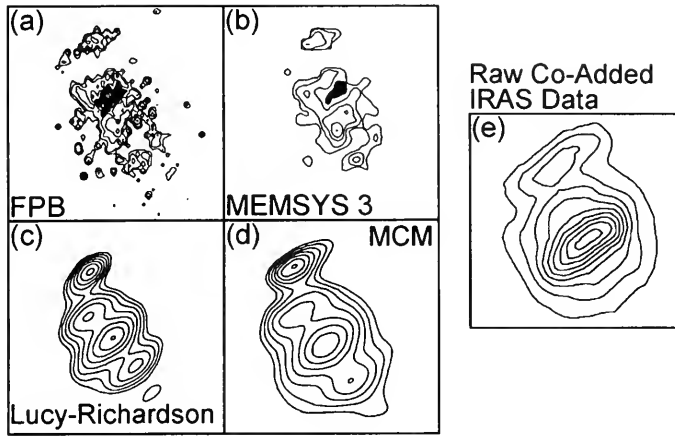


Figure 2. Image reconstruction of the interacting galaxy M51. (a) FPB-based reconstruction. (b) MEMSYS 3 Reconstruction. (c) Richardson-Lucy reconstruction. (d) MCM reconstruction. (e) Raw, co-added IRAS 60μm data. Figure of panel (b) reproduced from Bontekoe et al. (1991), by permission of the authors. Figures of panel (c) and (d) reproduced from Rice (1993) by permission of the author. This figure is reproduced from Puetter and Piña 1993b.

emission seen by IRAS. Hence a comparison of the two maps should provide an excellent test of the reality of structures found in our reconstruction. Also identified in Fig. 3 are several prominent optical sources and Hα knots.

As can be seen, the reconstruction indicates excellent correlation with the radio. The central region of the main galaxy and its two brightest arms align remarkably well, and the alignment of the radio emission from the north-east companion and the IRAS emission is excellent. Furthermore, for the most part, whenever there is a source in the reconstruction which is not identifiable with a radio source, it can be identified with either optical or Hα knots. An excellent example is the optical source in the hole of emission to the north-east of the nucleus of the primary galaxy or the bright optical source to the north-west of the nucleus (both labeled "Opt" in Fig. 3). Because of the excellent correlation with the radio, optical, and Hα images, we are quite confident that all of the features present in our reconstruction are real.

Aside from the fact that most of the sources can be identified with emission at other wavelengths, the residual errors in our reconstruction are much smaller than in the MEMSYS 3 reconstruction. As pointed out by Bontekoe et al. (1991) the peak flux in the MEMSYS 3 reconstruction is 2650 units. The residual errors are correlated with the signal and lie between 0 and 430 units. By contrast, the peak value in the FPB reconstruction is 3290 units, the residuals are uncorrelated with the signal, and the residuals lie between -9 and 17 units. (The contour levels for the MEMSYS 3 and FPB reconstructions of Fig. 3 are identical and are 150, 300, 600, 1200, and 2400 units.) Furthermore, the large deviation residuals in the FPB reconstruction are due to systematic errors involving incomplete scan coverage of M51. Fortunately, these errors do not lie under the significant flux emitting portions of the M51 image. The residual errors associated with emitting regions in M51 are significantly smaller ($\sigma \approx 1$ unit) and show a roughly Gaussian distribution. Full appreciation of the sensitivity of our technique is only obtained once the reconstruction has been flux calibrated. Formally, the residual error over the majority of the image is 2.7 mJy. This compares with the 280 mJy, 90% completeness limit for the IRAS Faint Source Survey. The largest residual systematic

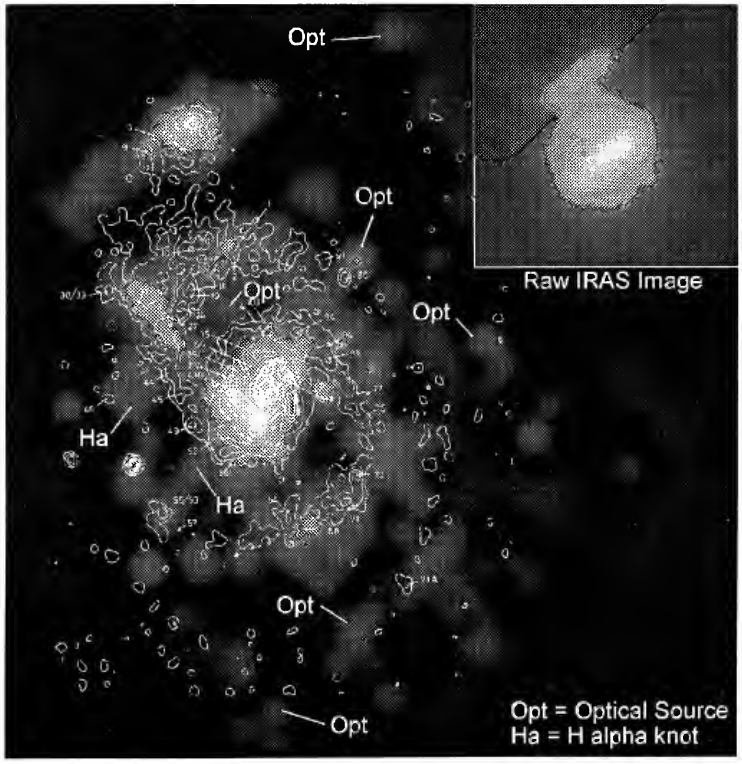


Figure 3. Details of the FPB reconstruction of the IRAS 60μm survey of M51. The overlaid contours are the 5 GHz data of van der Hulst et al. (1988). Also noted are several of the stronger 60μm features that can be identified with optical features or Hα knots. Figure reproduced from Puetter and Piña (1994).

errors associated with incomplete sampling of the M51 region correspond to 50 mJy. This is still a factor of more than five times fainter than the IRAS Faint Source Survey limit.

5. Conclusions

In conclusion, we have introduced a new concept, the pixon, the use of which provides large improvements in the ability to reconstruct images from blurred, noisy data. The pixon is the fundamental and indivisible unit of information required to describe the underlying signal within the accuracy allowed by the data. In this regard it is an idealized concept. However, this paper also presents a practical and capable approximation to this ideal for a broad range of problems, i.e., the fuzzy, Fractal-Pixon Basis (FPB). This basis uses the local spatial scale relevant to the underlying signal to constrain the inversion of the equations governing the measurement process. In doing so, this method provides performance superior to pure GOF (Maximum Likelihood) and ME methods. Some of the advantages of pixon-based methods are the elimination of signal-correlated residuals

and the production of spurious sources typical of other methods. Practical examples from the realm of astronomical image reconstruction show that pixon-based methods can offer large improvements in resolution as well as the detection of extremely weak features in the data.

Acknowledgments. The authors would like to thank a number of people for their valuable contributions to this work. We would especially like to thank Nick Weir of Caltech for numerous fruitful discussions regarding image processing and for graciously performing the multi-channel MEMSYS 5 reconstructions presented in this paper. We would also like to thank Romke Bontekoe and Do Kester for providing the M51 IRAS test data set as well as for many helpful conversations. Finally, the authors would like to thank Steven Gull and John Skilling for a number of conversations that greatly expanded our understanding of Bayesian methods in general, and how the pixon fits into Bayesian theory in particular. This work was supported by NASA and the National Science Foundation.

References

Bontekoe, T. R. 1991, in *Maximum Entropy and Bayesian Methods*, W. T. Grady, Jr. & L. H. Schick, eds., Kluwer Academic Publishers, Dordrecht, 319

Bontekoe, T., R., Kester, D. J. M., Price, S. D., de Jonge, A. R. W., & Wesselieus, P. R. 1991, A&A, 248, 328

Bontekoe, T. R., Koper, K., & Kester, D. J. M. 1994, A&A, in press

Gull, S. F. 1989, in *Maximum Entropy and Bayesian Methods*, J. Skilling, ed., Kluwer Academic Publishers, Dordrecht, 53

Gull, S. F., & Skilling, J. 1991, *MemSys5 Quantified Maximum Entropy User's Manual*, Maximum Entropy Data Consultants, Ltd., Royston

Lucy, L. B. 1974, AJ, 79, 745

MacKay, D. J. C. 1992a, Neural Computation, 4 (3), 415

MacKay, D. J. C. 1992b, Neural Computation, 4 (5), 698

Piña, R. K., & Puetter, R. C. 1993, PASP, 105, 630

Puetter, R. C., & Piña, R. K. 1993a, SPIE Proceedings, 1946, 405

Puetter, R. C., & Piña, R. K. 1993b, in *Science with High Spatial Resolution Far-Infrared Data*, Infrared Processing and Analysis Center, California Institute of Technology, Pasadena, in press

Puetter, R. C., & Piña, R. K. 1994, Proceedings 1993 Max Ent Conference, Santa Barbara, in press

Rice, W. 1993, AJ, 105, 67

Sibisi, S. 1989, in *Maximum Entropy and Bayesian Methods*, J. Skilling, ed., Kluwer Academic Publishers, Dordrecht, 389

Skilling, J. 1989, in *Maximum Entropy and Bayesian Methods*, J. Skilling, ed., Kluwer Academic Publishers, Dordrecht, 45

Skilling, J. 1991, in *Maximum Entropy and Bayesian Methods*, W. T. Grady, Jr. & L. H. Schick, eds., Kluwer Academic Publishers, Dordrecht, 267

van der Hulst, J. M., Kennicutt, R. C., Crane, P. C., & Rots, A. H., 1988, A&A, 195, 38

Weir, N. 1991, in *3rd ESO/ST-ECF Data Analysis Workshop*, P. Grosbøl & R. H. Warmels, eds., European Southern Observatory, Garching, 115

Weir, N. 1993a, J. Opt. Soc. Am., in press

Weir, N. 1993b, private communication

The Restoration of HST Images and Spectra II
Space Telescope Science Institute, 1994
R. J. Hanisch and R. L. White, eds.

Iterative/Recursive Deconvolution with Application to HST Data

James M. Coggins

Department of Computer Science, CB 3175 Sitterson Hall, University of North Carolina at Chapel Hill, Chapel Hill, NC 27599-3175

Laura Kellar Fullton and Bruce W. Carney

Department of Physics and Astronomy, CB 3255 Phillips Hall, University of North Carolina at Chapel Hill, Chapel Hill, NC 27599-3255

Abstract. A new deblurring algorithm has been developed involving both iteration and recursion and that is linear, flux-conserving, noise resistant, and faster to converge than extant iterative deblurring methods. Mathematical analysis shows that the recursive component of the algorithm provides the accelerated convergence. A demonstration is provided using a simulated star field image blurred using an approximation to the point spread function of the Hubble Space Telescope.

1. Introduction

The problem presented by the blurred images obtained from the Hubble Space Telescope is a common one throughout science. Given a digital signal degraded by (approximately) stationary blurring and an estimate of the point spread function defining the blur, we seek to recover an improved approximation to the unblurred signal. Such blurs occur in 1-D chemical spectra and 2-D and 3-D images from microscopes, telescopes, photographs, CT and MRI scanners, satellite sensors, and scintigrams (nuclear medicine images). Lucy (1974) lists a series of problems in astronomy that depend on the ability to remove blur from image or spectral measurements. Jansson and his associates (1968, 1970a, 1970b) have studied extensively the problem of removing or characterizing the blurring function of spectrographs.

The same basic iterative deconvolution algorithm has been independently developed in astronomy and medical image processing and has migrated from those fields to many others. This iterative algorithm is highly sensitive to image noise and to errors in the estimate of the point spread function. Similar modifications to the basic algorithm have been proposed in both fields. One common modification involves introducing a nonlinear relaxation factor into the algorithm to accelerate convergence and suppress noise, but then image flux is no longer conserved. Thus, in astronomy for example, the deblurred image is no longer suitable for photometry since nonlinear, spatially varying changes in intensity are produced by the modified deblurring procedures.

This paper introduces a new iterative and recursive deconvolution procedure that is flux-conserving, linear, and relatively insensitive to noise or to error in the estimate of the point spread function. Results of a mathematical analysis of the iterative/recursive method will be presented, and a demonstration of the algorithm will be presented based on the blurring function of the Hubble Space Telescope.

2. Blurring

The blurring characteristics of a shift-invariant linear system are characterized by the point spread function (PSF) of the system. The observed image, i, is the result of convolving the PSF, h, with

the true image, t:

$$i = t * h. \tag{1}$$

Since convolution in space is equivalent to multiplication in frequency, the blurred image can be also be obtained as follows:

$$\begin{aligned} I &= TH \\ i &= F^{-1}(I) \end{aligned}$$

where capitals T and H denote the Fourier spectra of images t and h, respectively.

A diffraction-limited Hubble Space Telescope (HST) would have a PSF just over 2.5 pixels in radius (0.″1 at 633 nm) which would include half of the second bright Airy ring. The actual PSF is over 68 pixels in radius (3.″0) due to the spherical aberration caused by malformation of the HST's primary mirror (Burrows 1991).

The PSF of an imaging system can be estimated mathematically or empirically. The computational construction of a PSF involves complicated and computationally demanding ray tracing methods based on a complete description of the optical system. Empirical estimates of the PSF can often be obtained by imaging a relatively bright, isolated point source. For example, an empirical estimate of the PSF of a fluorescence microscope can be obtained by imaging a tiny polystyrene bead coated with a fluorescent marker dye (Fay 1986). PSFs derived by either the computational or empirical approach can be used in iterative deblurring methods. Inaccuracies in the PSF estimate will cause some aspects of the true image data to be interpreted as image noise and limit the degree to which the image can be deblurred.

A PSF can be summarized compactly by *encircled energy*, the integral of the PSF as a function of distance, ρ, from the center of the PSF. Indexing the PSF using polar coordinates (ρ, θ) with the origin at the center of the PSF, the encircled energy is

$$E(\rho) = \int_0^{\rho} \int_0^{2\pi} i(\rho, \theta)\, d\theta\, d\rho \tag{2}$$

Since optical systems do not add energy to a signal, $E(\rho) \to 1$ as $\rho \to \infty$. Thus, a spatially extensive PSF attenuates the signal at the center, in-focus location, complicating detection of faint structures in complex images. In crowded fields, the blurred images of nearby objects can interfere with each other, requiring detection algorithms to distinguish intensity peaks due to objects from intensity peaks of artifacts arising from constructive interference between the blurred images of separate objects. Nonlinear algorithms are prone to create false objects at intensity peaks due to constructive interference or to overfitting particularly bright noise peaks.

The effect of a deblurring algorithm can be summarized by plotting the encircled energy of a point source as a function of radius. This measure is computed in the same manner as $E(\rho)$ above with the polar coordinate system centered on the correct location of the point source.

3. Inverse Filtering

Since blurring can be implemented by a product in the frequency domain, it may be possible to deblur an image by a division in the frequency domain,

$$T = \frac{I}{H} \tag{3}$$

This works if the modulation transfer function (MTF), H, contains no zeros, but such a situation rarely occurs in practice because MTFs of most image acquisition systems do contain zeros throughout the measured spatial frequency range. When H contains no zeros but has components with very small values, limited numerical precision can create significant errors in the image estimate. Such a blur irretrievably discards some information about the structure of the true image, and there is

no longer a unique solution to the deblurring problem. In practice, constraints on the form of the reconstructed image (e.g., smoothness constraints) can be used to infer appropriate values for the deleted components.

When inverse filtering is not possible, the objectives of deconvolution are modified. Given an observed, blurred image, i_0, and an estimate of the PSF, h, derive an image $\hat{t}$ such that $i_0 = \hat{t} * h$. The image $\hat{t}$ is the estimate of the true image, t. This operation is called pseudoinverse filtering, or approximate deconvolution, which is usually implemented using iterative methods.

4. Iterative Deconvolution

Iterative deconvolution begins by guessing what the true image, t_0, might be. This initial guess is denoted $\hat{t}_1$. If this guess is correct, then the convolution $\hat{t}_1 * h$ will produce the observed image, i_0. If the guess is wrong, it can be corrected based on the residual between the observed image and the blurred guess: $(i_0 - \hat{i}_1)$. In fact, the correction might be simply to add that difference to $\hat{t}_1$.

The observed image will serve as the initial guess. The first step in an iterative deblurring method, then, is to blur the observed image again. While this might seem surprising, it is correct because the observed image is the closest data we have to the true image. A flat field could be used as the initial guess, but then the correction factor would be the observed image, so the estimate used for the second iteration would be the observed image anyway.

Formally, the Basic Iterative Deconvolution (BID) procedure is defined as follows:

$$\hat{t}_1 = i_0 \tag{4}$$

$$\hat{i}_k = \hat{t}_k * h \tag{5}$$

$$\hat{t}_{k+1} = \hat{t}_k + (i_0 - i_k) \tag{6}$$

This approach is essentially the Jacobi method for solving simultaneous linear systems as applied to signal processing by Van Cittert (1931), extended by Jansson (1968, 1970a, 1970b) and independently developed by Iinuma (1967a, 1967b). It converges to the correct inverse filter if one exists; otherwise it can be terminated after a finite number of iterations to obtain an approximation to the true image. The iterative method can also be understood as computing a power series expansion of the inverse filter $1/H$ based on the identity

$$\frac{1}{x} = \sum_{i=0}^{n} (1 - x)^i \tag{7}$$

Convergence is slow after the first few iterations, with rapidly diminishing returns. The approach also is very sensitive to noise in the signal or error in the estimate of the PSF.

Mathematically, the effect of the BID algorithm is most easily understood via the frequency domain. This discussion follows Metz (1969) and Kawata (1980a, 1980b). Transforming equations 5 and 6 into the frequency domain gives

$$\hat{T}_{k+1} = \hat{T}_k + (I_0 - \hat{T}_k H) \tag{8}$$

Reordering the terms and factoring yields

$$\hat{T}_{k+1} = \hat{T}_k(1 - H) + I_0 \tag{9}$$

This equation shows how each iteration is obtained from the previous one. Combining this form of the iteration with the initial condition in Eq. 4, one can construct successive iterates

$$\begin{aligned} \hat{T}_1 &= I_0 \\ \hat{T}_2 &= I_0(1 + (1 - H)) \\ \hat{T}_3 &= I_0(1 + (1 - H) + (1 - H)^2) \end{aligned} \tag{10}$$

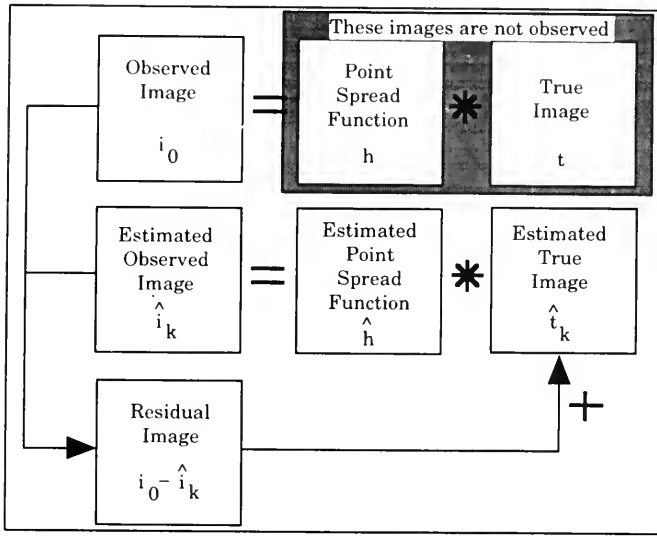


Figure 1. The basic iterative deconvolution (BID) algorithm

to find the general form for the successive estimates,

$$\hat{T}_k = I_0 \sum_{i=0}^{k-1} (1 - H)^i \tag{11}$$

which contains the power series expansion for the inverse filter. This expression for the kth estimate of the true image can be reduced to a more compact form using the series identity

$$\sum_{i=0}^{k-1} x^i = \frac{x^k - 1}{x - 1} \tag{12}$$

to yield

$$\hat{T}_k = I_0[\frac{(1 - H)^k - 1}{(1 - H) - 1}] = I_0[\frac{1 - (1 - H)^k}{H}] \tag{13}$$

When $k = 1$, the estimate of the true image's spectrum is I_0. As $k \to \infty$, the filter approaches $1/H$ and the estimate approaches the correct answer.

Define

$$\mathbf{R}_0^{(k)} = \frac{1 - (1 - H)^k}{H}. \tag{14}$$

$\mathbf{R}_0^{(k)}$ is the effective filter after k iterations of the BID algorithm. The estimate of the true image's spectrum after k iterations of the BID algorithm can be written

$$\hat{T}_k = \hat{T}_1 \mathbf{R}_0^{(k)} = \mathbf{R}_0^{(k)}(\hat{T}_1) \tag{15}$$

5. Constrained Iterative Deconvolution

The BID algorithm converges slowly, so its convergence might be accelerated by multiplying the residual by a constant before adding it into $\hat{t}_k$, thereby boosting the correction factor. The reason this method fails requires examination of the content of the residual image.

The residual image contains (a) the aspects of the image that have not yet been accounted for as instances of the PSF, and (b) any structure in the observed image that arises from any process other than a convolution of the original image with the PSF. Any such extraneous information is *noise*. Sources of such noise include data errors occurring after the optical image is formed, errors in the estimate of the PSF, variations in the real PSF across the image field, Poisson-distributed photon noise in imagery of faint sources, or image effects caused by signals (cosmic rays in astronomy, scattered X-rays in medical imaging) impinging on the detector that are not subject to the blurring effects of the optical system. The portion of the residual image attributable to the PSF decreases in each iteration, but the noise content passes through essentially unchanged. Multiplying the residual image by some factor to boost the speed of convergence also strengthens the noise being added back into the estimate of the object. This noise quickly dominates the signal and makes the procedure unstable.

The instability can be fought by introducing nonlinear constraints intended to suppress corrections due to noise. Typical constraints include clamping the values in t_k so that they lie within a known reasonable dynamic range, or limiting the amount of adjustment that can occur at any pixel (i.e., clamping $(i_0 - i_k)$). These limits may be imposed by setting thresholds or by introducing a damping, or relaxation function.

Such constraints have two problems. First, they actually just delay the inevitable; noise problems are still serious. Second, the nonlinearities introduced in the constraints may cause artifacts in the estimate of the true image, destroying the quantitative information that is the point of many scientific studies. The mathematical analysis of the nonlinear methods is much more difficult; depending on the nonlinearity employed, a closed form for the estimate may not exist since different pixels may be treated in different ways.

6. A New Enhancement to the Iterative Method

We propose a new method for deconvolution that extends the BID algorithm by recursively invoking itself on the residual at each iteration.

The residual image, $i_0 - i_k$, used to correct $\hat{t}_k$ in the BID algorithm is the difference between *blurred* images, therefore *it is a blurred image itself.* The correction to $\hat{t}_k$ that is needed is a deblurred version of $i_0 - i_k$. An estimate of this deblurred residual can be obtained by recursively invoking this deblurring procedure on the residual image along with the same estimate of the PSF. The residual image of the previous recurrence serves as both the observed image $i_0^{(r+1)}$ and the initial estimate, $\hat{t}_1^{(r+1)}$, in the new recurrence. The new recurrence blurs its observed image and computes the residual as before. The procedure recurs again to deblur this residual image. The recursion continues for as many levels as desired. Finally, at the deepest recursion level, the BID algorithm is applied for some small number of iterations. The resulting $\hat{t}_k^{(r)}$ is passed up to the next higher recurrence and used to correct $\hat{t}_i^{(r-1)}$.

This algorithm has several advantages over the previous approaches. First, there are no nonlinear steps in the algorithm, so its results are free of artifacts; in particular, flux is conserved. Second, by applying only a few iterations at each level (except perhaps the top level) the algorithm converges rapidly because no recurrence of the procedure is allowed to proceed far enough to enter the range of minimal returns. Third, the method is stable because during each iteration the restoration is stopped, a simpler, related problem is defined, and the restoration continues. The Discussion section will elaborate on the stability of the algorithm.

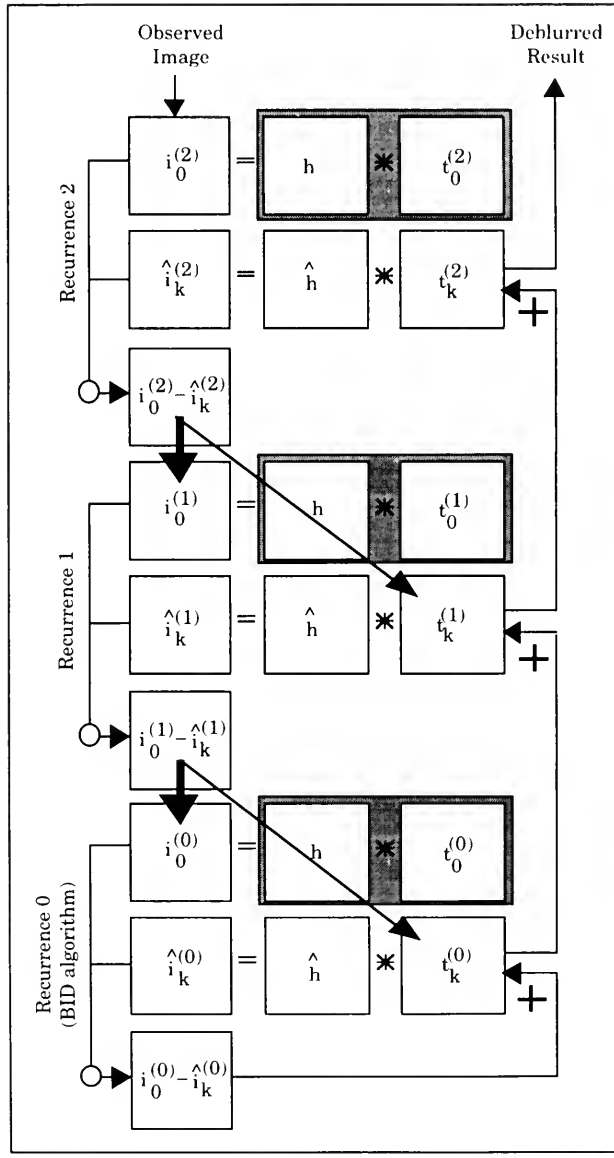


Figure 2. Iterative/Recursive Deconvolution

7. Mathematical Analysis of the Iterative/Recursive Algorithm

This section presents the results of a mathematical analysis of the Iterative/Recursive Algorithm. (For details of the analysis procedure, see Coggins 1993.) This analysis is based on an iterative/recursive deblurring procedure with k iterations at each of r recursion levels. The number of iterations could be different at different recursion levels, but keeping them equal simplifies this analysis. An expression will be derived for the effective linear filter applied by the entire iterative/recursive deblurring process for several values of k and r. The examples will be developed in the frequency domain, with capitals representing the Fourier spectra of the image given by the corresponding lower case letter.

The deblurring algorithm at the lowest recursion level is the BID algorithm, the effect of which is given in Eq. 14.

The key to understanding the iterative/recursive algorithm is to expand the restoration function $\mathbf{R}_r^{(k)}(x)$ for various values of r and k. We have worked out these effects for various (r, k) pairs. The method involves algebraically bootstrapping the expressions for (r, k) from the expressions for $(r-1, k)$ and $(r, k-1)$. This analysis unrolls the iterations and unfolds the recursions, resulting in a complicated expression in terms of BID iterations. The contributions of iteration and recursion to this algorithm can be compared by converting the BID iterations into truncated power series form and then algebraically simplifying into a sum of powers of $(1 - H)$. That is, we can express the restoration functions in the form

$$\mathbf{R}_r^{(k)}(x) = x[\sum_{i=0}^{f}(1 - H)^i] \tag{16}$$

where the limit of the summation, f, is a function of r and k. We have worked out the summation relationships shown in Table 1. Line 1 describes the BID algorithm. Line 2 describes an algorithm consisting of recursions alone. Subsequent lines explore the interaction of recursions and iterations. Clearly, the iterative/recursive algorithm is producing a closer approximation to $1/H$, as indicated by the rapidly growing summation limit, than either the recursions or the iterations alone.

Table 1. Summation Limits f for r Recursion Levels with k Iterations Each

r	k	f
0	N	N-1
N	1	N
1	2	4
1	3	9
2	2	10
3	2	22

This analysis demonstrates that the iterative/recursive algorithm does converge toward the inverse filter and that it provides a better approximation to the inverse filter for a given amount of computation than the BID algorithm.

8. Test Results

Since the most demanding test of a deconvolution technique is to apply it to an image which contains many objects, we have generated an artificial image of NGC 6293, a globular star cluster for which we have HST Planetary Camera (PC) data. The PC is a mosaic of 4 800×800 Texas Instruments CCD detectors (numbers 5–8). The filters used for our observations were F555W and F785LP

which are similar to the V and I filters of the ground-based Johnson photometric system. Our artificial image simulates a 600 second exposure from PC 7 with the F555W filter. Busko (1993) presents results from this algorithm for the simulated cluster image available from ST ScI. We present restoration results for real HST data in a separate paper (Fullton 1993).

To generate the artificial image, we used the color-magnitude diagram (CMD) of NGC 6293 as determined from ground-based photometry (Janes 1991). We used the stellar coordinates as measured by Janes and Heasley in their images (Janes, private communication), to generate a 300×300 pixel (about 13×13 arc second) image in which the measured magnitudes, V, were converted into photon counts and the stellar coordinates were used as the location of point sources with these fluxes. It was necessary to correct for the difference in pixel scale between the ground-based CCD images and those obtained with HST.

Because the photometry of Janes and Heasley is not complete at fainter magnitudes, i.e., $V > 21$, it was necessary to correct the observed data to account for those stars which were missed either because they were too faint to be detected or because of crowding effects. In order to do this, we used the luminosity function of the globular cluster M92 (Stetson and Harris 1988) which tells us how many stars we expect to see as a function of magnitude. M92 was chosen because its luminosity function has been well-determined and because its CMD morphology is similar to that of NGC 6293, indicating that the two clusters are likely similar in age and chemical composition. The number of artificial stars generated in each 0.25 magnitude bin from $V = 21$ to $V = 24.25$ was determined from the luminosity function and corrected for the observed area of NGC 6293 by a multiplicative factor determined by requiring that the luminosity function reproduce the observed counts from Janes and Heasley brighter than $V = 21$. The magnitude and position of each star were selected randomly using an algorithm for generating uniformly-distributed random numbers based on Convex's pseudo-random number generator `RAN`. The generated magnitudes were constrained to fall within the given magnitude bin for each star. The final frame contained 425 artificial stars which were added to a uniformly bright image whose intensity was chosen to be representative of the sky background in actual HST images of the cluster.

The image was blurred by convolution with an estimated HST PSF computed using the Tiny Tim software (Krist 1992). Additive Poisson noise was simulated in the blurred image by adding to each pixel a random value drawn from a Poisson distribution with a mean equal to the square root of the intensity at each pixel. The resulting image is shown in Fig. 3. No attempt was made to simulate cosmic ray hits or saturation effects which are often present in real HST data. These effects create portions of an image that are not the result of convolution of a light source with the PSF, so such areas will need to be removed or ignored when using the iterative/recursive deblurring algorithm on real HST images.

The simulated image was deconvolved using the iterative/recursive method coded in C++ on a Convex C-240. The restoration reported here used 5 recursion levels with 3 iterations per level. Thus, $3^5 = 243$ iterations of the BID algorithm at the lowest recursion level were performed. Restoration of the 300×300 image took approximately 38 CPU minutes, but this computation involved forward and inverse discrete Fourier transforms in each iteration. This time can be drastically shortened by performing the entire computation in the frequency domain, but it is then not possible to observe intermediate results during the restoration.

The restored image is shown in Fig. 4. The Fourier transform-induced ringing may appear exaggerated due to contrast stretching for this display. Fig. 5 shows intensity plots through the same row in the blurred and deconvolved images. The star located at column 163 in Fig. 5 has a simulated magnitude of 23.44. The star at pixel 194 has magnitude 21.14. The plot from the deconvolved image demonstrates the low amplitude of the ringing compared to the flux in the central peak of the star. This ringing does not violate flux conservation and does not interfere with intensity measurements of the stars in the restored images. The ringing may interfere with detection of stars at the faintest limits. As a star's magnitude decreases well below the sky brightness, it becomes increasingly difficult to distinguish the star from the star combined with the sky background. However, detection in the deconvolved image will be easier since the light from the star is not blurred over such a large area.

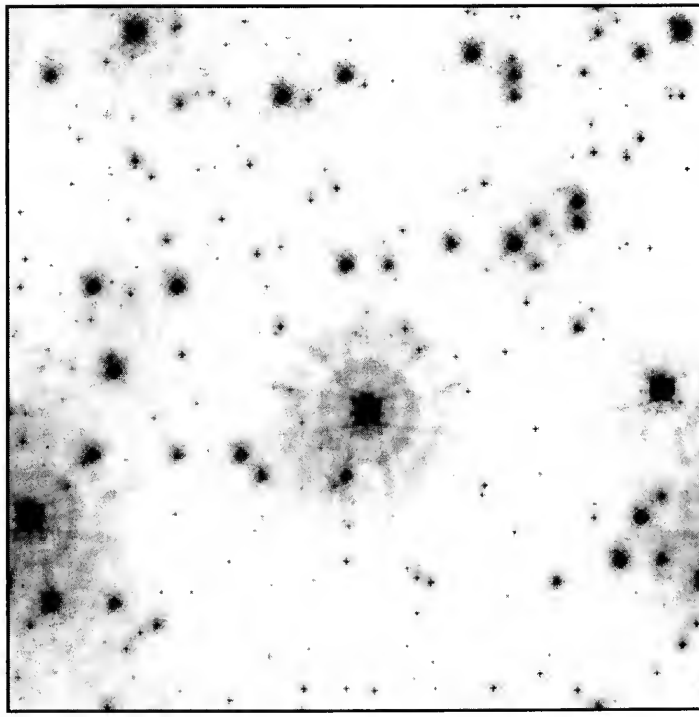


Figure 3. A 300×300 pixel simulated image of globular cluster NGC 6293. Stellar magnitudes and positions from Janes and Heasley (1991) for the brighter stars, magnitudes estimated from the luminosity function of M92 (Stetson 1988) and position generated randomly for the fainter stars. Point sources with the above locations and magnitudes were convolved with an HST PSF calculated using Tiny Tim (Krist 1992).

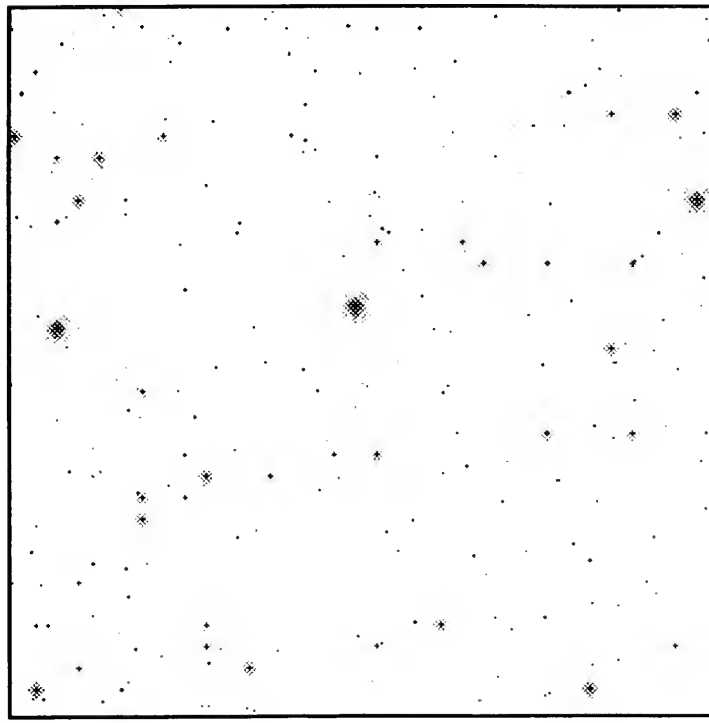


Figure 4. The deconvolved simulated image of NGC 6293 after 3 iterations at each of 5 recursion levels of the iterative/recursive deconvolution algorithm described in this paper.

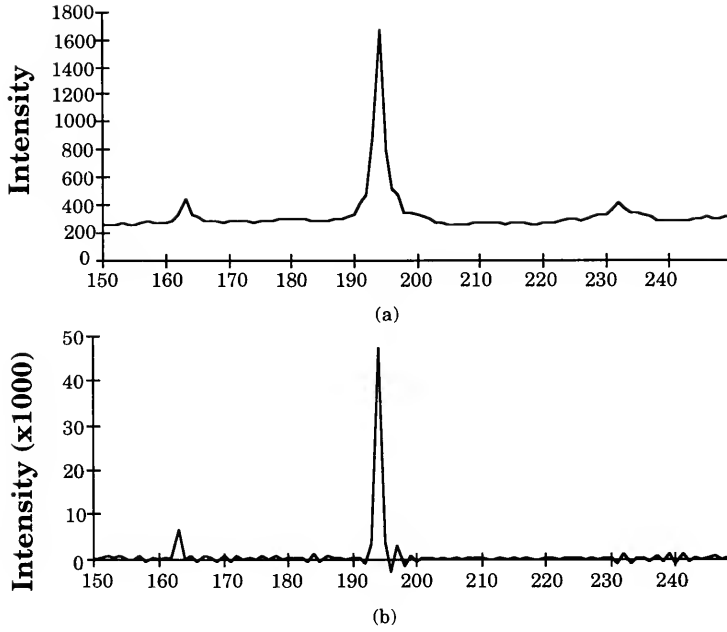


Figure 5. (a) An intensity plot of row 106 in the simulated image shown in Fig. 3. The centers of two stars lie in this row, a 23.44 mag star at pixel 163 and a 21.14 mag star at pixel 194. (b) An intensity plot of the same row in the deconvolved image. The signal at pixel 232 is from a star centered in a nearby row, so its signal disappears in the deblurred image. Note the different intensity scales in these two plots.

To test the suitability of this deconvolution technique for astronomical purposes, we have measured the brightnesses of all the stars in the deconvolved output image. For each star in the input star list, the sum of the pixel intensities in circular apertures of various radii centered on the stars' input coordinates was computed. The brightness of the uniform sky background was estimated by taking the mean of the median pixel values in several "star-free" regions in the deconvolved image. This sky value was subtracted from each pixel within the stellar apertures to correct for the flux contributed by the sky.

Fig. 6 is a plot of input magnitude (the magnitude used to create the artificial image) vs. the magnitude measured in an aperture of radius 3 pixels centered on each star in the output image after correcting for the sky background. A 45 degree line indicates perfect agreement between input and output magnitudes. A constant additive aperture correction was applied to the magnitudes since an aperture of only 3 pixels does not properly account for all the stellar flux due to the ringing apparent in the image. The aperture correction was determined by plotting the measured magnitudes of 10 bright, relatively uncrowded stars as a function of aperture radius. It was estimated from these growth curves that at an aperture of radius 6 pixels the ringing had subsided sufficiently that an accurate measurement of the stellar flux could be made. Thus the average difference between the magnitude in the 6 pixel aperture and the 3 pixel aperture (0.008 ± 0.001 mags) was defined to be the aperture correction and was added to the magnitudes of all the stars shown in Fig. 6. This correction corresponds to less than 1% of the stellar flux. Note that in the smaller aperture, all the stars were measured too bright. Thus, while one might expect the smaller aperture to yield a fainter estimate of the stellar brightness, this is not necessarily the case in the restored image. Due to the requirement of flux conservation, the central pixels are enhanced in brightness to compensate for the ringing, which may include negative flux values. However, as shown in Fig. 6, the aperture correction is constant with magnitude. Thus the percentage overestimate of brightness is the same for all the stars. Six stars with input coordinates within 1 pixel from another star were omitted from Fig. 6. Twenty-nine other stars had coordinates so near the edges of the frame that the 3 pixel radius aperture extended outside the boundaries, so these stars were also omitted.

There are two important observations to be made concerning Fig. 6. First, it is obvious that the relationship between input and output magnitudes is linear. This is to be expected since there are no nonlinear operations in the iterative/recursive algorithm. Second, the agreement between input and output magnitudes is extremely good, with most stars falling on the line of slope unity. Stars whose nearest neighbor was less than 2 pixels away are plotted as squares, those whose nearest neighbor was greater than 2 but less than or equal to 5 pixels away are plotted as crosses, and those whose nearest neighbor was over 5 pixels away are plotted as dashes. Almost all of the stars whose output magnitudes deviate substantially from the input magnitudes are stars with nearby neighbors, some of whose light is being included in the 3 pixel aperture used. This problem would be worse if a larger aperture had been used to encompass all the light from the star. All the affected stars are measured systematically too bright because of this added light from nearby stars. It is important to note that this is a result of our use of aperture photometry in crowded conditions and is not an artifact of the deconvolution.

9. Comparison With Other Techniques

There have been several investigations of the photometric properties of HST image restoration techniques utilizing a range of different algorithms. While we do not intend to describe each in detail, it is interesting to compare the results of our simulation with those obtained using some other methods. Holtzman et al. (1991) studied the use of the Richardson-Lucy method (Richardson 1972, Lucy 1974) in restoration of PC images. They used that technique to restore a simulated image of a crowded field. The scatter in the magnitudes measured in the deconvolved data was comparable to that in the original blurry image, so the restoration did not provide a quantitative improvement. Additionally, they found a systematic nonlinearity of several tenths of a magnitude over a 4 mag range. In their deconvolution of an actual image of stars in the 30 Doradus nebula, the errors in the

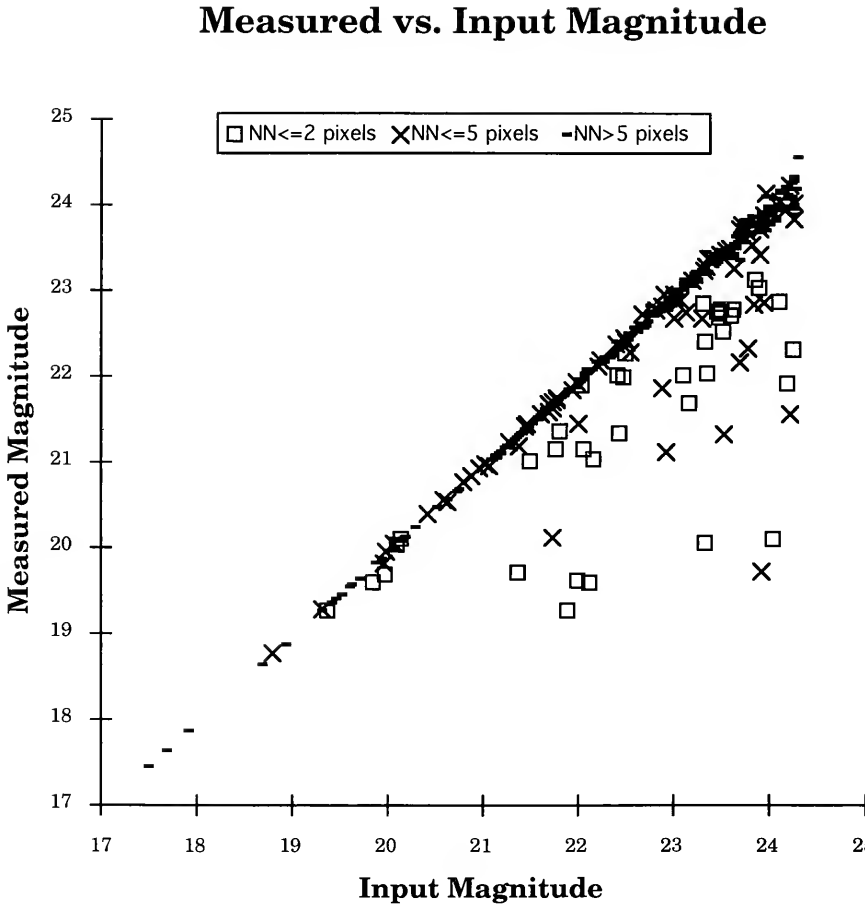


Figure 6. A plot of input magnitude vs. the magnitude derived from a 3 pixel aperture radius in the deconvolved image. An additive aperture correction determined from 10 bright stars has been applied to all the magnitudes. Stars whose nearest neighbor was within 2 pixels are plotted as squares. Stars whose nearest neighbor was greater than 2 but less than or equal to 5 pixels away are plotted as xs. Stars whose nearest neighbor was over 5 pixels away are plotted as small dashes.

deconvolved photometry judged by calculated sigmas and residuals were worse than those in the original image. Additionally, the deconvolved image was noisier and revealed no new sources not present in the original image.

Cohen (1991) investigated the reliability of the Maximum Entropy method (Skilling 1989, Gull 1989) in restoring astronomical images while preserving photometric accuracy. Her investigations involved restoring simulated images created by blurring point sources with a 3 pixel FWHM Gaussian. She discovered that the Maximum Entropy software systematically recovered the fainter objects as too faint. This nonlinearity was a function of the degree of contrast of the object with the sky and the degree of crowding. She concluded that then currently available Maximum Entropy software could not be used where photometric reliability was required.

Finally, while it is more difficult to compare the previous results with those from Faint Object Camera (FOC) data, primarily because of its lower signal-to-noise ratio (SNR), Thomson et al. (1992) studied photometric techniques with FOC images. They used Maximum Entropy and did one test with the Richardson-Lucy method. They found that they needed a magnitude-dependent aperture correction, a nonlinearity correction (to account for the nonlinearity of the Maximum Entropy method which systematically underestimated the flux in point sources by about 1 standard deviation) and a zero-point correction which depended on the SNR of the observed image. Their photometric error was typically about 0.1 mag from the deconvolved data, comparable to what they derived using standard profile-fitting photometric techniques on the original images (see Stetson 1987, 1993).

Busko (1993) has performed a systematic investigation of the photometric properties of several restoration algorithms including the iterative/recursive algorithm presented here. He found that photometry performed on images restored with the iterative/recursive algorithm produced the best results in terms of linearity and reduction in confusion between neighboring stars. Thus, it is not surprising that in comparison with the studies mentioned above, our simulation indicates that photometric results from the iterative/recursive procedure may be more reliable (and certainly less complicated) than those obtained with either the Richardson-Lucy or the Maximum Entropy algorithms. An application of the iterative/recursive algorithm to HST images of the globular cluster NGC 6352 is presented elsewhere in this volume (Fullton 1993) and demonstrates the excellent photometric restoration possible with this algorithm.

10. Discussion

The iterative/recursive algorithm at all but the lowest-level recurrence interrupts each iteration of the BID algorithm, defines a new, related deblurring problem, and then executes the algorithm on the related problem. The power of the algorithm lies in the way the new problem is defined.

Repeated blurring by the PSF produces a progressively simpler image, eventually blurring the image to a constant. Only a Gaussian blur is guaranteed not to create spurious local structure during progressive blurring (Koenderink 1992), but any realistic PSF will eventually blur an image to a constant.

Each difference image $(I_0 - I_k^{(\tau)})$ estimates *the derivative of image intensity with respect to blurring by the PSF*. This kind of derivative is central to the multiscale geometric image decompositions used in recent computer vision research; Coggins (1992) has called the computer vision approach based on progressive blurring *spatial spectroscopy*. Each recursion level implements the next higher order of this differentiation. The revised problem defined in each recursion, then, is to deblur a higher-order derivative of the image with respect to blurring by the PSF.

Blom (1993) has recently shown that the computation of higher-order derivatives is more robust against noise than the computation of lower-order derivatives in the presence of (white) noise. This result is surprising because the ill-posedness of differentiation has become an informal axiom of image and signal processing. Blom's result can be partly explained by the larger spatial extent of higher-order derivative kernels. The larger spatial extent, required to fit in all of the ripples of the high-order derivative operator, regularizes the data by smoothing out small scale noise. In the

present application, the higher-order derivatives are simpler, more regular images, so the iterative deblurring algorithm in the last recursion level is executed on a simpler deblurring problem having less noise contamination. The deblurred result is then used to correct the estimates of the true image at the previous recurrence.

Some reblurring algorithms (Kawata 1980b) that blur the observed image before iterating in order to suppress noise appear to be roughly equivalent to performing one recursion level of the iterative/recursive algorithm.

The iterative/recursive restoration algorithm will still encounter numerical problems if the noise is severe enough to persist in the lowest recursion level. However, the recursions make the iterative/recursive algorithm more robust than the BID algorithm, linear and flux-conserving, faster to converge toward the inverse filter than the BID algorithm, and faster to run than other deblurring algorithms (Busko 1993). The power of this algorithm suggests that research into restoration of chemical spectra, medical images, and other restoration problems should be revisited using this new algorithm.

References

Blom, J., B. M., ter Haar Romeny, A. Bel, & J. J. Koenderink 1993, J. Vis. Comm. & Image Rep., in press

Blom, J. 1991, Ph.D. Dissertation, University of Utrecht, The Netherlands

Burrows, C. J., et al. 1991, ApJ, 369, L21

Busko, I. 1994, this volume

van Cittert, P. H. 1931, Z. Phys, 69, 298

Coggins, J. M. 1992, in *2nd International Conference on Visualization in Biomedical Computing*, R. Robb, ed., SPIE, 1808, 145

Coggins, J. M., Fullton, L., & Carney, B. 1994, Comp. Vis. Graph. & Image Proc., submitted

Cohen, J. G. 1991, AJ, 101, 734

Fay, F. S., Fogarty, K. E., & Coggins, J. M. 1985, in *Optical Methods in Cell Physiology*, P. de Weer & B. Salzburg, eds., John Wiley and Sons, New York, 51

Fullton, L. K., Carney, B. W., Janes, K. A., Coggins, J. M., & Seitzer, P. 1994, this volume

Gull, S. F. 1989, in *Maximum Entropy and Bayesian Methods*, J. Skilling, ed., Kluwer, Dordrecht, 53

Holtzman, J. A., et al. 1991, ApJ, 369, L35

Iinuma, T. A., & Nagai, T. 1967a, Int. J. Appl. Rad. & Iso., 18, 325

Iinuma, T. A., & Nagai, T. 1967b, Phys. Med. Biol., 12, 501

Janes, K. A., & Heasley, J. N. 1991, AJ, 101, 2097

Jansson, P. A., Hunt, R. H., & Plyler, E. K. 1968, J. Opt. Soc. Am., 58, 1665

Jansson, P. A., Hunt, R. H., & Plyler, E. K. 1970a, J. Opt. Soc. Am., 60, 596

Jansson, P. A. 1970b, J. Opt. Soc. Am., 60, 184

Kawata, S., & Ichioka, Y. 1980a, J. Opt. Soc. Am., 70, 762

Kawata, S., & Ichioka, Y. 1980b, J. Opt. Soc. Am., 70, 768

Koenderink, J. J. 1992, *Solid Shape*, MIT Press, Cambridge

Krist, J. 1992, "The Tiny Tim User's Manual," Space Telescope Science Institute, Baltimore

Lucy, L. B. 1974, AJ, 79, 745

Metz, C. E. 1969, Ph.D. dissertation, Dept. of Biomedical Electronic Engineering, University of Pennsylvania

Richardson, W. H. 1972, J. Opt. Soc. Am. 62, 52

Skilling, J. 1989, in *Classical Maximum Entropy, Maximum Entropy and Bayesian Methods*, J. Skilling, ed., Kluwer, Dordrecht, 45

Stetson, P. B. 1987, PASP, 99, 191

Stetson, P. B. 1994, this volume

Stetson, P. B., & Harris, W. E. 1988, AJ, 96, 909

Thomson, R. C., Schade, D. J., Elson, R. A. W., Mackay, C. D. & Wilkins, T. N. 1992, MNRAS, 259, 104

The Restoration of HST Images and Spectra II
Space Telescope Science Institute, 1994
R. J. Hanisch and R. L. White, eds.

Adaptive Least Squares Image Restoration Using Whitening Filters of Short Length

B. Bundschuh and D. Schneider

Institut für Nachrichtenverarbeitung, Universität Siegen, D-57068 Siegen, Germany

H. Schulz

Radioastronomisches Institut, Universität Bonn, D-53113 Bonn, Germany, and
Astronomisches Institut der Ruhr–Universität Bochum, D-44780 Bochum, Germany

Abstract. This paper presents a regularized least squares algorithm for the restoration and reconstruction of images. A whitening filter of short length provides the regularization function. An adaptive version of the algorithm is developed by matching a weighting function to the regularization function. The adaptive regularization simultaneously leads to proper noise suppression and enhanced resolution of discontinuities.

1. Introduction

The restoration or reconstruction of signals which are degraded by a low pass type system represented by the point spread function $h(j_1, j_2)$ and corrupted by noise is a typical ill-posed problem. Nevertheless, image restoration has proven to be a valuable tool to compensate for the degradation of the imaging quality of the Hubble Space Telescope (HST). Fig. 1 schematically shows the generation of the measured data. Basically we want to estimate the signal **s**, e.g., an image written as a vector, from the measured data **g** which contains a noise term **n**. This signal can not be measured directly but only after having passed a linear system represented by the matrix **H**:

$$\mathbf{g} = \mathbf{H} \cdot \mathbf{s} + \mathbf{n} \,. \tag{1}$$

In the case of a shift invariant system **H** is a Toeplitz matrix.

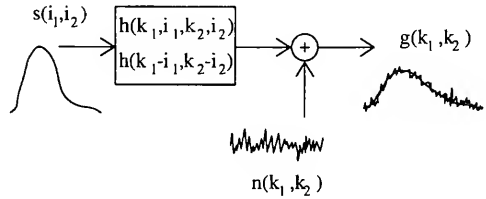


Figure 1. Generation of the measured data.

2. Restoration Based on Adaptively Weighted Regularization

Since linear shift invariant algorithms only lead to an unsatisfactory compromise between resolution and smoothness of the estimate $\hat{\mathbf{s}}$ (Andrews and Hunt 1977), we derive an adaptive method. This

is done by incorporation of the inverse weighting matrices $\mathbf{V}^{-1}$ and $\mathbf{A}^{-1}$ into the quadratic error term E_Q. The Toeplitz matrix $\mathbf{F}$ determines a linear shift invariant regularization function:

$$E_Q = \hat{\mathbf{s}} \cdot \mathbf{F}^T \mathbf{V}^{-1} \mathbf{F} \cdot \hat{\mathbf{s}} + (\mathbf{g} - \mathbf{H}\hat{\mathbf{s}})^T \cdot \mathbf{A}^{-1} \cdot (\mathbf{g} - \mathbf{H}\hat{\mathbf{s}}) = \text{min.}\,, \tag{2}$$

$$\hat{\mathbf{s}} = (\mathbf{H}^T \mathbf{A}^{-1} \mathbf{H} + \mathbf{F}^T \mathbf{V}^{-1} \mathbf{F})^{-1} \cdot \mathbf{H}^T \mathbf{A}^{-1} \cdot \mathbf{g}\,. \tag{3}$$

The optimization of $\mathbf{V}$ and $\mathbf{A}$ is carried out by minimization of the expectation D_Q of the squared distance between the original signal and the estimated signal:

$$D_Q = \left\langle (\hat{\mathbf{s}} - \mathbf{s})^T \cdot (\hat{\mathbf{s}} - \mathbf{s}) \right\rangle = \text{min.}\,, \tag{4}$$

$$\partial D_Q / \partial V = 0 \quad \wedge \quad \partial D_Q / \partial A = 0\,. \tag{5}$$

Application of matrix differentiation rules and matrix algebra leads to the same matrix equation in both cases which means that the results are not independent of each other:

$$\mathbf{F}^T \mathbf{V}_{\text{opt}}^{-1} \mathbf{F} \cdot \langle \mathbf{s}\mathbf{s}^T \rangle \cdot \mathbf{H}^T = \mathbf{H}^T \mathbf{A}_{\text{opt}}^{-1} \cdot \langle \mathbf{n}\mathbf{n}^T \rangle \tag{6}$$

Since there are infinitely many solutions we are free to select one of the matrices such that the simplest solution is obtained:

$$\mathbf{A}_{\text{opt}} = \langle \mathbf{n}\mathbf{n}^T \rangle \,| \quad \mathbf{V}_{\text{opt}} = \mathbf{F} \cdot \langle \mathbf{s}\mathbf{s}^T \rangle \cdot \mathbf{F}^T \tag{7}$$

The matrices are determined by the statistics of the original signal and the noise which are described by the autocorrelation matrices. If Eq. (7) is inserted into Eq. (3), the resulting estimate is similar to the estimate provided by a shift variant Wiener filter (Andrews and Hunt 1977). If the signal and the noise are Gaussian distributed it is also similar to a Maximum a Posteriori estimate (Andrews and Hunt 1977). Theoretically this is the optimal least squares solution of our restoration problem. But there are two reasons why it does not work in practical applications. First, the autocorrelation matrix of the original signal must be known quite accurately. Therefore, the algorithm is extremely sensitive against modeling errors and usually $\langle \mathbf{s}\mathbf{s}^T \rangle$ can not be estimated from the data at hand with sufficient accuracy. Second, its size is prohibitively large. In the case of a 512×512 pixel image the size of the matrix is 262144×262144, which exceeds even the memory capacity of today's computers.

A suboptimal but practical algorithm is obtained if we replace the general weighting matrices $\mathbf{V}$ and $\mathbf{A}$ by the diagonal matrices $\mathbf{V}_D$ and $\mathbf{A}_D$. Now the number of parameters of the statistical model is drastically reduced and this makes the algorithm much more robust. Eq. (2) and Eq. (3) must be replaced by

$$E_Q = \hat{\mathbf{s}} \cdot \mathbf{F}^T \mathbf{V}_D^{-1} \mathbf{F} \cdot \hat{\mathbf{s}} + (\mathbf{g} - \mathbf{H}\hat{\mathbf{s}})^T \cdot \mathbf{A}_D^{-1} \cdot (\mathbf{g} - \mathbf{H}\hat{\mathbf{s}}) = \text{min.}\,, \text{and} \tag{8}$$

$$\hat{\mathbf{s}} = (\mathbf{H}^T \mathbf{A}_D^{-1} \mathbf{H} + \mathbf{F}^T \mathbf{V}_D^{-1} \mathbf{F})^{-1} \cdot \mathbf{H}^T \mathbf{A}_D^{-1} \cdot \mathbf{g}\,. \tag{9}$$

The optimization of $\mathbf{A}_D$ and $\mathbf{V}_D$ is again carried out by minimization of the expectation D_Q of the squared distance between the original signal and the estimated signal:

$$D_Q = \left\langle (\hat{\mathbf{s}} - \mathbf{s})^T \cdot (\hat{\mathbf{s}} - \mathbf{s}) \right\rangle = \text{min.}\,, \tag{10}$$

$$\partial D_Q / \partial v_{ii} = 0 \;\forall\, i \quad \wedge \quad \partial D_Q / \partial a_{kk} = 0 \;\forall\, k\,. \tag{11}$$

Application of matrix differentiation rules and some matrix algebra show that the diagonal elements of the optimal diagonal matrices must be equal to the diagonal elements of the previously determined optimal general weighting matrices so that

$$a_{kk\text{opt}} = \langle \mathbf{n}\mathbf{n}^T \rangle \,|_{kk} \quad v_{ii\,\text{opt}} = \mathbf{F} \cdot \langle \mathbf{s}\mathbf{s}^T \rangle \cdot \mathbf{F}^T \,|_{ii}\,. \tag{12}$$

Now we can derive a linear adaptively regularized restoration algorithm. Eq. (3) is rearranged, and using components we obtain a linear system of equations:

$$\begin{aligned}
&\partial E_Q/\partial \hat{s}(q_1,q_2) = 0 \;\forall\; q_1,q_2 \Rightarrow \\
&\textstyle\sum_{i_1}\sum_{i_2}\hat{s}(i_1,i_2)\sum_{j_1}\sum_{j_2} w(j_1,j_2)f(j_1-i_1,j_2-i_2)f(j_1-q_1,j_2-q_2)+ \\
&+\textstyle\sum_{i_1}\sum_{i_2}\hat{s}(i_1,i_2)\sum_{k_1}\sum_{k_2}\sigma_n^{-2}(k_1,k_2)h(k_1-i_1,k_2-i_2)h(k_1-q_1,k_2-q_2) = \\
&= \textstyle\sum_{k_1}\sum_{k_2}\sigma_n^{-2}(k_1,k_2)g(k_1,k_2)h(k_1-q_1,k_2-q_2)
\end{aligned} \tag{13}$$

$a_{kk\,\mathrm{opt}}$ is determined from the local noise variance. Basically this means that in the case of HST data, the Poisson distributed noise is approximated by Gaussian distributed noise with signal dependent variance:

$$a_{kk\,\mathrm{opt}} = \langle \mathbf{n}\mathbf{n}^T \rangle\,|_{kk} = \sigma_n^{-2}(k_1,k_2) \tag{14}$$

$\sigma_n^{-2}(k_1,k_2)$ is estimated from the measured data. Since noise is Poisson distributed its variance is proportional to the local signal value. The HST point spread function leads to strongly smoothed images so that some additional smoothing using the operator $\mathrm{sm}(k_1,k_2)$ causes only minor errors in the estimation of the underlying noise-free signal. Therefore we estimate:

$$\sigma_n^{-2}(k_1,k_2) \propto \left(g(k_1,k_2) * \mathrm{sm}(k_1,k_2)\right)^{-1} \tag{15}$$

The reciprocals of the diagonal elements of the inverse weighting matrix determine the optimal weighting function $w(i_1,i_2)$:

$$\begin{aligned}
v_{Dii\,\mathrm{opt}} &= \left\langle \left(\sum_{j_1}\sum_{j_2} f(j_1,j_2)s(i_1-j_1,i_2-j_2) \right)^2 \right\rangle \\
\Rightarrow\; w(i_1,i_2) &= \left\langle \left(\sum_{j_1}\sum_{j_2} f(j_1,j_2)s(i_1-j_1,i_2-j_2) \right)^2 \right\rangle^{-1}
\end{aligned} \tag{16}$$

Obviously the weighting function is matched to the regularization function in a remarkably simple way. Unless a simulation is performed, $s(i_1,j_1)$ is not known a priori! Therefore, an estimate must be determined from the available data. Before the iteration starts the measured data themselves provide the best available estimate of the original signal. If we assume that during the iteration the estimate converges to the original signal, estimates of the optimal weighting function can be determined successively from intermediate results of the iteration. In practical applications this tricky approach converges. The algorithm is self-updating at least up to a certain degree (Bundschuh 1992):

$$w(i_1,i_2) = \left(\sum_{j_1}\sum_{j_2} f(j_1,j_2)\hat{s_\mu}(i_1-j_1,i_2-j_2) \right)^{-2} \qquad \text{iteration no. } \mu+1\,. \tag{17}$$

3. Optimization of the Regularization Function

In order to complete the design of the restoration algorithm we have to select the regularization function $f(i_1,i_2)$. In Eq. (8) diagonal weighting matrices were introduced which means that correlations are neglected. The result of the convolution operation $\mathbf{Fs}$ should be uncorrelated, or in other words, "white". Therefore $f(i_1,i_2)$ is also called a "whitening filter" (Yaroslavsky 1985). Approximations to this type of filter are differential operators (Twomey 1963), the Laplace operator (Pratt 1978), and the linear predictor (Bundschuh 1993):

$$s_{\mathrm{white}}(i_1,i_2) = f_{\mathrm{opt}}(i_1,i_2) * * s(i_1,i_2) \tag{18}$$

The application of the optimal whitening filter leads to a constant power density spectrum. Obviously the modulus of the transfer function of the optimal whitening filter must be reciprocal to the modulus of the transfer function of the original signal. Its phase still remains to be determined by some other optimality criterion. The resulting whitening operator is obtained by means of an inverse Fourier transform:

$$F_{\text{opt}}(m_1, m_2) = |S(m_1, m_2)|^{-1} e^{j\gamma(m_1, m_2)} \tag{19}$$

$$f_{\text{opt}}(i_1, i_2) = \frac{1}{I_1 I_2} \sum_{m1} \sum_{m2} |S(m_1, m_2)|^{-1} e^{j\gamma(m_1, m_2)} e^{2\pi j(m_1 i_1 / I_1 + m_2 i_2 / I_2)} \tag{20}$$

The size of the whitening operator which is used in practice should be as small as possible in order to save computer time. In this paper an operator with 5×5 coefficients is used for the processing of simulated as well as measured data. Basically this truncation reduces the decorrelation properties. The optimality criterion for the phase function is minimal spatial dispersion which minimizes the truncation error. Partial differentiation with respect to the unknown phase function and equating the result to zero leads to a system of non-linear equations:

$$\sum_{i_1} \sum_{i_2} i_1^2 i_2^2 f_{\text{opt}}(i_1, i_2) f^*_{\text{opt}}(i_1, i_2) = \min. \tag{21}$$

$$\frac{\partial}{\partial \gamma(k_1, k_2)} \sum_{i_1} \sum_{i_2} i_1^2 i_2^2 f_{\text{opt}}(i_1, i_2) f^*_{\text{opt}}(i_1, i_2) = 0 \; \forall \; k_1, k_2 \; . \tag{22}$$

A general solution cannot be found analytically, but some special solutions can easily be obtained. The restriction to real operators enables the use of symmetry properties of the phase function. A special solution which leads to minimal dispersion is

$$\gamma(k_1, k_2) = 0 \; . \tag{23}$$

From Eq. (19) we see that basically the power density spectrum $\Phi_{ss}(m_1, m_2)$ and the autocorrelation function $\varphi_{ss}(i_1, i_2)$ of the original signal should be known. If available, some a priori knowledge of the statistics of the original signal can be used. In many cases images can be modeled quite well as two-dimensional first order Markov processes (Pratt 1978):

$$\Phi_{ss}(m_1, m_2) = \frac{c_1}{1 + c_2(m_1^2 + m_2^2)} \; . \tag{24}$$

Since the coarse structure of the underlying signal can be seen from the measured data, the power density spectrum of the measured data was used for the processing of simulated as well as measured data instead of the unknown power spectrum of the original signal:

$$\Phi_{ss}(m_1, m_2) \cong |G(m_1, m_2)|^2 \; . \tag{25}$$

Eventually the self-updating iterative procedure which is used for the weighting function $w(i_1, i_2)$ using the intermediate results of the iteration can be used for the power density spectrum:

$$\Phi_{ss}(m_1, m_2) \cong |\hat{S}_\mu(m_1, m_2)|^2 \qquad \text{iteration no.}\, \mu + 1 \; . \tag{26}$$

Generally a variation of the regularization function within the image should enable the adaptation to local statistical image properties. Fig. 2 shows an example for whitening operator $f(j_1, j_2)$. Obviously it is a high pass filter which is reasonable for regularization functions. Qualitatively it shows some similarity to the Laplace operator.

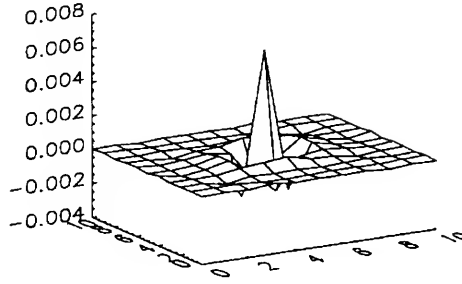


Figure 2. Whitening (decorrelation) operator.

4. The Boundary and Resampling Problem

As sketched in Fig. 3, registered images are influenced by fractions of the original signal from outside the domain of measured data. We distinguish three characteristic regions:

Inner Region Points in this region are covered by the complete point spread function. Therefore we can use the maximum available information on these data and should be able to obtain a reasonable restoration.

Outer Region Points in this region are not covered by the point spread function at all. Therefore we do not have any information on these data and they cannot be restored.

Intermediate Region Points in this region are covered by only a fraction of the point spread function. If we move from the inner region towards the outer region, the available information on these data decreases more and more. Therefore, the quality of the restoration decreases in the same manner.

If the region of the restoration covers only the inner region, it is assumed implicitly that the original signal has a region of support which is equal to the support of the restored data. But in reality the underlying original signal (the whole sky in astronomy) has an unlimited region of support. Therefore the assumption of a limitation can lead to inconsistencies and to oscillations at the boundaries of the region of the restoration.

A reduction of the boundary problem is possible by suitable pre-processing of the measured data, e.g., by even continuation (Yaroslavsky 1985). A rigorous solution which completely avoids the boundary problem is the extension of the region of restoration by the extent of the point spread function. If the extent of the measured data is 512×512 and the extent of the point spread function is 256×256, the extent of the region of the restoration should be (512 + 256)×(512 + 256) = 768×768. In this way the measured data are not influenced at all by original data from outside the region of restoration and no inconsistencies can occur. Basically this means that the measured data are padded with zeroes up to the extent of the real measured data plus twice the extent of the point spread function. The extent of the region of virtual measured data should be (512 + 256 + 256)×(512 + 256 + 256) = 1024×1024. The concentration on the real measured data is obtained by means of the weighting function $\sigma_n^{-2}(k_1, k_2) = 0$ outside the region of real measured data. We

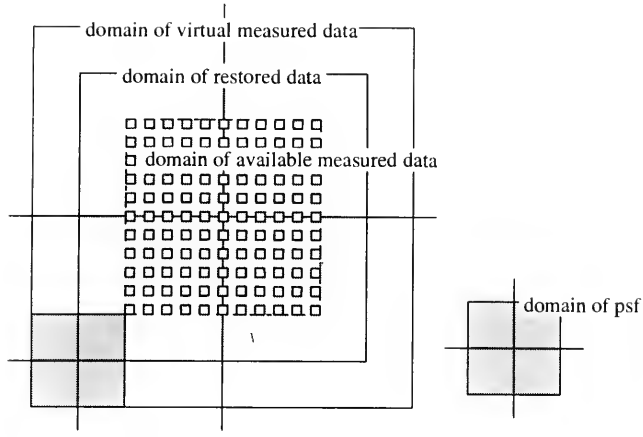


Figure 3. Characteristic categories of data.

would obtain the same effect if all the padded data were available with infinite variance of the noise outside the region of real measured data.

A fact which first appears to be peculiar is that the domain of restored data must be *larger* than the domain of available measured data!

The weighting function $\sigma_n^{-2}(k_1, k_2)$ can also be used to incorporate resampling on a finer grid directly into the restoration process. The measured data are again padded with zeroes, but now the zeroes are positioned between the real data points (see Fig. 3). Again we assume that the variance of the noise tends to infinity where we do not have measured data. Of course this kind of resampling is only possible if the point spread function is known completely on the finer grid. Basically it should even be possible to incorporate an interpolation directly into the restoration process if the measured data are given on a non-uniform and/or a non-rectangular grid.

5. Processing of Simulated Data

Fig. 4 shows a simulated original signal which consists of a random collection of clouds and impulses. The clouds are created by threshholding a superposition of two arrays of strongly correlated and weakly correlated random numbers so that a pre-specified fraction of the image remains black. The stars are simulated as a random collection of impulses which are distributed randomly over the image. The measured point spread function from the data file F372M.FITS, which was used to generate the simulated measured data, is also displayed.

Fig. 5 shows the simulated measured data together with the restored image. The noise is white and Poisson distributed with an SNR of 10 dB. The degree of blurring as well as the noticeable amount of noise is clearly visible. The restoration of the clouds is excellent, while the stars remain somewhat blurred. The reason for this difference is the use of the power density spectrum of the measured data which overestimates the correlation inherent in the data.

Figure 4. Random test signal and measured point spread function.

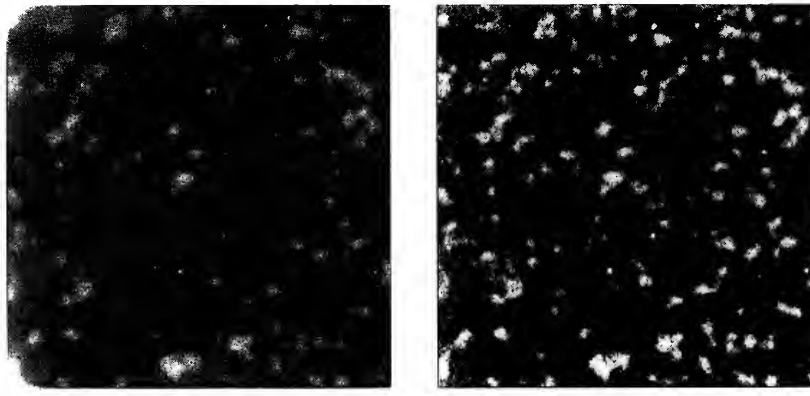


Figure 5. Simulated measured data and restored image.

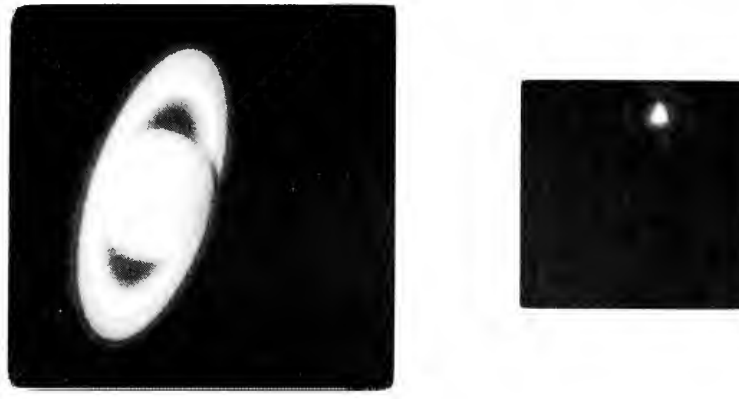


Figure 6. Measured Saturn image and point spread function.

6. Processing of Observed Data

Fig. 6 shows an image of Saturn which was recorded with the Wide Field / Planetary Camera. It is stored in the file SATURN.FITS. The measured point spread function from the data file SATURN_PSF.FITS is also displayed.

Fig. 7 shows the result of restoration. The restored image appears much clearer, with higher contrast than the measured data. The edges, which are considerably blurred in the measured data, can now be located exactly. Therefore the improvement provided by the restoration is obvious.

7. Conclusions

The adaptive least squares restoration method presented in this paper is based on three fundamental properties. First, the regularization operator should be selected such that it removes the correlation contained in the original signal as far as possible. Therefore it should approximate a whitening filter as close as possible. Second, the function which provides the locally adaptive weighting of the degree of regularization must be matched to the regularization operator. Third, a suitable weighting of the noise term depending on the reliability of the measured data is introduced. The weighting function can also be used to incorporate the solution of the boundary problem as well as of the resampling problem directly into the restoration. The resulting adaptive restoration method leads to proper suppression of noise as well as of oscillations without strong blurring of discontinuities. The whitening operators can be calculated approximately from the data at hand and we do not have to select a particular regularization operator in advance.

Acknowledgments. A special acknowledgment is due to Prof. Leonid P. Yaroslavsky from the Russian Academy of Sciences who provided the basic idea to use whitening filters as regularization operators in adaptive least squares restoration.

References

Andrews, H. C., & Hunt, B. R. 1977, *Digital Image Restoration*, Prentice–Hall, Englewood Cliffs

Bundschuh, B. 1992, in *Remote Sensing and Visualization of Information*, Third International Seminar on Digital Image Processing in Medicine, Riga, Latvia, April 1992

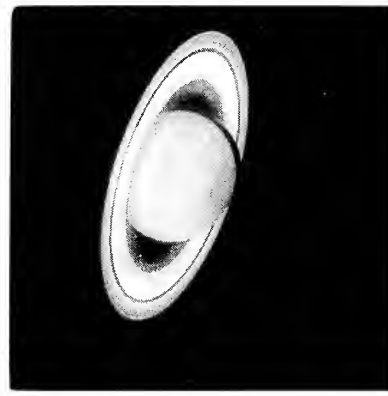


Figure 7. Restored Saturn image.

Bundschuh, B. 1993, in *5th International Conference on Computer Analysis of Images and Patterns* (CAIP'93), Budapest, Hungary, September 1993

Pratt, W. K. 1978, *Digital Image Processing*, Wiley and Sons, New York

Twomey, S. 1963, J. ACM, 10

Yaroslavsky, L. P. 1985, *Digital Picture Processing, an Introduction*, Springer, Berlin

The Restoration of HST Images and Spectra II
Space Telescope Science Institute, 1994
R. J. Hanisch and R. L. White, eds.

Image Restoration and Super-Resolution by Novel Applications of a Neural Network

Gordon Chin

NASA/GSFC, Code 693, Laboratory for Extraterrestrial Physics, Greenbelt, MD 20771

Stephen L. Mahan and William E. Blass

Department of Physics and Astronomy, University of Tennessee, Knoxville, TN 37996-1200

Abstract. A new instrument paradigm is proposed based on the discovery of a method to determine a robust inverse point spread function for a scientific observing instrument modeled as a linear system. As a result of this discovery, it is possible to contemplate an instrumental extension which results in the recovery of a major portion of lost resolution due to the blurring effects of the PSF. Implementation of the instrumental extension and the resulting resolution enhancement is independent of prior knowledge of or access to the observed data. The method is applied to HST images as well as several one dimensional spectral data sets. Results of HST recoveries are compared to Richardson-Lucy recoveries.

1. Introduction

All scientific instruments share a common basis when viewed as a linear system. An observing system I, for example, has an input Θ, and an output D, a response (or point spread function) p, and additive noise n, as represented in Eq. 1:

$$\Theta \longrightarrow \boxed{\text{I(p)}} \longrightarrow D + n \tag{1}$$

We can characterize such a system mathematically using the definition in Eq. 1, which describes a linear instrument model in continuous space. We can define an extended instrument, as in Eq. 2, by adding another linear processing system L, having a response function q that produces an output ϕ given the input D

$$\Theta \longrightarrow \boxed{\text{I(p)}} \overset{D+n}{\longrightarrow} \boxed{\text{L(q)}} \longrightarrow \phi \tag{2}$$

The idea of an extended instrumental system such as Eq. 2 gives us a powerful way of looking at the processes that we wish to describe in this paper.

Using the extended linear system instrument model, our goal is to obtain an enhanced instrument with useful, data independent properties; ϕ may be, for example, a simple estimator of q, in which case our extended instrument model can be used for resolution recovery for both spectra and images. Or the model can be used to transform from one type of point spread function to another such as a positive semi-definite estimator of q in order to make possible an accurate determination of the continuum level from a sinc apodized spectrum as from an Fourier Transform Spectrometer. We will first give examples of these applications below. Then we will briefly explain our methodology in obtaining the linear filter q from a novel application of a neural network, specifically a convolution network.

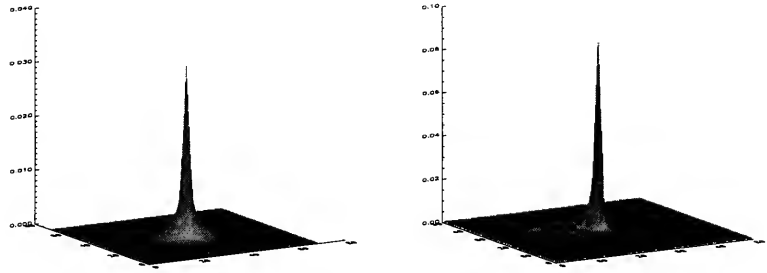


Figure 1. Left: Gaussian based HST-like PSF. Right: Tiny TIM based HST PSF.

2. Image and Spectrum Recovery

2.1. General Procedure

The procedure used to recover both spectra and images or, alternatively, to map from one type of PSF to another, is the same and is outlined below:

1. Generate a model instrumental point spread function or obtain an observed one.
2. Determine the net instrument response desired. That is, determine what $p \otimes q$, the effective PSF of the extended instrument should be.
3. Use a numerical algorithm, the convolution network (discussed in detail below), to search for the solution that provides the desired mapping of the observed data to the desired form. That is, find a q such that $p \otimes q$ is the desired extended instrument PSF. Note that the extended instrument is implemented totally independent of the specific output data of the instrument, i.e., q is data independent.
4. Apply the convolution network's solution to the observations to be mapped, i.e., compute $D \otimes q = D_{\text{enhanced}}$.

2.2. HST-Like and HST Image Reconstruction

We first made preliminary studies investigating recovery of images that have been broadened by the HST's mirror by modeling its PSF as shown in Fig. 1. At left is a simple PSF based on a superposition of two Gaussians, and at right is a typical Tiny TIM PSF for HST's Wide Field/Planetary Camera.

The results of applying our procedure outlined above are given in Fig. 2. At left is the original ground-based image of M82. The middle image shows the M82 data degraded by our model PSF. At right is the recovered estimate of the image using the procedures outlined above. Note that the estimated image does not use any knowledge of the original image. Fig. 3 shows how well the recovery process works in the presence of artificially added noise.

We have also applied our recovery procedure to HST images accessible from the ST ScI image restoration test data sets. Our results for HST images are shown in the following figures. Fig. 4 is an HST image of Jupiter, our enhancement, and a comparison with a Richardson-Lucy enhancement (provided by Bob Hanisch). The PSF used is a Tiny Tim model PSF. Fig. 5 shows our recovery of an HST image of Saturn using an observed 53x53 point HST PSF. Fig. 6 is based on an IRAF M51 image from the ST ScI test data set. The Tiny Tim-PSF processed image used as a starting point in our processing is also from the image restoration test data set. Our recoveries from test images with and without noise are shown with a Richardson-Lucy recovery with noise (also from the ST ScI test data set). We believe that our method compares well with the Richardson-Lucy method, but have not made any quantitative comparisons.

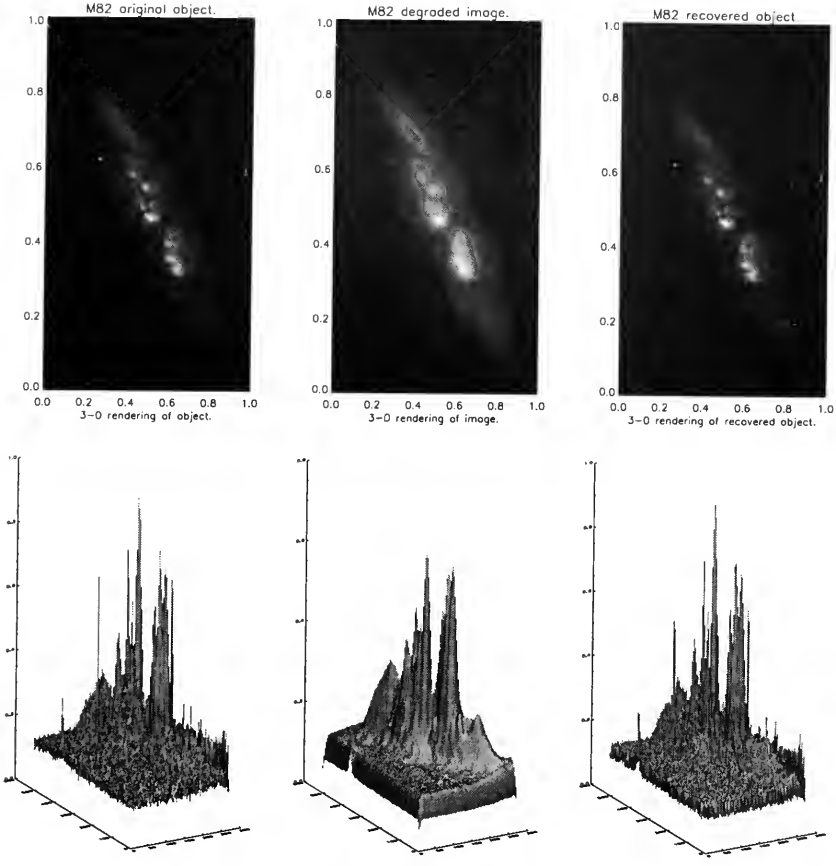


Figure 2. Recovery of M82 based on a noise free blurring of the image by an Hubble-like PSF.

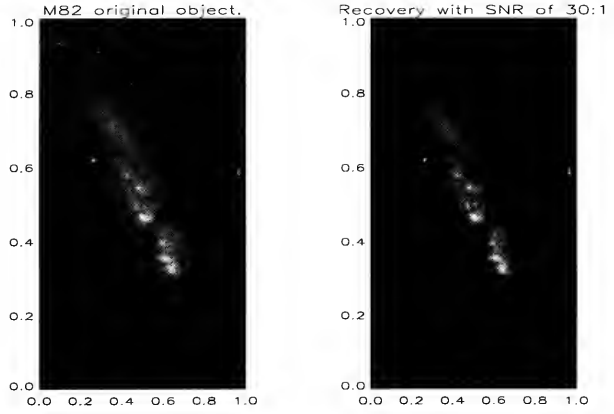


Figure 3. Recovery of M82 based on a random noise blurring of the image by an Hubble-like PSF.

Figure 4. Sample recovery of an HST image of Jupiter compared with a Richardson-Lucy recovery.

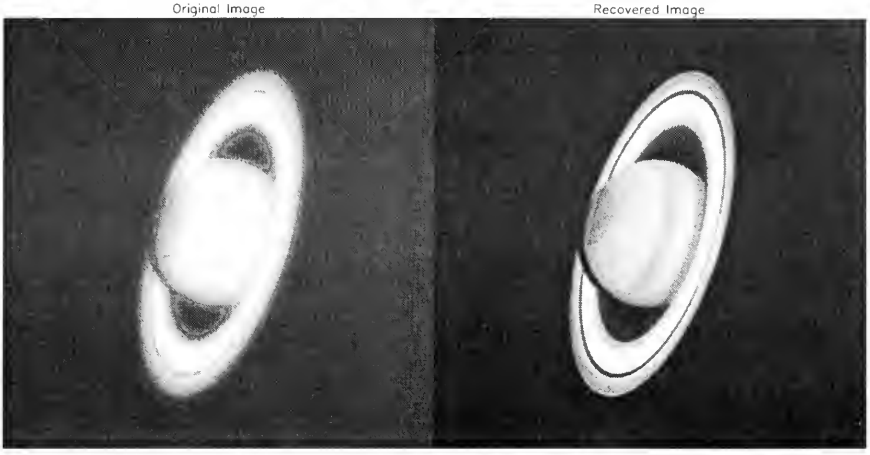


Figure 5. Sample recovery of an HST image of Saturn.

2.3. Recovery of Spectra

Tunable Diode Laser Spectroscopy Our recovery procedure was applied also to spectral data, in this case to a tunable diode laser (TDL) spectrum. The TDL observation of 13 μm propane is shown in Fig. 7. The second row shows an example of a constrained signal space deconvolution of this spectrum, using a nonlinear method (after Blass and Halsey 1981 and Jansson 1984). The other plots are examples of resolution enhancement using the our procedure with the target function being Gaussians having a full width at half maximum (FWHM) of 2 data points, 3 data points, and a unit impulse function, respectively. The TDL PSF is modeled with a Gaussian having a FWHM of 6 data points. The use of targeted PSFs is discussed in more detail below.

Numerical Apodization Fig. 8 shows that it is practical to map a spectrum from a sinc apodization to a positive semi-definite distribution using our numerical procedure. A sinc point spread function (top, Fig. 8) is mapped to a sinc^2 function (middle, Fig. 8), and to a Gaussian function (bottom, Fig. 8) using the linear filters derived from our procedure.

Numerical Apodization Plus Spectral Recovery: FTS Spectrum One benefit of applying our extended instrument model is the ability to cascade a series of filter functions to achieve a desired end result. For example, we can model the response of a Fourier Transform Spectrometer (FTS) spectrum whose PSF is a sinc function and to map its response to a sinc^2 apodized spectrum using our procedure. The numerically apodized spectrum was then deconvolved (restored) using a nonlinear constrained signal space deconvolution algorithm. The "true" spectrum (before observation) is shown at the top of Fig. 9. The second row of Fig. 9 shows an observation from a simulated FTS spectrometer. The third row shows the sinc^2 apodized spectrum. An estimate of the true spectrum derived from the apodized spectrum is shown in the bottom row by applying a further inverse filter.

2.4. Deconvolution by a Novel Neural Network

The instrumental convolution is characterized by Eq. 1, a Fredholm integral equation of the first kind. In principle, one can analytically solve for q by using Fourier transform theory. The instrument model equation is $d = p \otimes \theta + n$ and its Fourier Transform then is $D = P\Theta + N$. Rearranging and

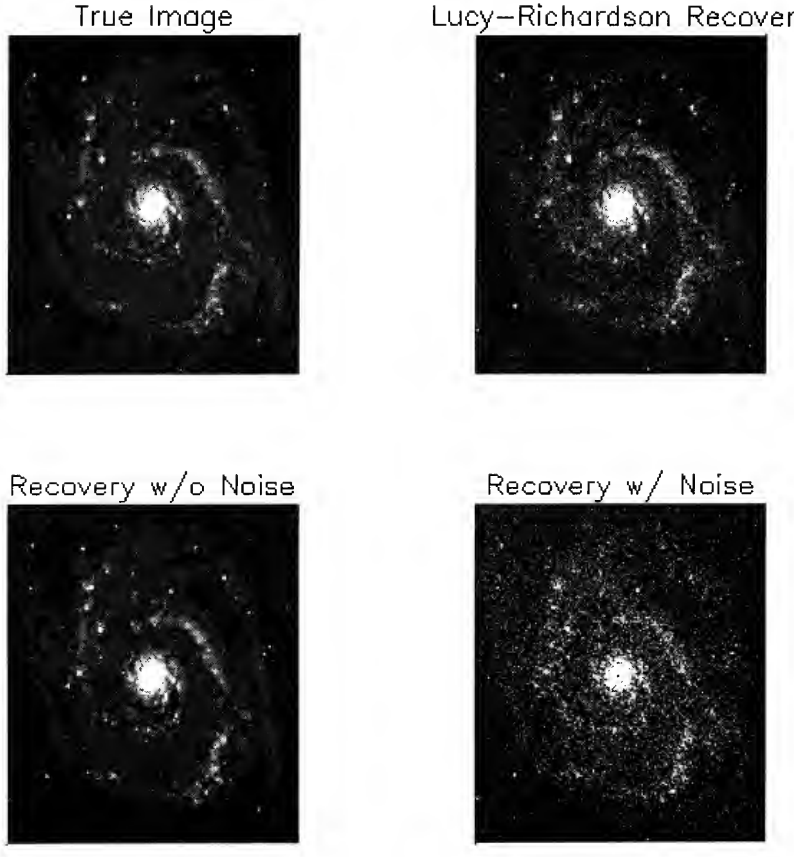


Figure 6. Example recovery of a simulated HST image based on the IRAF test image of M51.

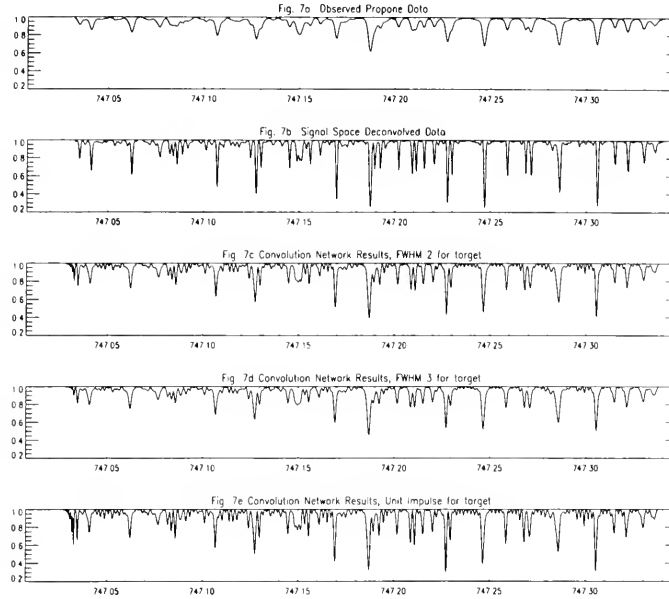


Figure 7. Resolution enhancement of a diode laser spectrum of propane. Top row: original data. Second row: constrained signal space deconvolution. Third row, fourth, and bottom rows: examples of restorations using our procedure with various target functions.

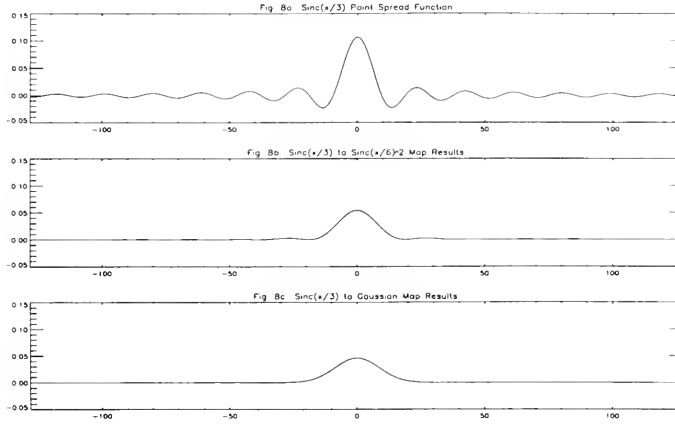


Figure 8. Examples of numerical apodization.

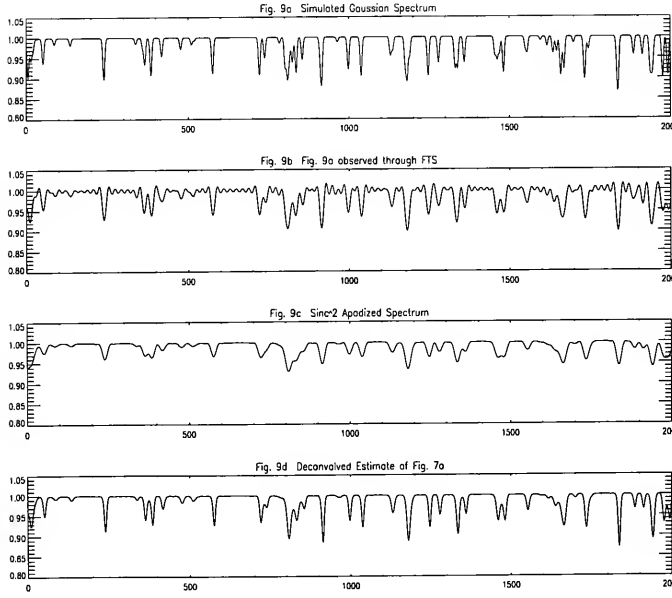


Figure 9. Simulated deconvolution of high resolution spectrum.

using the inverse Fourier Transform, one recovers an estimate of

$$\theta(\vec{x}) = \int_{-\infty}^{\infty} \frac{D(\vec{k}) - N(\vec{k})}{\sqrt{(2\pi)^n} P(\vec{k})} e^{-i\vec{k}\cdot\vec{x}} d^n k \tag{3}$$

This direct inversion has serious problems: $\frac{D(\vec{k})-N(\vec{k})}{P(\vec{k})}$ is not well behaved, in particular it amplifies noise. The convolution network searches for an approximate q that minimizes an error function, $\epsilon^2 = (p \otimes q - T)^2$, where T is the target function. The convolution network is derived from the error-backpropagation neural network model with a slightly more complex, yet more natural connection scheme for imaging systems. Many deconvolution algorithms use a correction term, e, of the form:

$$\epsilon = p \otimes d_{\text{enhanced}} - d_{\text{observed}} \tag{4}$$

The convolution network involves a correction that has the form:

$$\epsilon = p \otimes q - t_{\text{desired}} \tag{5}$$

where p is the point spread function and t_{desired} is the desired result of the mapping (i.e., for a $\text{sinc}(x)$ function mapped to a $\text{sinc}^2(x/2)$ function, t_{desired} is the $\text{sinc}^2(x/2)$ function, p is a $\text{sinc}(x)$ function, and q is the function that minimizes ϵ). After determining q, the application of the convolution network is computationally inexpensive since it involves only a convolution. It is feasible to build an instrument that can be tailored for real-time data enhancement/processing via several filter functions, q_i, as shown in Fig. 6.

$$\Theta \longrightarrow \boxed{\text{I(p)}} \xrightarrow{D+n} \boxed{\text{L(q)}} \begin{cases} \xrightarrow{q_1} \omega & \text{Positive Definite} \\ \xrightarrow{q_i} \tau & \text{Arbitrary Mapping} \\ \xrightarrow{q_n} \phi & \text{Resolution Recovery} \end{cases} \tag{6}$$

3. Conclusions

As a result of our explorations, we can set forth several points in conclusion.

- The convolution network has proven to be a robust method for determination of an effective solution to a class of Fredholm integral equations of the first kind.
- The convolution network was originally developed to search for an efficient, data independent method of image reconstruction of Hubble Space Telescope imagery. This has proven to be highly satisfactory and work is continuing.
- The convolution network has proven successful in resolution enhancement of many imaging systems.
- The convolution network can successfully be applied for the processing of sinc apodized spectra to obtain a positive semi-definite representation of the spectra. This provides a means for obtaining a baseline for the sinc apodized spectra.

References

Blass, W. E., & Halsey, G. W. 1981, *Deconvolution of Spectra*, Academic Press, New York.

Jansson, P. A., ed. 1984, *Deconvolution with Applications in Spectroscopy*, Academic Press, New York.

The Restoration of HST Images and Spectra II
Space Telescope Science Institute, 1994
R. J. Hanisch and R. L. White, eds.

Model Updating in the MEM Algorithm

Nailong Wu[1]

Space Telescope Science Institute, 3700 San Martin Drive, Baltimore, MD 21218

Abstract. The model updating technique used in the Maximum Entropy Method (MEM) task for image restoration in IRAF/STSDAS is described in detail. The properties of the method, including faster convergence in iteration and reduced nonlinearity of the solution in photometry, are discussed and shown by statistics and illustrations.

1. Introduction

The `mem0` package version B in IRAF was released in October 1992. Since then the task `irme0b` for image restoration by MEM in this package has undergone major changes. The latest version of this task, with the name `mem`, will be included in the STSDAS V.1.3.1 patchkit scheduled to be released in late 1993. Currently its executable code `cmem.e` for Sun workstations, together with the parameter and help files, are stored in the NOAO anonymous FTP site (iraf.noao.edu, ~/contrib/).

The model updating (MU) technique has greatly improved the performance of the task, temporarily named `cmem`, of the new version. The iterations converge much faster, and the photometric nonlinearity of the results have been reduced.

In the following we will elaborate on the MU technique and its implementation in `cmem`. Statistics and illustrations will be used to show the aforementioned merits.

2. Model Updating

The entropy expression generally accepted in image restoration is (in 1-D notation)

$$S(\mathbf{h}) = \sum_j h_j - m_j - h_j \log(h_j/m_j) \tag{1}$$

where $\mathbf{h} = \{h_j\}$ represents the image to be restored, and $\mathbf{m} = \{m_j\}$ is the model or prior estimate of the image.

The model $\mathbf{m}$ in the entropy expression plays a special role in MEM image restoration. In most cases, e.g., using the χ^2 data constraint, the MEM solution is unique, depending on the model. More specifically, once the model is given, the solution is unique. However, any change of the model will result in the change of the solution even if the data constraints remain the same. This will be shown specifically in the next section.

Let us consider the following specific formulation in the MEM algorithm. The entropy of an image defined as in Eq. (1) is maximized subject to the data fitting constraint in the image domain

$$E = \chi^2 = \sum_j |\sum_k p_{jk} h_k - d_j|^2 / v_j = E_c$$

[1]Also with Beijing Astronomical Observatory, CAS, Beijing 100080, China.

and the total power constraint

$$F = \sum_j h_j = F_{\text{ob}}$$

where $\{d_j\}$ represents the data, i.e., degraded image, $\{v_j\}$ the noise variance, and (p_{jk}) the PSF (point spread function) combined with the ICF (intrinsic correlation function); the subscript c stands for critical value, and ob for observational value.

The method is implemented by forming the objective function

$$J = S - \frac{1}{2}\alpha E - \beta F$$

and using the Newton method to find ME solutions, i.e., to maximize J for particular values of the Lagrange multipliers α and β (inner iteration). The desired values of E and F are achieved by choosing appropriate α and β (outer iteration). In what follows we will concentrate on α and pay minimum attention to β since the total power constraint is of little interest in the problem.

The iterates are updated according to

$$\mathbf{h}^{(k+1)} = \mathbf{h}^{(k)} + \mu\Delta\mathbf{h}$$

where μ is an optimal step determined by a one-dimensional search. In order to calculate the change $\Delta\mathbf{h} = \{\delta h_j\}$, we need to solve a set of linear equations of very large size:

$$(-\nabla\nabla J)\Delta\mathbf{h} = \nabla J$$

where the gradient and matrix elements are calculated from $\mathbf{h}^{(k)}$, i.e., the current $\mathbf{h}$ and given by, respectively,

$$(\nabla J)_j = -\log(h_j/m_j) - \alpha\sum_n(\sum_k p_{nk}h_k - d_n)/v_n \cdot p_{nj} - \beta \tag{2}$$

and

$$(-\nabla\nabla J)_{jk} = h_j^{-1}\delta_{jk} + \alpha\sum_n p_{nj}p_{nk}/v_n\ . \tag{3}$$

This is a time consuming operation.

In the underlying algorithm for the program, the non-diagonal elements of the PSF matrix are simply ignored in calculating the Hessian $\nabla\nabla J$ under the assumption that the diagonal elements dominate, so that the matrix becomes a diagonal one:

$$(-\nabla\nabla J)_{jk} = h_j^{-1} + \alpha q_j \quad \text{for} \quad j = k; \quad = 0 \quad \text{otherwise}$$

where

$$q_j = \sum_n p_{nj}^2/v_n\ .$$

In this way the solution of the equations becomes a simple operation, specifically,

$$\delta h_j = \frac{1}{h_j^{-1} + \alpha q_j} \cdot (\nabla J)_j\ .$$

The scheme described above is attractive because of its simplicity. However, it has the following drawback: At the beginning of iteration, α is zero and $\mathbf{h} = \mathbf{m}$. Normally the model $\mathbf{m}$ is a flat image. In order to reduce the value of E, α gradually steps up in the course of the iteration as depicted schematically by the lines located in the upper part of Fig. 1. As such, it is fairly easy to find the ME solutions for the particular α's at the early stage of iteration since the ignored non-diagonal elements in Eq. (3) are relatively insignificant. However, as the iteration advances, it becomes more and more difficult to find the ME solutions for the particular α's because the ignored non-diagonal

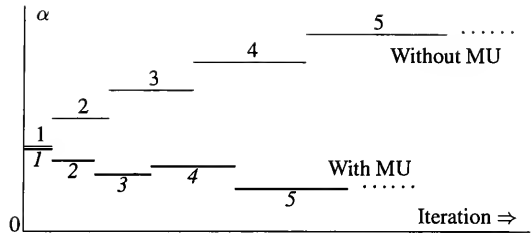


Figure 1. The change of α in iteration. The outer iteration numbers are as annotated. Each outer iteration has a fixed α, and contains a number of inner iterations.

elements (proportional to α) become larger and larger relative to the diagonal elements (depending partly on α). (Of course, this kind of difficulty will be encountered by any algorithm — generally speaking, convergence slows down at the later stage of iteration.) From the above analysis we see that if the values of α can be reduced, then the convergence will be faster. Can we do it? To achieve the desired value of E, we need to increase α to a certain value. α and E are in a one-to-one relationship. At first glance, it seems to be impossible.

The key of the new technique is to update the model in the iteration. At the beginning of the iteration, as usual, $\alpha = 0$ and $\mathbf{h} = \mathbf{m}$. α is increased to some value to start the first outer iteration, and an ME solution is obtained at the end. Then this ME solution is used as the model to start the second outer iteration. Accordingly, α is increased to some value different from zero but not from the value used in the first outer iteration. The later outer iterations proceed in the same manner. In this way α will jump up and down, but not progressively step up in the course of the iterations, as depicted by the thick lines located in the lower part of Fig. 1. As a consequence, the values of α are kept small. Better approximation in the solution of the equations is achieved, and therefore the convergence is speeded up.

In the program `cmem` a parameter called `m_update` is used to control the frequency of model updating. Specifically, the model will be updated every `m_update`'th outer iteration. Setting `m_update` to 1 realizes the above described model updating procedure and achieves the most rapid convergence.

3. Merits

3.1. Faster Convergence

By virtue of the MU technique, in order to achieve some critical value E_c for data fitting, not only the number of outer iterations but also the number of inner iterations in each outer iteration is reduced. Consequently, the total number of iterations is greatly reduced. The computational time is reduced accordingly.

Using synthetic images (Gaussian peaks on a flat background), ordinary images (portraits), and astronomical images (point sources, extended objects), experiments show that compared with `m_update` = 3000 (without model updating), the CPU time is typically reduced for `m_update` = 1 (the most frequent model updating) by factors of 3 to 10.

3.2. Reduced Nonlinearity in Photometry

Let us look into the iteration procedure in some more detail. For simplicity we assume that some "critical value" E_c is achieved after two outer iterations. This value is common to the two cases: without and with model updating. They share the same first outer iteration (see Fig. 1), after which

the ME solution $h_j^{(1)}$, determined from $0 = (\nabla J)_j$ in Eq. (2) with $m_j \equiv 1$, satisfies

$$h_j^{(1)} = \exp(-\alpha_1 \sum_n r_n^{(1)} p_{nj} - \beta_1)$$

where α_1 and β_1 are α and β used in the iteration, and the residual

$$r_n^{(1)} = (\sum_k p_{nk} h_k^{(1)} - d_n)/v_n \ .$$

Without Model Updating The second outer iteration starts off with the initial value $h_j^{(1)}$, and ends up with the solution $h_j^{(2)}$ satisfying

$$h_j^{(2)} = \exp(-\alpha_2 \sum_n (\sum_k p_{nk} h_k^{(2)} - d_n)/v_n \cdot p_{nj} - \beta_2) \ . \tag{4}$$

Note that we still have $m_j \equiv 1$.

With Model Updating The ME solution from the first outer iteration is used as the model, i.e., m_j in Eq. (2) is equal to $h_j^{(1)}$. The initial value in the second outer iteration is still $h_j^{(1)}$. Now the solution $h_j^{(2)\prime}$ satisfies

$$\begin{aligned} h_j^{(2)\prime} &= m_j \exp(-\alpha_2' \sum_n (\sum_k p_{nk} h_k^{(2)\prime} - d_n)/v_n \cdot p_{nj} - \beta_2') \\ &= \exp(-\sum_n (\alpha_1 r_n^{(1)} \cdot p_{nj} + \alpha_2' (\sum_k p_{nk} h_k^{(2)\prime} - d_n)/v_n \cdot p_{nj}) - (\beta_1 + \beta_2')) \ . \end{aligned} \tag{5}$$

Note that $\alpha_1 r_n^{(1)} p_{nj}$ is fixed in the iteration.

The χ^2 values are equal for both cases:

$$E_2 = \sum_j |\sum_k p_{jk} h_k^{(2)} - d_j|^2/v_j = E_2' = \sum_j |\sum_k p_{jk} h_k^{(2)\prime} - d_j|^2/v_j = E_c.$$

The equality of E_2 and E_2' means that the two solutions $\mathbf{h}^{(2)} = \{h_j^{(2)}\}$ and $\mathbf{h}^{(2)\prime} = \{h_j^{(2)\prime}\}$ are located in the same hyper-ellipsoid of $\chi^2 = E_c$. They are two distinct points when $E_c \neq 0$. Assuming that exact data fitting could be achieved, $E_c = 0$, then the hyper-ellipsoid would shrink to a single point representing both $\mathbf{h}^{(2)}$ and $\mathbf{h}^{(2)\prime}$. Even in this case, $\mathbf{h}^{(2)}$ and $\mathbf{h}^{(2)\prime}$ would be two solutions that are not equal because the "solution space" (a set formed by all the possible solutions) without model updating is different from that with model updating. This can be understood by comparing Eq. (4) with Eq. (5) and noticing the term $\alpha_1 r_n^{(1)} p_{nj}$ in the latter's exponent. The situation here is similar to that in which we have two different coordinate systems in a space; the same point in the space has different coordinate values in different systems.

Now we revert to the case where $E_c \neq 0$. Based on the above discussion, it is easy to see that $\mathbf{h}^{(2)}$ and $\mathbf{h}^{(2)\prime}$ are different.

As a digression we would like to point out that the consistency of MEM solution ensures that the final solution depends only on the given data but not on the order in which these data are used (Wu 1991). However, the argument given there does not apply to the current case concerning image restoration.

In the case where more than two outer iterations are performed, it is clear from the above discussion that the final solutions for a certain value, say N_{data}, of E_c are different with and without model updating. But we are not able to quantify theoretically the difference between them.

We have observed in experiments that the solutions with and without model updating are similar in morphology but noticeably different in photometry. The former is less biased or has reduced nonlinearity in photometry. Why is that? It is generally accepted that better models will result in solutions better in photometry. The flat model used throughout the iteration is by no means the best one. In the iteration, as α is increasing and data fitting becomes tighter and tighter, the large-scale structure of image is built up first and the small-scale structure comes later. The model is improving by updating it in a reasonable way. This explains the improvement of the solution in photometry. Model updating does make difference to the solution. Fortunately, it becomes better, not worse.

4. Tests Using a Simulated Image

A simulated globular star cluster for the WFC of HST before the servicing mission has been used as a test case. The image was retrieved from a set of test data created for image restoration work in the ST ScI anonymous FTP area.

The image contains 470 stars with approximately the maximum and minimum pixel values (2500 and 10 DN, respectively), covering a range of magnitude of 6. A standard (not subsampled) PSF is used. The image size is 256×256 pixels.

4.1. Convergence Speed

Convergence speed is expressed in terms of the number of FFTs needed to obtain the final MEM solution. The number listed in Table 1 is decomposed into two factors as (the total number of iterations) × (the number of FFTs per iteration). `cmem` is the latest version of my MEM program. Values of `m_update` = 1 and 3000 correspond to the most frequent model updating and no model updating, respectively. Comparisons have been made with the `lucy` task (Richardson-Lucy method) in the IRAF/STSDAS package and `MEM/MemSys5`, a commercial software package.

Table 1. The number of FFTs for convergence.

`cmem`		`lucy`	`MEM/MemSys5`	
m_update=1	=3000		historic	classic
108x4	1881x4	302x4	375x2	not converged after 10016x2

From the table we can see that the convergence is speeded up dramatically because of model updating. With this technique, `cmem` converges fastest.

4.2. Photometry

In Fig. 2 "Difference" gives the stars' magnitudes estimated from the true image ("True magnitude") minus those from the restored image. The radius for aperture photometry is 3 pixels. Without errors in photometry, the points would lie on a horizontal line as additionally drawn in each plot. Without bias, the points would symmetrically distribute below and above the line.

In Fig. 2 we see that the restored image C by `lucy` exhibits quite good linearity in photometry. `cmem` with model updating (A) is comparable to `lucy` except near the fainter end of the magnitude range where `cmem` shows some systematic errors. The extreme differences between A and C are +0.20 and −0.45 mag.

The result in the case of no model updating (Fig. 2B) is noticeably worse. The result from `MEM/MemSys5 historic` (D) is worse still. The rather large amount of offset in D is due partly to the lack of total power constraint in the program. (The total power errors correspond to offsets of 0.011, 0.017, 0.035 and 0.12 in A, B, C, and D, respectively.) The extreme differences are

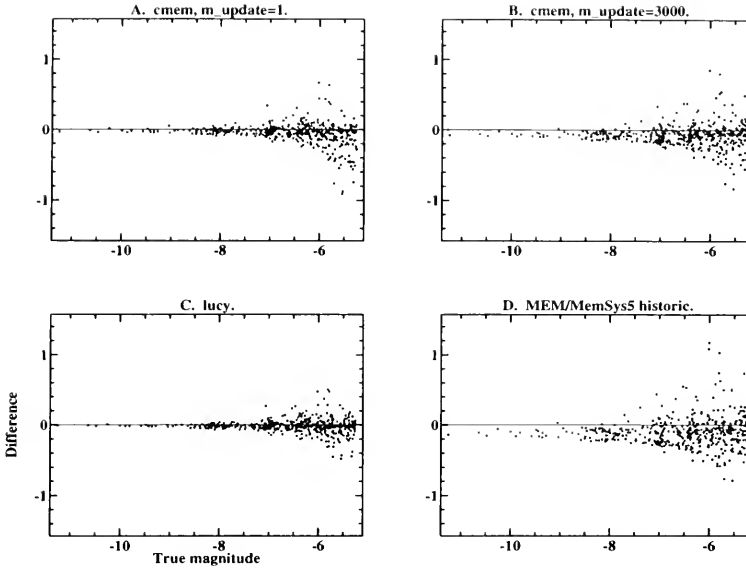


Figure 2. Difference between the true and restored images in photometry.

$+0.60/-0.35$ mag between B and C, and $+0.90/-0.45$ mag between D and C. Intermediate values of differences are scattered to a much greater degree compared with those between A and C.

5. Conclusion

It is a common view that MEM is slow in convergence and highly nonlinear in photometry. Now this view should be changed. With the model updating technique, we are able to make things with MEM better: the iteration converges fairly fast, and it is comparable to the Richardson-Lucy method in photometry.

References

Wu, N. 1991, "Bayesian Inference, MEM, and Consistency," presented at the 4th International Meeting on Bayesian Statistics, Valencia, Spain

The Restoration of HST Images and Spectra II
Space Telescope Science Institute, 1994
R. J. Hanisch and R. L. White, eds.

Interactive Deconvolution with Error Analysis

K. Bouyoucef, D. Fraix-Burnet, and S. Roques

Laboratoire d'Astrophysique, CNRS URA 285/GdR 134, Université Toulouse III, 14 Avenue Edouard Belin, 31400 Toulouse, France

Abstract. In the general framework of regularization of inverse problems, we define the main contours of the reconstruction algorithm IDEA (*Interactive Deconvolution with Error Analysis*) that is developed in our group, giving only some methodological principles. The deconvolution problem is stated in terms of weighted spectral interpolation: the amount and the nature of the interpolation to be performed is related — in a quantitative manner — to the choice of a synthetic coverage linked to the target resolution. We illustrate this deterministic viewpoint on the SN1987A image of HST and compare the result with two Bayesian approaches: the Richardson-Lucy Method (RLM) and the Maximum Entropy Method (MEM) in terms of gain in resolution, error propagation, speed of convergence, and some relevant astrophysical criteria.

1. Introduction

The underlying principle used to restore an image is known generally as a reconstruction or inverse problem. This usually involves finding a method of inverting certain classes $y = Ax$ of integral equations. In the case of a deconvolution problem, these equations are of the type $\phi_i(z) = \iint h(z - z')\phi_o(z')\, dz'$ where h is the kernel associated with the operator A. This is basically an ill-posed problem (Tikhonov and Arsenin 1977) because it does not fulfill the three Hadamard conditions of existence, uniqueness, and stability of the solution. The last condition of stability causes the main problems because if A^{-1} is not continuous, the solutions x are unstable with respect to slight variations of initial data y. This means in practice that a slight error in the data may lead to a very large error in the solution.

When the problem is discretized, its transposition into finite dimension certainly eliminates the question of continuity of A^{-1}, but the difficulties of the underlying infinite problem result again in numerical instabilities due to the ill-conditioned nature of the matrix of the system. As a result, the noise is enhanced and the solutions may be singular. Let δy be some variation of the data y. Denoting by δx the corresponding variation of x, one has $\|\delta x\|/\|x\| \leq C\, \|\delta y\|/\|y\|$, where $C = \mu_{max}/\mu_{min}$, μ^2_{max} and μ^2_{min} being the largest and the smallest eigenvalues of the imaging operator. The condition number C provides then a measure of the difficulty about the resolution of equation $y = Ax$, and hence a measure of the robustness of the problem under consideration.

To circumvent the ill-conditionedness nature of this problem, one is led to postulate that the properties of the solution are not entirely contained in the equation to be solved. One is then led to introduce a priori information on the solution which takes a regularizing part in the deconvolution process. This information is the basis for classes of linear or non-linear methods which are known as deterministic or Bayesian ones.

2. IDEA: Methodological Principles

The reader will find extensive studies in Lannes et al. (1987a-b) where the method summarized here is presented in a closed form, and some astrophysical applications in Fraix-Burnet et al. (1989), Roques et al. (1993) and Bouyoucef et al. (1994). To a first approximation, the experimental data

ϕ_i are related to the original object ϕ_o, the intensity of the source at some high level of resolution, by an experimental transfer relation:

$$\phi_i(z) = (\phi_o * h)(z) + \text{error term}$$

where h is the Point Spread Function. The error term includes random or systematic errors (e.g., errors on the determination of the PSF, linearity assumption of the imaging relation, image sampling) and signal-uncorrelated random noise (telescope, detectors, atmosphere, guiding, etc.).

The accuracy of the approximation considered when writing the convolution equation is controlled in Fourier space by a pointwise image-error bound $\sigma_i(u)$ such that $|\widehat{\phi}_i(u) - h(u)\widehat{\phi}_o(u)| \leq \sigma_i(u)$. Moreover, ϕ_o can be regarded as a member of a certain family of objects. For each spatial frequency u, it is possible to exhibit a suitable upper bound of the Fourier transform of ϕ_o: $|\widehat{\phi}_o| \leq \sigma_0(u)$.

Due to the noise-type and systematic errors characterized by the pointwise Signal-to-Noise Ratio (SNR) in the frequency space:

$$\text{SNR}(u) = \frac{|\widehat{\phi}_i(u)|}{\sigma_i(u)}$$

it is preferable to give up trying to determine ϕ_o at its highest level of resolution. One then defines the "object to be reconstructed" ϕ_s as a smoothed version of ϕ_o by a relation of the form $\widehat{\phi}_s(u) = \widehat{s}(u)\widehat{\phi}_o(u)$, where $\widehat{s}(u)$ is some synthetic transfer function small in the mean square sense outside some extent H_r. This domain is the frequency coverage to be synthesized and regularizes the effective support H of the transfer function $\widehat{h}$. The choice of its diameter is of fundamental importance because it is closely related to the resolution limit of the reconstruction process. Intuitively, the greater is H_r with respect to H, the greater is the gain in resolution, but at the same time, the less stable is the deconvolution process. The problem is then to define the best compromise between the resolution to reach and the stability of the solution. The support of ϕ_s is contained in some finite region V a priori known, which size and shape, determined in an interactive manner, will prove to play, together with H_r, an essential part.

This filter $\widehat{s}$ fulfills three conditions:

1. As the support of $\widehat{h}$ is essentially in H, the support of $\widehat{s}$ is essentially in H_r (say for instance 95% of its energy in H_r). One imposes a condition of the form:

$$\frac{1}{\|\widehat{s}(u)\|^2} \int_{H_r} |\widehat{s}(u)|^2 du = \xi^2$$

2. The support of s must be as small as possible (with respect to the choices of H_r and ξ) to ensure the best possible resolution for ϕ_s.

3. $\widehat{s}(0) = 1$ what secures flux conservation.

A first approximation $\widehat{\phi}_t$ of $\widehat{\phi}_s$ of the object to be reconstructed is:

$$\widehat{\phi}_t(u) = \begin{cases} \dfrac{\widehat{s}(u)}{h(u)} & \text{if SNR}(u) \geq \alpha_t \\ 0 & \text{otherwise} \end{cases}$$

where α_t is some threshold value of the order of unity.

In the frequency domain where the SNR is considered as good (SNR$(u) \geq \alpha_t$), $\widehat{\phi}_t$ is more or less reliable. One then assigns to the data a weight depending on the values of SNR. Let α_t' be the threshold value beyond which the whole information contained in $\widehat{\phi}_t$ is considered as entirely

reliable. Let $g(u)$ be the "weight function" characterizing the weight attached to this information. The function $g(u)$ is defined as follows:

$$g(u) = \begin{cases} 1 & \text{outside } H_r \\ 1 & \text{in } H_r \text{ if } \mathrm{SNR}(u) > \alpha_t' \\ & (\textit{information entirely reliable}) \\ \dfrac{\mathrm{SNR}(u) - \alpha_t}{\alpha_t' - \alpha_t} & \text{in } H_r \text{ if } \alpha_t \leq \mathrm{SNR}(u) < \alpha_t' \\ & (\textit{information partially reliable}) \\ 0 & \text{in } H_r \text{ if } \mathrm{SNR}(u) < \alpha_t \\ & (\textit{information non reliable}) \end{cases}$$

One supposes that $\mathrm{SNR}(u) < \alpha_t$ outside H_r. So defined, g characterizes the way of grounding the reconstruction on the information. Our deterministic procedure based on a least squares minimization defines the reconstructed object as the function which minimizes the functional:

$$q(\phi) = \|g(\widehat{\phi}_t - \widehat{\phi})\|^2.$$

The stability of the reconstruction problem is conditioned by the smallest eigenvalue μ_{min}^2 of the imaging operator. It can be analytically estimated by examining some physical parameters: v, characteristic functions of V, and $w = 1 - g^2$ related to the choice of H_r and to SNR. Indeed, this eigenvalue is a function of the "interpolation parameter" :

$$\eta = \tau\nu = \left(\int_V v(z)dz\right)^{1/2} \left(\int_{H_r} (w(u)du\right)^{1/2}$$

characterizing the amount the interpolation to be performed both is real and in Fourier spaces. One has the following relation:

$$\mu(\eta) = 1 - \eta^2 \sum_{m=0}^{\infty} (-1)^m k_m \, \eta^{2m}$$

where the k_m's depend on v and w and are of the kind "moment of inertia" relatively to V and H_r. This equation provides useful approximation of the minimum eigenvalue of the imaging operator occuring in the expression of an upper limit Θ of the reconstruction error:

$$\Theta = \frac{\|\Delta\phi\|}{\|\phi\|} \leq \frac{1}{\mu_{min}} \frac{\rho}{\|\phi\|}.$$

When this analysis is implemented before the reconstruction, one has only an estimation μ_e of μ_{min} and the estimation ϕ_t of ϕ. The upper bound of the quadratic reconstruction error is given by:

$$\Theta_e = \frac{1}{\mu_e} \frac{\varepsilon}{\|\phi_t\|}.$$

In these two expressions, ρ and ε can be regarded as errors resulting from the noise analysis and from systematic errors related to the choice of H_r.

Note that these majorants correspond to the case where the overall error would be concentrated in the eigenspace associated with the minimum eigenvalue, what is not the case in practice (the error is generally distributed amongst all the eigenspaces).

Once the stability conditions are fulfilled, the solution can then be obtained by several well known minimization iterative methods. We apply the conjugate gradients method (Hestenes et al.,

1952) because a particular implementation allows to compute the spectral decomposition of the imaging operator (then a more precise error can be calculated). The convergence is superlinear, and the least squares solution ϕ is reached in a number of iterations of the order of the number of degrees of freedom of the reconstruction process.

3. Example of Deconvolution: The HST Image of SN1987A

For illustration of IDEA, we present the deconvolution of an HST image of SN1987A because of its relatively simple structure (a bright core together with a well delimited extended object, the ring). We compare the result of the deconvolution with IDEA with two Bayesian methods: the Richardson-Lucy method (RLM) and the Maximum Entropy method (MEM). For these two methods, we use softwares available in IRAF at STScI.

This 128x128 pixel image is an 822s FOC exposure (f/96 mode, pixel size 0.″022) through the F501N (OIII) filter, obtained on 1990 August 24. It is described in Jakobsen et al. (1991) and a RLM deconvolution of this image is presented in Panagia et al. (1991). We use a PSF obtained on a star on 1990 August 28 in the same telescope conditions as for SN1987A.

We show in Fig. 1 the original image, and the solutions obtained with IDEA, RLM and MEM. The RLM deconvolution is obtained after 50 iterations. For the MEM deconvolution, the result is obtained after 30 iterations with an uniform noise of 2.56, a Poisson noise coefficient of 0.35, a quadratic noise coefficient of 0.03, and after final smoothing by 2 pixels. For the IDEA deconvolution, we choose a gain in resolution of 1.9 yielding an upper limit of 17% to the quadratic error. As noticed above, the least squares solution is reached in IDEA in a number of iterations corresponding to the number of degrees of freedom of the reconstruction process (8 in this case).

The supernova star is anisotropic in all images: the RLM and MEM solutions show several spikes which are residuals of the PSF. But on the IDEA solution this anisotropy is essentially an elongation at P.A.$\simeq$ 45° that could be physically related to the elliptical projected shape of an envelope around the supernova (see below).

At first glance, the RLM solution seems to reach the highest resolution. A comparison of the profiles of the supernova between the different images is shown in Fig. 2. These profiles are obtained by plotting the intensity of each pixel as a function of its distance to the center. A polynomial fit of the profile of the PSF is presented on Fig. 2a, together with the profile of the faint star visible SE of the supernova. The two bright stars close to the supernova cannot be used as PSF because of distortions introduced by saturation effects in the camera, but they have nevertheless radial profiles indistinguishable from that of the PSF. The first ring caused by the spherical aberration is clearly seen. The inner part of the PSF has FWHM= 0.″06 corresponding to the unaberrated telescope. On the five other plots, the profile of the supernova on the raw and deconvolved images is compared to the profile of the PSF.

Clearly, *the supernova on the raw image does not have the profile of the PSF*, meaning that it is *non-stellar* (FWHM= 0.″13). This has already been found by Jakobsen et al. (1991), and its width indeed corresponds to the expected size of the envelope that was ejected in the 1987 explosion. Hence, the deconvolution should preserve the profile of the supernova. This is exactly the case with IDEA (FWHM= 0.″12, Fig. 2b) and MEM (FWHM= 0.″11, Fig. 2c). In addition, even if we try to increase the gain in resolution with IDEA (at the expense of the quadratic error), the profile remains larger than the PSF (FWHM= 0.″10). By contrast the supernova on the RLM solution (Fig. 2d) has exactly the profile of the PSF, revealing over-resolution. Because one could argue that the number of iterations on the RLM solution is too high, we present results with 10 and 30 iterations (Fig. 2e and f): the supernova is already over-resolved at 10 iterations. So whatever the stopping criterion that could be implemented in RLM, the solution would be over-resolved.

The total width of the bright structure of the ring W of the supernova (Fig. 1) is 0.″09 with IDEA and 0.″08 with RLM (50 iterations). This shows that this structure is slightly resolved. This confirms that the "large" profile of the supernova seen on the IDEA solution is really due to an envelope around the star and not to a lack of resolution. We also note that the smoother shape of

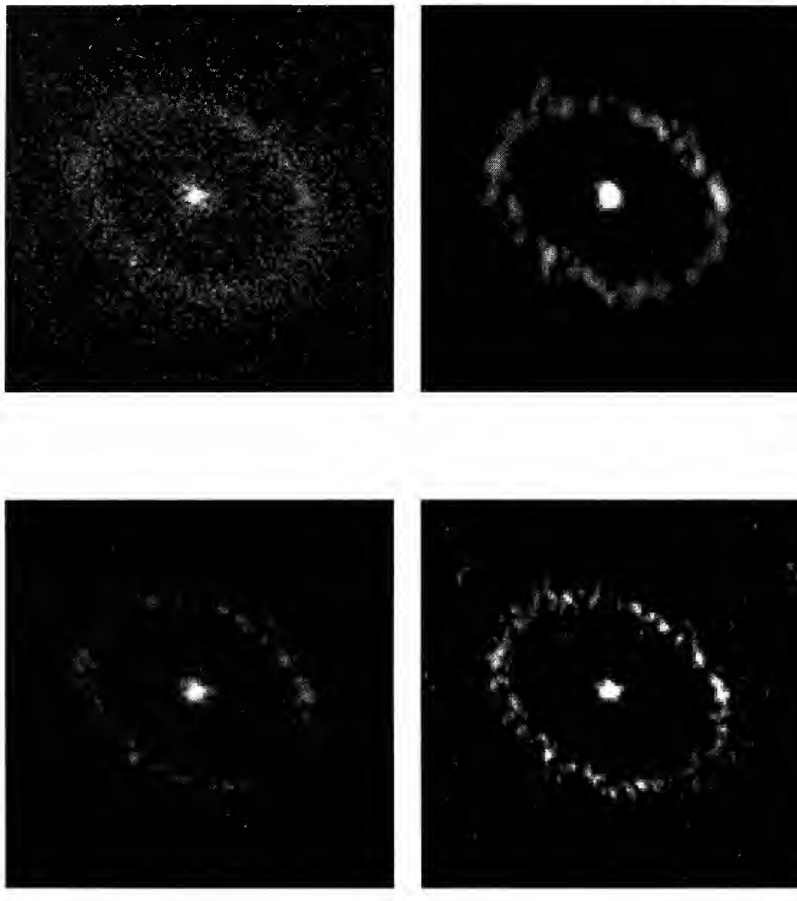


Figure 1. Top left: raw image of SN1987A. Top right: IDEA deconvolution. Lower left: MEM deconvolution. Lower right: RLM deconvolution. The lut is linear and the IDEA and RLM images are thresholded at the same levels.

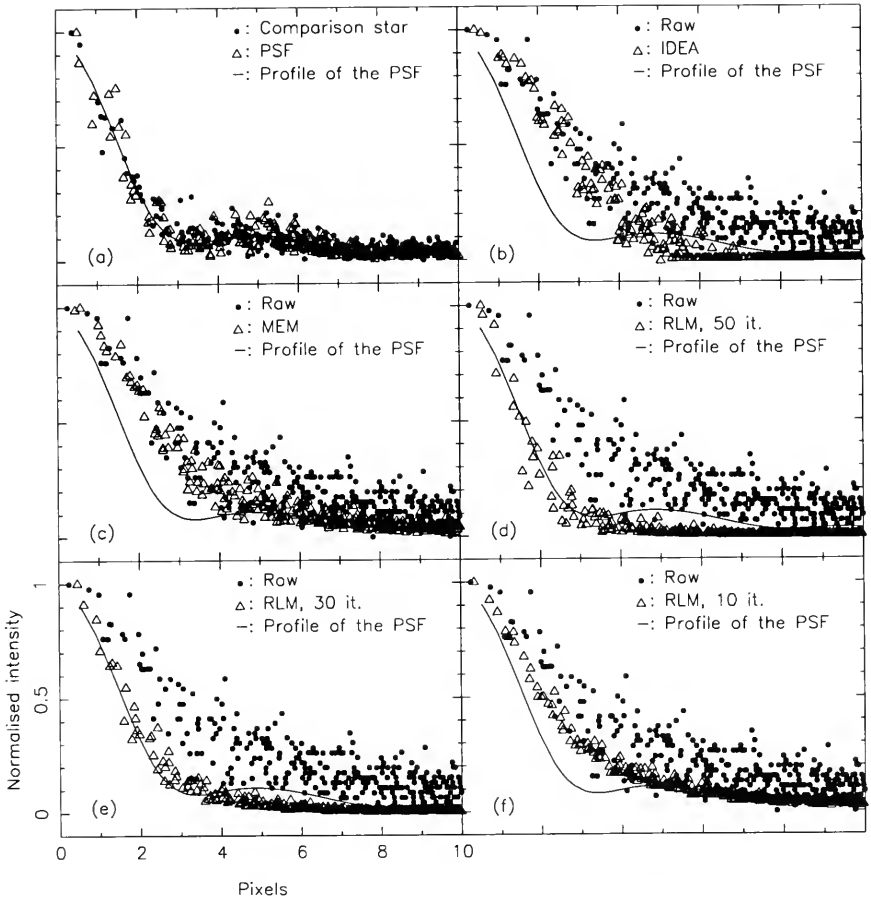


Figure 2. a: radial profiles of the PSF and the star SE of SN1987A. The curve is a polynomial fit to the PSF. On the other plots: profile of the PSF and profile of the supernova on the raw image compared with profiles of the supernova after the different deconvolutions (b: IDEA, c: MEM, d: RLM 50 iterations, e: RLM 30 iterations, f: RLM 10 iterations).

the supernova after deconvolution with IDEA is more reminiscent of an envelope around the star than on the other deconvolutions.

On the IDEA solution, the ring is more filamentary and less noisy than on the other deconvolutions. It is rather blobby on the RLM solution suggesting some over-resolution as shown in the analysis above. It is very noisy on the MEM solution probably because MEM fails to gather the information from the wings of the PSF which are of very low signal-to-noise ratio.

The fluxes of the supernova and the ring measured within the supports used in IDEA are listed in Table 1. For the raw image, we use the values estimated by Jakobsen et al. (1991) for comparison.

Table 1. Absolute fluxes in 10^{-14} erg s^{-1} cm^{-2}. The values for the raw image are taken from Jakobsen et al. (1991).

Image	Supernova	Ring
Raw	5.9	30
IDEA	5.2	22
RLM	4.3	18
MEM	1.9	10

IDEA is expected to rigorously conserve the photometry and this is verified in this case, taking into account the uncertainty on the photometry with such a PSF. On another hand, MEM and RLM clearly loose a significant fraction of the total energy from the wings of the PSF, although the peak value of the supernova on the RLM deconvolution is much higher than on the other deconvolutions because of over-resolution.

4. Conclusion

The approach of IDEA allows an interactive choice for the compromise between gain in resolution and stability of the solution. In addition, the photometry is theoretically conserved. These two points have been illustrated on the difficult example of HST images and the error control allowed in IDEA clearly provides confidence in the quantitative analysis of the deconvolved images.

The analysis presented in this communication should lead to a better understanding of the guiding ideas of the whole regularized restoration methods that can be applied to HST images, and shows what can and cannot be done in this field whatever the particular technique implemented.

A more extensive presentation of IDEA and a study of HST images of the optical extragalactic jets in M87 and 3C66B is presented in Bouyoucef et al. (1994). A comparison between our deterministic approach and a regularized RLM and an algorithm of maximum entropy on the mean are in progress. Simulations are also used in order to determine the photometry conservation for low signal-to-noise ratios.

We would like to emphasize that IDEA has been developed independently from the spherical aberration of the HST, so that it can be used with any image, as for instance with ground-based astronomical images (see Fraix-Burnet et al. 1989).

Current developments of IDEA include the introduction of multiresolution analysis in the determination of the signal-to-noise ratio.

Acknowledgments. DFB acknowledges financial support from the Institut National des Sciences de l'Univers (INSU) for this workshop.

References

Bouyoucef, K., Fraix-Burnet, D., Lannes, A., & Roques, S. 1994, to be submitted to A&A

Fraix-Burnet, D., Nieto, J. L., & Roques, S. 1989, A&A, 217, 387

Goult, R. J., Hoskins, R. S., Milner, J. A., & Pratt, M. J. 1984, in *Computational Methods in Linear Algebra*, Stanley Thornes Publishers

Hestenes, M. & Stiefel, E. 1952, J. Res. Nat. Bur. Stand., 49, 409

Jakobsen, P. et al. 1991, ApJ, 369, L63

Lannes, A., Casanove, M. J., & Roques, S. 1987b, J. Mod. Opt., 34(3), 321

Lannes, A., Roques, S, & Casanove, M. J. 1987a, J. Mod. Opt., 34(2), 161

Panagia, N., Gilmozzi, R., Macchetto, F., Adorf, H.-M., & Kirshner, R. P. 1991, ApJ, 380, L23

Roques, S., Bouyoucef, K., Touzillier, L., & Vigneau, J. 1993, in *Progess in Wavelet Analysis and Applications*, Meyer & Roques, eds., Editions Frontières, Gif sur Yvette, France, 563

Tikhonov, A. N. & Arsenin, V. Y., 1977, in *Solutions of Ill-Posed Problems*, Winston, Washington, DC

The Restoration of HST Images and Spectra II
Space Telescope Science Institute, 1994
R. J. Hanisch and R. L. White, eds.

Towards HST Restoration with a Space-Variant PSF, Cosmic Rays and Other Missing Data

Hans-Martin Adorf

Space Telescope – European Coordinating Facility, European Southern Observatory, Karl-Schwarzschild-Strasse 2, D-85748 Garching bei München, Germany

Abstract. The open problem of restoring full Wide Field and Planetary Camera (WF/PC) image frames is considered. Several novel algorithms or modifications to existing algorithms are described useful for restoring undersampled (multi-) frames degraded by a space-variant point spread function (SV-PSF) and by irregular faults. The algorithms comprise a faithful rotation operator for sufficiently sampled images, an alternative method for treating regular and/or irregular missing data, and an algorithm for SV-PSF restoration combining the classical sectioned restoration method of Trussel & Hunt with the generalized "co-addition" restoration method of Lucy & Hook. With these algorithms a comprehensive full-frame WF/PC restoration now appears to be feasible.

1. Introduction

The restoration of images obtained by the Hubble Space Telescope's first Wide Field and Planetary Camera (WF/PC-1), is still an unsolved problem. This is mainly due to the spatial variability of its point spread function (PSF), in conjunction with a variety of other factors. The purpose of this contribution is to investigate the algorithmic requirements for a comprehensive space-variant PSF (SV-PSF) restoration procedure, and to show that, based on the developments of the past three years, a full-frame WF/PC restoration has become feasible.

A comprehensive WF/PC restoration process (Fig. 1) has to address the following problems:

1. The point-spread function is spatially variable and relatively large. Restoration with a SV-PSF is known to be notoriously difficult, and generally demanding in terms of CPU-power. The PSF has to be computed from a parametric model, or somehow estimated from the data frame(s). An efficient algorithm must be used in view of the large data and PSF frame formats.

2. The data frames are undersampled. Observations must therefore be carefully designed in terms of splits, sub-pixel shifts and rotations. Often the registration parameters must be estimated post-facto, and different overlapping frames have to be combined.

3. Long exposures suffer from numerous cosmic ray (CR) hits. These have to be detected — a daunting task in the presence of undersampling — and the missing data have to be treated somehow.

4. There are other problems, such as detector contaminations dubbed "measles" and flat-field uncertainties.

In the following several algorithms addressing problems (1) to (3) above are presented. Firstly, the problem of restoring undersampled (multi-) frames is discussed. It is argued that the cosmic ray problem and the undersampling problem can be treated on the same footing. The proposed algorithm for irregular missing data with masks largely follows treatments by Adorf et al. (1993), White (1993), and recently by Freudling (1993), except for the up- and down-sampling aspects. Secondly, an algorithm for SV-PSF restoration is presented which combines the generalized "co-addition"

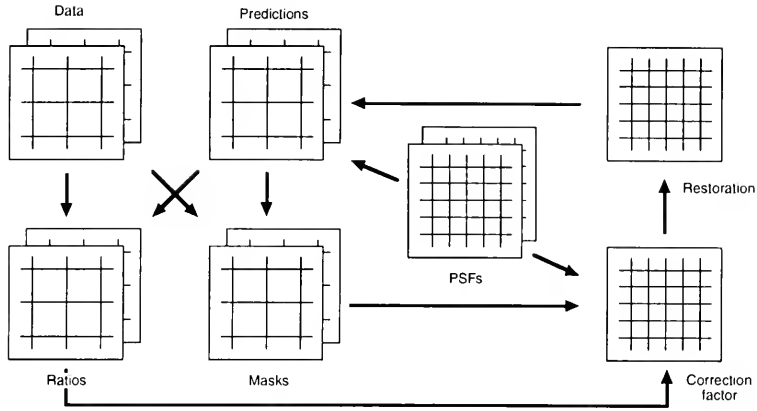


Figure 1. Flow-chart of a full-fledged WF/PC restoration process. The data (upper left) consists of several small, coarsely sampled patches ("sections") taken from one or more WF/PC exposures. Associated with each patch is a separate approximately space-invariant (SI) point-spread function (center). There is only one large, finely sampled joint restoration (right) from which several coarsely sampled predictions are computed with the help of the SI-PSFs. A prediction based algorithm for "on-the-fly" fault detection of cosmic ray hits and other defects is integrated into the restoration scheme.

algorithm of Lucy & Hook (Lucy 1991a; Lucy & Hook 1992, 1993) with the sectioned restoration method originally devised by Trussel & Hunt (1978a, 1978b). When designing these algorithms the following principles have been adhered to (in order of decreasing priority): preserve the integrity of the observational data; leave the Richardson-Lucy (RL) restoration method (Richardson 1972; Lucy 1974) intact as much as possible; strive for theoretically sound, yet efficient, algorithms rendering full-frame WF/PC restoration practical.

2. Treatment of Undersampled Image Frames

In principle, both WF/PC cameras undersample their PSF. In the case of WF/PC-1 the undersampling problem is usually obscured by the combined effects of spherical aberration and noise, rendering many WF/PC frames sufficiently sampled in effect. For WF/PC-2 it is expected that, in view of its sharp PSF, non-ideal pixel response function, and reduced detector noise (Burrows 1994), the undersampling problem will be much more prominent.

The anticipated problems associated with undersampled image frames have been addressed in a number of articles (Adorf 1989a, 1989b, 1990). Fosbury (1993) discusses several image sampling strategies for the proposed future Advanced Camera for HST, and lists recent references concerning undersampled multi-frames. The major difficulty with any undersampled data frame is that it does not represent a continuous image. Thus generally, it cannot be shifted, stretched or rotated.[1]

In order to overcome the undersampling problem, in the context of HST, it is generally recommended to obtain two or more data frames of the same field (i.e., multi-frames) with different

[1] Even when the data are sufficiently sampled, the conventional rotation algorithms implemented in the standard astronomical image processing packages are generally inadequate, since they are not faithful to the data. An algorithm is called "faithful" here if it preserves the information content (in particular, the high spatial frequency spectrum) of the data. Every faithful operator has an inverse operator.

(e.g., shifted or rotated) sampling patterns, and to subsequently combine them. The generic procedure for reconstructing an image from a set of undersampled multi-frames consists of carrying out a joint restoration following the standard Lucy-Hook recipe, which internally requires algorithms for shifting, stretching, or rotating the model (= best restoration) in order to map it onto the sampling patterns of the different observations. It is important that these mappings are carried out faithfully. Of these operators the rotation operator is the most difficult one to realize in a faithful manner.

2.1. A Novel Faithful Rotation Algorithm

The rotation algorithm described here strictly preserves the spatial Fourier-spectrum and is based on the fact that a 2×2 rotation matrix $R(\theta)$, which rotates an image by an angle θ, can be decomposed in several ways, including one consisting of three consecutive 2×2 shear matrices

$$R(\theta) = S_x(-\theta/2)\, S_y(\arctan(\sin(\theta)))\, S_x(-\theta/2) \tag{1}$$

Here $S_x(\theta)$ and $S_y(\theta)$ designate a shear by an angle θ in the x- and y-direction, respectively.

Since a sufficiently sampled image can be faithfully sheared using efficient 1D sinc interpolations (based on fast Fourier tranforms, FFTs), the whole rotation operation can also be carried out faithfully. The only problem to watch out for are potential Nyquist frequency components, which can only appear in even-sized frames. Since the three shears, out of which the rotation is composed, are carried out independently on individual rows (or columns), the faithful rotation algorithm is particularly suited to a vector-parallel computer architecture.

The rotation algorithm can be incorporated into the generalized Lucy-Hook "co-addition" scheme, in order to faithfully rotate the current best restoration from model to observational space, and then to rotate the correction factor back

$$\phi^k(x) = p(x) \circ [R(\theta) y^k(x)] \tag{2}$$

$$c^k(x) = R(-\theta)[\rho^k(x) * p(x)] \tag{3}$$

Note that neither the WF/PC PSF nor the pixel response function are isotropic.

2.2. Restoration to a Finer "Subsampled" Grid

The standard implementations of the RL-algorithm include a subsampling option (Lucy 1991b; White 1993), so the problem of restoring to a finer grid is widely considered to be "solved." However, it has not been proven so far that the RL-algorithm with subsampling is unique or has desirable properties when unaliasing or reconstructing missing high spatial frequency components.

When the RL-method with subsampling option is used to restore an undersampled "image" frame, it will create a restoration that, while consistent with the data, is not smooth, but rather maximally pixelated (see the Appendix). So the question emerges which method to use in the case of an undersampled optical PSF, and a detector with non-ideal pixel response, as expected for WF/PC-2.

The following observations show that there is indeed some opportunity for algorithmic development:

1. The sampling theorem states: It is possible to exactly reconstruct a continuous bandlimited signal from discrete samples, if the signal is sufficiently densely sampled (and noise-free). The sampling theorem, however, does *not* state: an undersampled continuous signal cannot be reconstructed from discrete samples.

2. An undersampled signal is data with missing values (e.g. every other data row and column is missing).

3. One possesses some prior knowledge, to help in filling in the missing values:
   - the continuous signal is non-negative (apart from noise);

– the observed signal is strictly bandlimited (apart from noise);

– the PSF is known, and since its Fourier-transform (the optical transfer function, OTF), has a finite support, it always stretches beyond more than one pixel (it is in fact infinitely large).

4. The interplay between the (linear) band-limitation and the (non-linear) non-negativity constraint is not yet completely understood. So there is latitude for theoretical developments and experimentation.

In order to overcome the pixelation effect of the classical RL-algorithm, the following modified RL-algorithm is suggested which operates solely on the fine grid:

1. Create a static mask $m(x)$ which is 1 where data values have been observed and 0 elsewhere.

2. Replace the predicted values $\phi^k(x)$ by the observed ones $\phi^*(x)$ wherever the latter are known, i.e., where the mask $m(x) = 1$ and apply other constraints such as non-negativity (projection onto convex set, POCS); a dynamically relaxing band-limit constraint should presumably be included here.

3. Proceed with a standard RL-restoration.

This algorithm allows some direct control over the extent of high frequency components introduced by the restoration algorithm in its unaliasing attempt.

2.3. Detection and Treatment of Cosmic Ray Hits

The cosmic ray (CR) problem involves two separate parts: detecting CR-hits and somehow "filling in" the missing data.

Several stand-alone algorithms have previously been published to detect CR-hits (Murtagh 1992; Murtagh & Adorf 1992; Adorf et al. 1993). However, it is clear that if a restoration is carried out anyway, the CR-contamination problem is best treated within that restoration process (Freudling 1993; White 1993). For sufficiently sampled images these algorithms work well. However, for undersampled images they are affected by the pixelation problem described above.

Let us, for a moment, consider how the human visual system would find cosmic ray hits: suspicious pixels are compared with their respective neighbourhoods. More precisely, the brain, using PSF-knowledge, predicts how the observation ought to look, and the eye compares this prediction with the actual data.

The essence of this detection method can be captured in mathematical terms, using Lucy's (1974) original notation, as: The prediction $\phi^k(x)$, computed from the current best restoration $\psi^k(x)$ via convolution with the PSF, represents our current understanding of the observation. Thus any faults in the observed data $\phi^*(x)$ are best detected by comparing the data $\phi^*(x)$ with the prediction $\phi^k(x)$. We are led to the following spatially adaptive, dynamic κ-σ clipping algorithm for detecting CR-hits (and potentially other faults):

1. Estimate a map of spatially variable standard deviations $\sigma(x)$. If the Poisson-noise assumption holds, one can simply take the square root of the signal counts. Otherwise one may estimate $\sigma(x)$ from the spatial variance within a small neighborhood around location x.

2. Flag all those pixels whose data values $\phi^*(x)$ are outside $\phi^k(x) \pm \kappa\sigma(x)$ thereby producing a dynamic mask $m^k(x)$, which is 1 where the data $\phi^*(x)$ is "good" judged via $\phi^k(x)$, and 0 elsewhere. Here κ denotes a threshold value.

Since the map of standard deviations $\sigma(x)$ may be corrupted by initially missed CRs, it is best made dynamic too by updating it regularly after every m^{th} RL-iteration. This double dynamization of the

variance estimation and CR-detection steps should allow to optimally detect CR-hits even on single exposures down into the noise sea. The problem to watch out for is dynamical instability.

Several methods are available to treat CR-hits once detected. One might replace the predicted values by the observed ones wherever the latter are known (projection onto convex set) and proceed with a quasi standard RL-restoration. In mathematical notation this algorithm reads:

$$\phi^k(x) = p(x) \circ \psi^k(x) \tag{4}$$

$$\phi^{*k}(x) = m^k(x)\phi^*(x) + \left(1 - m^k(x)\right)\phi^k(x) \tag{5}$$

$$r^k(x) = \phi^{*k}(x)/\phi^k(x) \tag{6}$$

$$c^k(x) = r^k(x) * p(x) \tag{7}$$

$$\psi^{k+1}(x) = c^k(x)\psi^k(x) \tag{8}$$

This method of filling in missing data values[2] might also be beneficial in situations when the information is not completely destroyed, e.g., in the case of saturated pixels, where the recorded value can kept as a lower bound for a number of iterations, but later on would be left free to be inferred from the data (White, pers. comm.).

3. Restoring a WF/PC Frame with a Space-Variant PSF

Let us now turn to the other main topic of this contribution, namely an algorithm for restoring a full WF/PC frame (or undersampled multi-frames) with its SV-PSF. The well-known "brute-force" SV-PSF restoration method is the "sectioned" algorithm (Trussel & Hunt 1978a, 1978b) which works as follows:

1. Subdivide the observation into a regular set of M patches (the "sections") with data $\phi^*_\mu(x)$.
2. Restore each section under the hypothesis that there is a PSF $p_\mu(x)$ which is space-invariant (SI) within patch μ.
3. Sew the M restorations $\psi^k_\mu(x)$ together.

The spatial density of patches (more precisely patch centre locations) necessary to carry out a restoration depends on the degree of spatial variability of the PSF. It is clear that the patches must overlap each other by at least the effective PSF diameter in order to exclude border effects, thereby limiting the minimum size of the patches.

The main problem with the simple sectioned approach is the danger of discontinuities occurring in the restoration at the patch boundaries, particularly when the PSF is asymmetric and changes its morphology, as in the case at least for WF/PC-1.

Here is an elegant way out: if one considers the patches as "independent" overlapping observations of the same field, each with its own space-invariant (SI) PSF, then one can simply use the Lucy-Hook "co-addition" algorithm to construct a single restored image $\psi^k(x)$ from the multiple data patches $\phi^*_\mu(x)$. A spatially variable, "fuzzy" weighting function $w_\mu(x)$ can be used in order to combine the correction factors $c^k_\mu(x)$ belonging to the different patches μ

$$c^k(x) = \sum_\mu w_\mu(x) c^k_\mu(x) \tag{9}$$

[2]This simple and efficient heuristic algorithm can be (and has been) criticized for violating the first principle of data analysis "Don't modify the data." However, this criticism can be rebutted, because bad pixels never attain the same status as good ones.

It is obvious that the formula above may be generalized to multiple input data frames.

Presently, it is an open question which weighting function $w_\mu(x)$ to best use. A "hard" hat-shaped function may again lead to the problem of potential discontinuities at the section boundaries. Triangular or cosine bell-shaped weighting functions are better behaved in that respect. One could possibly "derive" a weighting function by cross-correlating a series of interpolating PSFs and use the correlation coefficient as a guide to the spatial variability of $w_\mu(x)$. Or one might apply a minimum mean-square-error criterion to optimize the choice of the functional form of the $w_\mu(x)$.

The proposed SV-PSF algorithm can be generalized in various ways. One could adapt the spatial density (and corresponding size) of the sections to the local signal-to-noise (S/N) level, since it does not make sense to restore low S/N regions very accurately. Another way of saving CPU-time might be to use a denser patch grid where the PSF changes more rapidly, e.g., at the frame edges, or in areas of increased astronomical interest (White, pers. comm.).

Of course, the success of the proposed SV-PSF algorithm hinges upon the ability to produce suitable PSFs. Whether one has to rely on theoretical models or use one of the "blind deconvolution" PSF-estimation techniques (Lucy 1994) is currently an open question.

A final remark: The proposed SV-PSF algorithm is ideally suited to a coarse-grained vector-parallel computer architecture such as TMC's Connection Machine 5.

4. Summary

Two problems related to WF/PC imagery have been addressed: restoration in the presence of undersampling and restoration in the presence of a space-variant point-spread function.

It has been argued that the process of Richardson-Lucy (RL) restoration with subsampling is not fully understood in the presence of undersampling and that further work is required. A POCS-based variant of the RL-algorithm with subsampling has been suggested for further examination. Also an alternative model-based algorithm for detecting irregular random missing data such as cosmic ray (CR) hits has been proposed. The algorithm is based on a spatially adaptive, dynamic κ-σ clipping, which should be optimal, at least for single-pixel CR-hits.

A novel, easily parallelizable algorithm for faithful image rotation has been stated. It is based on a sequence of three successive shears which in turn can efficiently be implemented using fast Fourier transforms (FFTs). The rotation algorithm is required within a joint restoration of multiple rotated WF/PC frames.

Finally a novel algorithm has been proposed for restoring full-frame WF/PC images distorted by a space-variant point spread function (SV-PSF). The algorithm is efficient and parallelizable. The algorithm allows several interesting generalizations which promise to either increase performance or restoration quality or both.

With these algorithms a comprehensive full-frame SV-PSF restoration of single or multiple WF/PC frames now appears to be feasible.

Appendix

Some simulations have been carried out to investigate how much pixelation the conventional RL-method with sub-sampling option (where the data-to-prediction ratio is upsampled by pixel replication) imprints onto the restoration of an undersampled distorted image frame.

Two grids were used, a "fine" grid with 64×64 pixels and a "coarse" grid with 32×32 pixels, i.e., half the sampling rate of the fine grid. As an image model a single sine wave was used with a wavelength of 6 pixels on the fine grid, which is still sufficiently sampled on the coarse grid (2/3 of the coarse grid's Nyquist frequency). A value of 1 was added to create a non-negative image.

Four different PSFs were used: a delta function (PSF0) and 3 diffraction limited PSFs with "aperture" radii of 31, 15, and 7 pixels (PSF1 to 3), respectively. PSF1 is approximately two-fold

(four-fold) undersampled on the fine (coarse) grid. PSF2 is sufficiently sampled on the fine grid, but about two-fold undersampled on the coarse grid. PSF3 is sufficiently sampled on both grids.

The model was convolved with each of the PSFs, then 2×2 block-averaged and down-sampled (decimated) to 32×32 pixels, generating four simulated data frames (REST0 to 3). No noise was added. The ST-ECF's IRAF code for the RL-method with sub-sampling option was used to restore the four simulated "data" frames.

As expected, the restoration REST0 (after some 20 accelerated iterations) clearly showed the coarse-grid pixel pattern: pairs of two pixels on the fine grid had exactly the same value. The restoration REST3 on the other hand showed very little pixelation, if any. The intermediate restorations REST2 and REST1 showed a degree of remnant pixelation on the fine grid which increased with the degree of undersampling.

Acknowledgments. Thanks are due to Bob Hanisch, STScI, for suggesting the SV-PSF restoration problem and to Richard Hook, ST-ECF, who helped to improve the written version of this contribution. I am particularly indepted to Rick White, STScI, for enlightening discussions and several valuable suggestions.

References

Adorf, H.-M. 1989a, *ST-ECF Newsletter*, 12, 9

Adorf, H.-M. 1989b, in *1st ESO/ST-ECF Data Analysis Workshop*, P. J. Grosbøl, F. Murtagh, & R. H. Warmels, eds., European Southern Observatory, Garching, 215

Adorf, H.-M. 1990, in *Errors, Bias, and Uncertainties in Astronomy*, C. Jaschek & F. Murtagh, eds., Cambridge University Press, Cambridge, 71

Adorf, H.-M., Hook, R. N., & Lucy, L. B. 1993, in *Space Astronomical Telescopes and Instrumentation II*, SPIE, in press

Burrows, C. 1994, this volume

Fosbury, R. 1993, in *The Future of Space Imaging*, R. Brown, ed., Space Telescope Science Institute, Baltimore, 83

Freudling, W. 1993, *ST-ECF Newsletter*, 20, 8

Hook, R. N., & Lucy, L. B. 1992, *ST-ECF Newsletter*, 17, 10

Hook, R. N., & Lucy, L. B. 1993, *ST-ECF Newsletter*, 19, 6

Lucy, L. B. 1974, AJ, 79, 745

Lucy, L. B. 1991a, *ST-ECF Newsletter*, 16, 6

Lucy, L. B. 1991b, in *The Restoration of HST Images and Spectra*, R. L. White & R. J. Allen, eds., Space Telescope Science Institute, Baltimore, 80

Lucy, L. B. 1994, this volume

Lucy, L. B., & Hook, R. N. 1992, in *Astronomical Data Analysis Software and Systems I*, ASP Conference Series 25, D. M. Worrall, C. Biemesderfer, & J. Barnes, eds., 277

Murtagh, F. D. 1992, in *Astronomical Data Analysis Software and Systems I*, ASP Conference Series 25, D. M. Worrall, C. Biemesderfer, & J. Barnes, eds., 265

Murtagh, F. D., & Adorf, H.-M. 1992, in *Data Analysis in Astronomy IV*, V. Di Ges'u, et al., eds., Plenum Press, New York, 103

Richardson, W. H. 1972, J. Opt. Soc. Am., 62, 55

Trussel, H. J., & Hunt, B. R. 1978a, IEEE Trans. Acoust. Speech Sig. Proc., 26, 157

Trussel, H. J., & Hunt, B. R. 1978b, IEEE Trans. Acoust. Speech Sig. Proc., 26, 608

White, R. L. 1993, in *Restoration Newsletter*, R. J. Hanisch, ed., Space Telescope Science Institute, Baltimore, 1, 11

The Restoration of HST Images and Spectra II
Space Telescope Science Institute, 1994
R. J. Hanisch and R. L. White, eds.

Image Restorations of High Photometric Quality

L. B. Lucy[1]

Space Telescope – European Coordinating Facility, European Southern Observatory, Karl-Schwarzschild-Strasse 2, D-85748 Garching bei München, Germany

Abstract. An image restoration technique is described that achieves high photometric accuracy for both stars and distributed emission. The technique makes use of observer-supplied information that some objects in the field are point sources and thereby eliminates the ringing and photometric bias that arise with conventional restoration procedures. A hierarchy of codes is described based on this two-channel decomposition of astronomical images. The more sophisticated of these codes incorporate simultaneous estimation of the PSF and are thus especially relevant for ground-based imaging. Observing programs are described in which these codes effectively transfer the resolution of HST to ground-based images.

1. Introduction

Astronomical images are commonly taken with the intent of making photometric measurements of objects in the field. For pre-COSTAR HST images, the question then arises: should these measurements be made on the original or on the restored image? When the objects of interest are stellar, we might intuitively expect that higher accuracy will be achieved using the restored image. The reason for this is that the sharpening of point sources should allow the photometry to be carried out with a smaller aperture, thus reducing contamination by light from nearby sources and from the background.

Unfortunately, two effects have hindered the realization of this expected gain in accuracy. The first arises when the stars to be measured are superposed on an extended background source — nebular emission or numerous unresolved stars. In this circumstance, non-linear restoration techniques — e.g., Richardson-Lucy (R-L) or Maximum Entropy (ME) — give rise to Gibbs oscillations or "ringing" as they attempt the impossible task of recovering delta functions. Because of this effect, image restoration does not necessarily increase the rate at which a photometric measurement of a stellar object converges with increasing aperture.

The second effect is statistical bias and is of particular concern for ME restorations. When entropy is defined relative to a uniform prior (or default) image, the restored ME image departs from uniformity because of real structure in the observed image, but only by the amount required by the adopted goodness-of-fit criterion. Accordingly, the restored ME image is biased in favor of uniformity (flatness), and this implies that the highest intensity peaks are underestimated.

For stellar photometry, these problems have led to the suggestion that the restored HST image be used merely to identify and locate point sources, with the actual photometric measurements being then made on the original image. Here, however, the possibility is explored of developing more powerful restoration procedures that eliminate the above difficulties. Some years ago already, Frieden and Wells (1978) implemented a scheme that effectively solved the first of the above problems. Their scheme comprised the following discrete steps: 1) a model of the slowly varying background is constructed; 2) this model is subtracted from the original image and the resulting

[1] Affiliated to Astrophysics Division, Space Science Department, European Space Agency

residual image is deconvolved; 3) the background model is added to the deconvolved residual image to obtain the final restored image. In this paper, a somewhat related procedure is described, leading to a hierarchy of fully automatic codes of increasing sophistication which yield restorations of high photometric precision for both stars and distributed emission. The performance of these codes is tested and illustrated in Paper II (Hook & Lucy 1994).

2. Method

The method adopted might be described as image restoration incorporating designated sources having specified shape and known positions. For astronomical images, the designated sources are naturally taken to be point sources (stars, QSOs, AGNs), so the specified shape is the delta function. Clearly, an observer can commonly, with high confidence, identify many objects in an image as point sources. Accordingly, it makes sense to develop codes that use this extra information and, at least for these designated point sources, thereby eliminate the ringing associated with failed attempts to recover delta functions.

In 1-D vector notation, the adopted model for $\Psi(\xi, \eta)$, the intensity distribution on the celestial sphere, is

$$\Psi_j = \psi_j + \psi_j^* \,. \tag{1}$$

Here ψ_j represents distributed emission and ψ_j^* the designated point sources. This latter representation is achieved by defining ψ_j^* to be zero everywhere except at the locations $(j = j_1, j_2, \ldots, j_N)$ of N point sources.

This model of the object can be mapped onto the image plane using the equation of image formation, and the resulting prediction Φ_i for the intensity distribution in the image plane can then be compared to the observed distribution $\tilde{\Phi}_i$. The posed image restoration problem is thus to determine ψ_j and ψ_j^* in order to achieve an optimum fit of Φ_i to $\tilde{\Phi}_i$.

An obvious first thought is to adopt the R-L procedure and determine ψ_j and ψ_j^* with an iterative scheme that asymptotically yields Maximum Likelihood estimates subject to the appropriate non-negativity and normalization constraints. But this results in indeterminacy since an individual star can be represented either in the ψ_j^* vector or by the corresponding single-pixel peak added to the ψ_j vector, or indeed by any linear combination of these extremes.

To eliminate this indeterminacy, we must impose a constraint on ψ_j that excludes single-pixel peaks. In other words, we must impose a resolution limit on ψ_j that is somewhat larger than the limit implied by the numerical discretization. This is achieved by introducing regularization into the optimization problem for determining ψ_j and ψ_j^*. Of the many possible regularization terms (see, e.g., Titterington 1985), we choose the entropic form

$$S = -\Sigma \frac{\psi_j}{\sigma} \ln \frac{\psi_j}{\chi_j} \,, \tag{2}$$

where $\sigma = \Sigma\psi_j$ and where the *floating default* (Horne 1985, Lucy 1994)

$$\chi_j = \Sigma R_{jm} \psi_m \,. \tag{3}$$

Here the *resolution kernel* R_{jm} is a bell-shaped function whose width is the required resolution limit. Structure in ψ_j on a smaller scale results in a decrease in the entropy S and so is disfavored.

3. Code I: Known PSF

In image restoration, the PSF $P\,(x|\xi)$ is usually regarded as known. For this standard case, the ideas of §2 lead to the following optimization problem: the quantities ψ_j and ψ_j^* are determined by maximizing the objective function

$$Q(\psi_j, \psi_j^*; \mu) = H + \alpha S + \mu(\Sigma\Psi_j - 1) \,. \tag{4}$$

Here H denotes the reduced log-likelihood of the observed image and is given by

$$H = \Sigma \tilde{\Phi}_i \ln \Phi_i \,, \tag{5}$$

with

$$\Phi_i = \Sigma P_{ij} \Psi_j \,. \tag{6}$$

The regularization parameter α controls the weight given to the regularization of ψ_j, and the Lagrange multiplier μ allows overall flux conservation to be imposed.

4. Construction of Algorithms

The R-L algorithm (Richardson 1972; Lucy 1974) was initially derived from Bayes' theorem, but in fact its use is really justified by the subsequent discovery (Lucy 1974; Shepp and Vardi 1982) that every iteration increases the likelihood assigned to the observed image. Thus, in the usual case where no stars are designated point sources, the objective function that is asymptotically maximized by the R-L algorithm is

$$Q(\psi_j; \mu) = H + \mu(\Sigma \psi_j - 1) \,. \tag{7}$$

Knowing this, we can rewrite the R-L algorithm as an operation on Q. The result for the increment $\psi^{r+1} - \psi^r$ is

$$\Delta \psi_j = \psi_j \left(\frac{\partial Q}{\partial \psi_j} \right)^* \,, \tag{8}$$

where the asterisk indicates that μ has been evaluated by requiring that

$$\Sigma \Delta \psi_j = 0 \,. \tag{9}$$

Eq. (8) suggests an operational procedure for deriving potentially effective algorithms: write down the function Q whose optimization defines the restoration problem, then derive an iterative correction scheme by applying Eq. (8).

With Q given by Eq. (4) and S by Eq. (2), the resulting algorithm simplifies to

$$\Delta \psi_j = \psi_j \left(C_j - \alpha \frac{\partial S}{\partial \psi_j} \right) \tag{10}$$

and

$$\Delta \psi_j^* = \psi_j^* C_j \,, \tag{11}$$

where C_j is the R-L correction factor.

5. Comments

The use and performance of Code I will perhaps be clarified by the following remarks:

Initialization The positions of the designated point sources ($j = j_1, j_2, \ldots, j_N$) must be given as well as initial amplitudes for ψ_j and $\psi_{j\ell}^*$. These are taken to be constant and such that $\Sigma \psi_j = \Sigma \psi_{j\ell}^* = \frac{1}{2}$.

Normalization The integrated flux of the point sources (ψ_j^*) plus that of the distributed emission (ψ_j) is constrained to equal that of the observed image ($\tilde{\Phi}_i$). The allocation of flux between these two channels (initially 1:1, see above) is adjusted iteratively as Q is maximized.

Regularization Applies only to the distributed emission. This implies that the converged point-source amplitudes $\psi^*_{j\ell}$ are Maximum Likelihood estimates relative to the highly flexible model of background emission provided by ψ_j. This latter aspect is to be compared with the mathematically simple background models used in standard photometric software packages.

Spatially-Variant PSFs As implemented, Code I assumes a spatially-invariant PSF but only in order to use FFTs. As with the R-L algorithm, this assumption is not otherwise necessary.

Point Sources These are modelled as single active pixels in a sea of zeroed pixels. If more accurate centering is necessary, sub-pixelation is readily introduced (Lucy and Baade 1989; White 1990). Such representations of delta functions facilitate the use of discrete FFTs and lead to simple code. Nevertheless, for highest accuracy, continuous positions should be used.

Floating Default A ME restoration with uniform prior gives the most probable distribution of photons in the restored image for a specified goodness-of-fit to the observed image. Although intuitively appealing, this solution, as noted earlier, is photometrically biased as a result of redistributing photons from high to low intensity regions. Moreover, when the PSF is compact, this redistribution is physically inconsistent with the independence of different "picture elements." However, for a floating default computed with a compact kernel, this bias is greatly reduced (e.g., Lucy 1994) as is this inconsistency.

Because the entropic form with floating default is probably not derivable by combinatorial arguments, it should be regarded as a useful mathematical device and not ascribed a more fundamental status than other regularization functions (Titterington 1985).

Non-Negativity As implemented, Code I maximizes Q subject to the constraints $\psi_j \geq 0$ for all j and $\psi^*_{j\ell} \geq 0$ for all ℓ. There would be some merit in changing the second constraint to $\psi_j + \psi^*_{j\ell} \geq 0$ for all ℓ. This would then allow $\psi^*_{j\ell} < 0$, which, though unphysical, avoids bias when estimating magnitudes near the noise limit, as happens, for example, in following the fading of a SN. An algorithm that achieves this has been constructed.

Non-Designated Sources Given the resolution limit imposed on distributed emission, the question arises as to how well point sources are represented that are not designated as such. This is investigated in Fig. 1 where all emission derives from point sources but only one — the brightest — is designated as such. Despite the resolution limit, the restored background (ψ) shows sharp peaks at the positions of several stars. Accordingly, this model of the background could be used to identify and locate additional point sources and thus form a step in an iterative procedure that maximally decomposes the crowded field into individual stars, whose magnitudes are then derived from the final solution for $\psi^*_{j\ell}$.

6. Code II: Unknown PSF

In the standard case of known PSF (Code I), knowledge of the PSF is often derived from the very image to be restored. This is of necessity true for ground-based images where the seeing-dominated and thus time-variable PSF must be determined from suitably isolated stars in the field. For such cases, a natural generalization of Code I is a code that incorporates these "PSF stars" into the list of designated point sources and treats the PSF as an unknown function to be determined simultaneously with ψ_j and $\psi^*_{j\ell}$. Clearly, this generalization effectively mandates the assumption of a spatially-invariant PSF.

The objective function whose maximization defines the above restoration problem is

$$Q(\psi_j, \psi^*_j, \rho_k; \mu, \lambda) = H + \alpha S + \mu(\Sigma\Psi_j - 1) + \lambda(\Sigma\rho_k - 1)\,. \qquad (12)$$

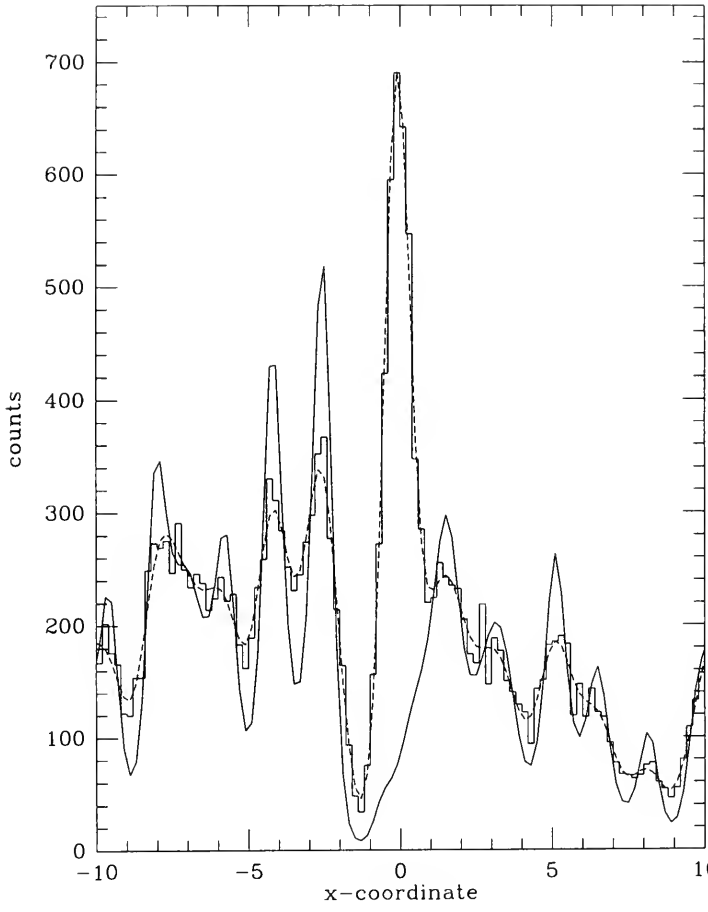


Figure 1. One-dimensional illustration of application of Code I to crowded star field. The brightest star at $x = 0$ is a designated point source, the others are not. The observed image ($\tilde{\Phi}$) is plotted as a histogram and the predicted intensity distribution in the image plane (Φ) as a dashed curve. The continuous curve is the restored and regularized model (ψ) of the distributed emission — background stars, in this case. Further individual stars can be recognized in ψ.

The new symbols introduced here are ρ_k the unknown PSF and a second Lagrangian multiplier λ to impose normalization of the PSF. Because the PSF is now assumed to be translationally invariant, the predicted intensity distribution in the image plane is

$$\Phi_i = \Sigma \rho_{i-j} \Psi_j \ . \tag{13}$$

As with Code I, an algorithm for maximizing Q is derived following the operational procedure of §4. Details are omitted.

6.1. Remarks

Code II is not without pitfalls. The first is that Q has multiple maxima of approximately equal height. In fact, if N point sources are designated, there are $N + 1$ such maxima, with each of the N spurious maxima corresponding to the entire image ($\tilde{\Phi}$) being attributed to the PSF-broadening of just one of the point sources. These spurious solutions ($m = 1, 2, \ldots, N$) thus have the form: $\psi_j = 0$ all j, $\psi^*_{j_\ell} = 0$ for $\ell \neq m$, $\psi^*_{j_m} = 1$, and $\rho_k = \tilde{\Phi}_{j_m+k}$. Fortunately, these unwelcome solutions are readily avoided with sensible initialization (see §5) and would in any case be readily recognized as spurious.

A more serious problem arises when all the designated point sources are superposed on distributed emission. The resulting PSF and the allocation of emission between point sources and background are then determined by and sensitive to the regularization procedure. Limited experiments suggest that sensible results in this circumstance require strong regularization with a rather broad resolution kernel in Eq. (3).

Code II might seem to be an example of blind deconvolution (Nisenson et al. 1990) since both the restored image and the PSF are derived simultaneously from a single observed image. However, in contrast to blind deconvolution, the designated point sources here play an essential role in making the PSF determinate. Accordingly, as noted above, this code should be regarded as consolidating conventional image-processing procedures into a single, automatic code.

7. Code III: Regularized PSF

Even when there are numerous designated point sources, the empirical PSF derived with Code II will be noisy at low intensity levels. For the seeing-dominated PSFs of ground-based images, this must be expected in the outer halo that surrounds the approximately Gaussian core (King 1971). An obvious next step, therefore, is to regularize the PSF in order to damp out such noise fluctuations, and this can be achieved simultaneously with the estimation of ψ_j, $\psi^*_{j_\ell}$ and ρ_k if an extra entropic term is added to Eq. (12). Moreover, if this new term also has a floating default, its kernel can be chosen to optimize the regularization by taking advantage of prior information about the PSF. For example, apart from diffraction spikes, a ground-based PSF should be circularly-symmetric. It is useful then to take the floating default to be the azimuthally-averaged ρ_k. The final ρ_k then departs from circular symmetry only if the stellar images demand it — perhaps due to tracking errors — and such distortions are corrected in the restored image.

Thus far only a 1-D experimental version of Code III has been tested.

8. Scientific Opportunities

The successful tests of Codes I and II reported in Paper II prompt thoughts as to the scientific programs that could benefit from their application. Some of these might be described as software-driven observing programs.

Host galaxies Conventional deconvolution of images of QSOs does not usefully contribute to morphological studies of host galaxies because of the ringing centered on the deconvolved quasars.

But with Codes I or II, the quasar can be treated as a designated point source. This not only eliminates ringing but also provides the restored and regularized image (ψ_j) of the host galaxy with quasar removed. With the alternative technique of PSF fitting and subtraction, the residual image is not restored and also suffers from the Poisson noise associated with the bright quasar and left behind on subtraction.

Measurement of H_0 Major programs underway with HST aim to discover Cepheids in relatively distant galaxies and thus to determine H_0 using the locally calibrated period–luminosity relation. In these programs, the HST resolves crowded fields into stars thus allowing Cepheids to be discovered and photometered.

An alternative approach is the following: First, a single HST image is used to resolve the field into stars and to determine accurate star positions. Multiple ground-based images of the field are then decomposed into stars using Codes I or II with the HST star positions as input. Finally, from the values ψ_{jt}^* on the different frames, Cepheids are discovered numerically and their periods found.

When so used, Codes I and II effectively transfer HST's resolving power to ground-based telescopes. Besides thus being highly economical of HST time, this technique has the further advantage that the photometry is in well-understood systems established with and for ground-based telescopes.

Imaging of Nebulae Many planetary nebulae and supernovae remnants occur in extremely crowded star fields, to the point that scientifically useful measurements are hard to make. Again Codes I or II could be used with star positions from one HST image to obtain star-free ground-based images of such nebulosities in diagnostically-important emission lines.

Supernovae Light Curves Because of variable seeing, conventional fixed-aperture photometric measurements of SNe are contaminated with variable amounts of light from the host galaxy. But with Codes I or II the SN can be designated as a point source; its intensity is then determined relative to a detailed model of the host galaxy with seeing taken into account.

References

Frieden, B. R., & Wells, D. C. 1978, J. Opt. Soc. Am., 68, 93

Hook, R. N., & Lucy, L. B., this volume (Paper II)

Horne, K. 1985, MNRAS, 213, 129

King, I. R. 1971, PASP, 83, 199

Lucy, L. B. 1974, AJ, 79, 745

Lucy, L. B. 1994, A&A, submitted

Lucy, L. B., & Baade, D. 1989, in *1st ESO/ST-ECF Data Analysis Workshop*, P. Grosbøl, R. H. Warmels, & F. Murtagh, eds., European Southern Observatory, Garching, 219

Nisenson, P., Standley, C., & Gay, D. 1990, in *The Restoration of HST Images and Spectra*, R. L. White & R. J. Allen, eds., Space Telescope Science Institute, Baltimore, 103

Richardson, W. H. 1972, J. Opt. Soc. Am., 62, 55

Shepp, L. A., & Vardi, Y. 1982, IEEE Trans. Med. Imag., MI-1, 113

Titterington, D. M. 1985, A&A, 144, 381

White, R. L. 1990, in *The Restoration of HST Images and Spectra*, R. L. White & R. J. Allen, eds., Space Telescope Science Institute, Baltimore, 139

The Restoration of HST Images and Spectra II
Space Telescope Science Institute, 1994
R. J. Hanisch and R. L. White, eds.

Image Restorations of High Photometric Quality. II. Examples

R. N. Hook & L. B. Lucy[1]

Space Telescope – European Coordinating Facility, European Southern Observatory, Karl-Schwarzschild-Str. 2, D-85748 Garching bei München, Germany

Abstract. A generalized form of the Lucy-Richardson restoration method is described in which images in the object plane are divided into two channels. One of these contains point sources (i.e., δ-functions) and the other models a smooth background distribution. The latter image is regularized by the use of an entropy term and hence has enforced smoothness. This approach avoids problems encountered in the photometry of the results of non-linear restoration methods. Examples of the use of this technique on standard simulated HST data frames are given, including some assessments of its photometric accuracy. Implementations running under both IRAF and MIDAS are available.

1. Introduction

The most popular and effective methods for the restoration of HST data are non-linear and hence redistribute flux in a way which often degrades the photometric content of the data. They also cause artifacts in the background to appear when large numbers of iterations are used, and the number of iterations which will give best results is often difficult to estimate. Several studies (e.g., Cohen 1988, Linde & Spännare 1993) have suggested that a workable compromise for point source photometry is to use restored images, which show better contrast and object discrimination, to identify objects and then to perform photometry by using standard PSF fitting packages (e.g., DAOPHOT, Stetson 1987) on the unrestored data. Other work (Busko 1993) has shown that there is often a tradeoff between photometric accuracy in the measurement of point sources and the quality of the background in the resultant restored image.

Many astronomical images do not contain a uniform spread of spatial frequencies. Instead they consist of many point sources and a smoother background intensity distribution. Standard restoration methods do not use this information and hence cannot significantly sharpen point sources without causing the background to break up into speckles. Other methods, such as those based on Maximum Entropy considerations, can impose smoothness but generally underestimate point source fluxes. We have adopted a different approach which separates the known point sources, which are forced to be δ-functions, from a background which is forced to be smooth. Hence we obtain the best of both worlds by using extra information about the image. As a further generalization, the method may be used to iteratively find the PSF using a form of "blind iterative deconvolution." A description of the logical steps in the development of this method and more detailed mathematics are given elsewhere in this volume (Lucy 1994). In this paper several illustrations of its use on different classes of problem are presented along with details of the implementation. The test images are taken from those prepared by the STScI Image Restoration Project.

This technique suggests several new possible observational projects using both space and ground based telescopes.

---

[1] Affiliated to Astrophysics Division, Space Science Department, European Space Agency

2. Technique

Let an estimate of true intensity distribution on the sky be represented as the sum of two components:

$$\Psi = \psi^* + \psi, \tag{1}$$

where ψ^* is the intensity of the point sources (mainly zero apart from δ-functions at the positions of designated sources) and ψ is the intensity of the background. The corresponding estimates for the components which are observed are $\phi^* = \psi^* \otimes P$ and $\phi = \psi \otimes P$ where P is the point spread function and $\otimes$ is the convolution operator. The predicted total observed intensity is $\Phi = \phi^* + \phi$. We wish to obtain the "best" estimates for ψ^* and ψ and choose to maximize the quantity:

$$Q = \sum_i \tilde{\Phi}_i ln \Phi_i + \alpha S \tag{2}$$

where $\tilde{\Phi}$ is the observed intensity distribution, and the summation is performed over all pixels in the image. The second term is only evaluated over the *background* image and it has a regularizing effect. The quantity α is a free parameter specified by the user and controls the strength of the regularization in comparison with the first term which is the logarithm of the likelihood. When α is zero the objective function becomes the same as that of the standard Lucy-Richardson technique. S takes the form of an entropy expression which may be calculated in two ways, relative to a flat background:

$$S = -\sum_i \phi_i ln \phi_i \tag{3}$$

or as relative to a *floating default* $\backslash$:

$$S_{fl} = -\sum_i \phi_i ln \frac{\phi_i}{\backslash_i} \tag{4}$$

Note that ϕ must be renormalized in calculating these entropy terms (Lucy 1994). The floating default is most conveniently chosen to be a smoothed version of the current estimate for the background. In practice the constant default works well for images in which the background is roughly flat such as those used for star cluster photometry. For cases where there is significant structure, such as stars superimposed on a galaxy, the floating default method works dramatically better and is essential for good results. Examples of both types of default are given below.

An iterative scheme has been developed to maximize the above objective function. It has also proved possible to accelerate this algorithm in a manner analogous to that which has been described for the standard Lucy-Richardson restoration scheme (Hook & Lucy 1993). The addition of the entropy term removes the automatic non-negativity conservation of the standard Lucy-Richardson method so care must be taken to avoid negative points which lead to gross instabilities. This may be done using the acceleration option which will act in these cases as a *deceleration*. The regularization of the background means that no ugly artifacts appear and the iterations may, and normally should, be allowed to continue as far as convergence.

3. Examples and Photometric Tests

This method has been tested on many kinds of data both real and simulated. Here we present three examples in which the structure of the background is of increasing interest and photometry of point sources of diminishing importance. All three examples use simulated data from the collection prepared by the STScI Image Restoration Project and available from the STEIS anonymous FTP service (Hanisch 1993). A final example illustrates how the "blind iterative restoration" option may be used to obtain a PSF image from a data frame.

3.1. A Crowded Field Observed With HST WFC

Photometry of crowded fields is a very important problem for which many highly developed and thoroughly tested codes exist. Most of these were developed for ground based images and there may be special problems with handling HST data where images are badly undersampled and the greater extent of the PSF wings makes the crowding problem very severe. The photometric performance of the new algorithm has been tested using the WFC star cluster simulation which is available from STEIS. We used just the lower left quadrant of the image which includes the very dense core. The image was expanded by a factor of four in x and y before processing. Fig. 1 shows the results of this and Fig. 2 shows the photometric accuracy and also a histogram of the background image. These can be compared to similar graphs given for different restoration methods by Busko (1993). The photometry is seen to be highly linear and to lack bias. The scatter is larger than predicted from pure noise considerations because of the high degree of crowding.

3.2. Stars Superimposed on a Galaxy

Many interesting problems in astronomy involve trying to do photometry of point sources on a highly structured background. An important example is the photometry of Cepheid variables in nearby galaxies. The new algorithm is particularly suited to this problem because it can effectively model any background simultaneously with measuring the point source magnitudes. There is no assumption about the background structure except that it must be smooth to some (user controllable) degree. A good test for this kind of work is shown in Fig. 3 where the input data is a ground-based image of the nearby galaxy M51 on which some point sources of known brightness have been added. Fig. 4 shows the quality of the resulting photometry. The output image (upper left) also illustrates how the regularization effectively suppresses the speckles in the background which are so prominent in the standard restoration (lower right).

3.3. A QSO "Fuzz" Simulation

In an important class of problem there is only a single point source but this is so bright as to swamp interesting structure lying underneath. Fig. 5 shows the results of processing a simulated PC image of a QSO with an underlying spiral galaxy. Although conventional restoration manages to recover the spiral arms well, there is still an inaccurate restoration of the regions close to the nucleus and the familiar speckles. This method has also been found to be among the most effective for this kind of work on real images obtained from the ground (Stockton & Ridgway 1993).

3.4. Blind Iterative Restoration

In many cases the PSF cannot be reliably predicted and must be obtained directly from the data image. The two channel method may be generalized to simultaneously improve both the estimate of the true object intensity distribution and the PSF. This is more stable and effective than other blind iterative methods because the point sources are at designated positions. Fig. 6 shows an example of obtaining an improved PSF from a subset of the simulated WFC image which was used in the crowded field example given above. More details of the method are given in Lucy (1994).

4. Implementation

This method has been implemented as a FORTRAN code called PLUCY which uses the F77/VOS as an interface to the IRAF package. Both stationary and floating defaults are available as well as acceleration, sub-pixelation, blind iterative restoration, and several other options. It is not fully supported but there is a detailed help file and it is made available to anyone who wishes to try it out. All the files and instructions necessary to build a local version are available on the ST-ECF anonymous FTP archive (ecf.hq.eso.org = 134.171.11.4) in the directory `pub/swlib/restoration/iraf`. The same core code is also available to run within the MIDAS system and this may be collected from the

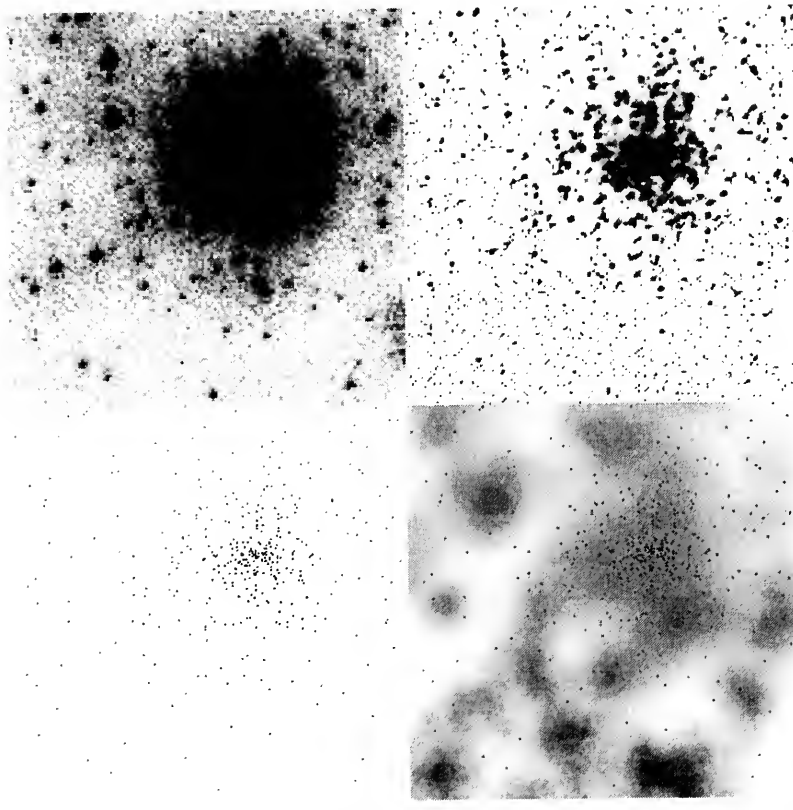


Figure 1. *Top left:* A section from a simulated HST WFC image of a globular cluster. The PSF is spatially invariant and the positions of all stars are known exactly. *Top right:* the result of 100 iterations of the standard Lucy-Richardson restoration. *Bottom left:* the result of a two channel photometric restoration as described in the text. A floating prior with a smoothing kernel with $\sigma = 5$ pixels was used and 50 (accelerated) iterations were applied. *Bottom right:* the same image as at bottom left except with the display levels ten times as deep to reveal the residual (smooth) background structure.

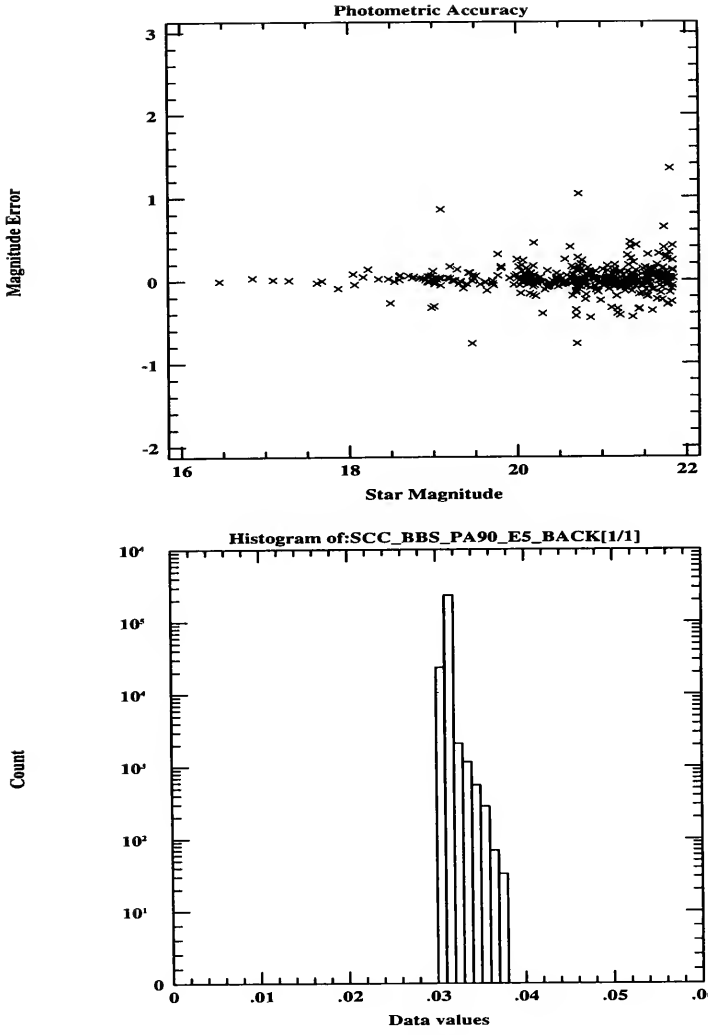


Figure 2. *Upper plot:* the photometric errors from the two channel restoration shown in Fig. 1. The actual star magnitudes are arbitrary but the errors are accurate. All stars have been included, even those close to the edges of the edge of the frame. *Lower plot:* the histogram of the background image.

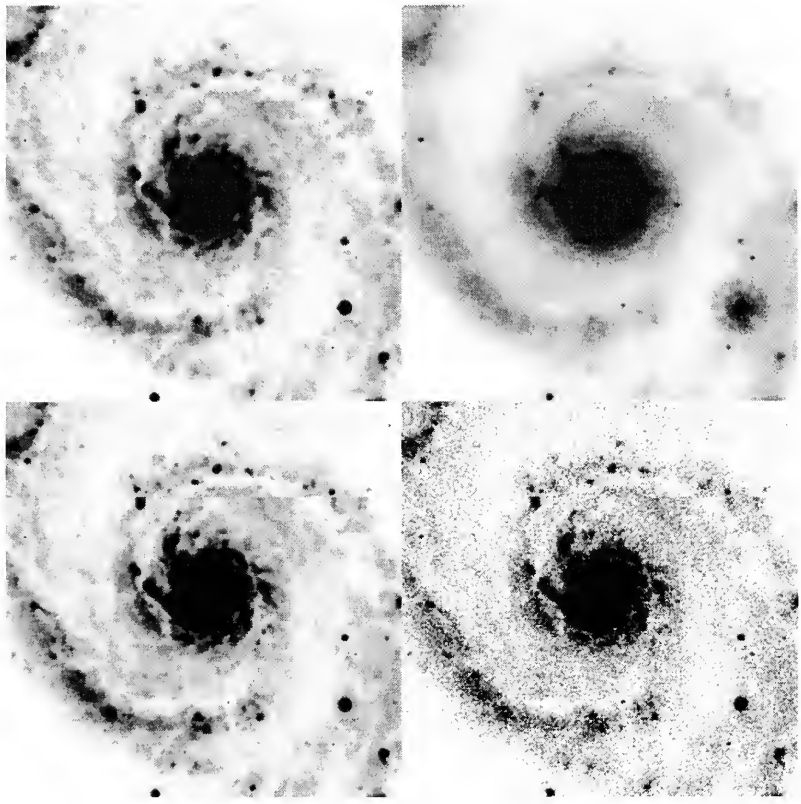


Figure 3. *Bottom left:* a section of a simulation of a spiral galaxy which has had several point sources of known brightness added as δ-functions. *Top right:* the result of subsequently convolving this with the HST WFC PSF and giving appropriate noise. *Bottom right:* the result of applying 200 iterations of the standard Lucy-Richardson restoration method to the image at upper right. *Top-left:* the result of 100 iterations of the two-channel restoration method using the known positions of the stars.

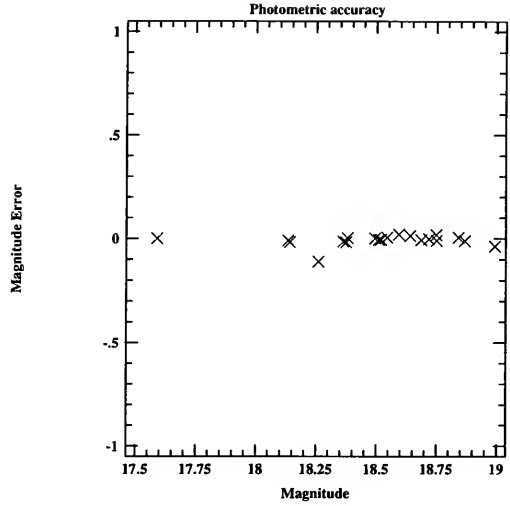


Figure 4. The photometric accuracy of the measurements of the stars shown in Fig. 3. The magnitudes on the x axis are arbitrary.

same FTP site in the directory `pub/swlib/restoration/midas/lucy`. It is planned to incorporate these codes into future releases of both STSDAS (in the *contrib* package) and MIDAS.

References

Busko, I. 1993, in *Restoration Newsletter*, R. J. Hanisch, ed., Space Telescope Science Institute, Baltimore, 1, p. 60

Cohen, J. G. 1991, AJ, 101, 734

Hanisch, R. J. 1993, in *Restoration Newsletter*, R. J. Hanisch, ed., Space Telescope Science Institute, Baltimore, 1, p. 76

Hook, R. N. & Lucy L. B. 1993, in *Science with the Hubble Space Telescope*, P. Benvenuti & E. Schreier, eds., European Southern Observatory, p. 245

Linde, P. & Spännare, S. in *Proceedings of the 5th ESO Data Analysis Workshop*, P. Grosbøl & R. C. E. de Ruijsscher, eds., p. 131

Lucy, L. B. 1974, AJ, 79, 745

Lucy, L. B. 1994, this volume

Stetson, P. B. 1987, PASP, 99, 191

Stockton, A. & Ridgway, S. 1993, personal communication

Richardson, W. H. 1972, J. Opt. Soc. Am., 62, 55

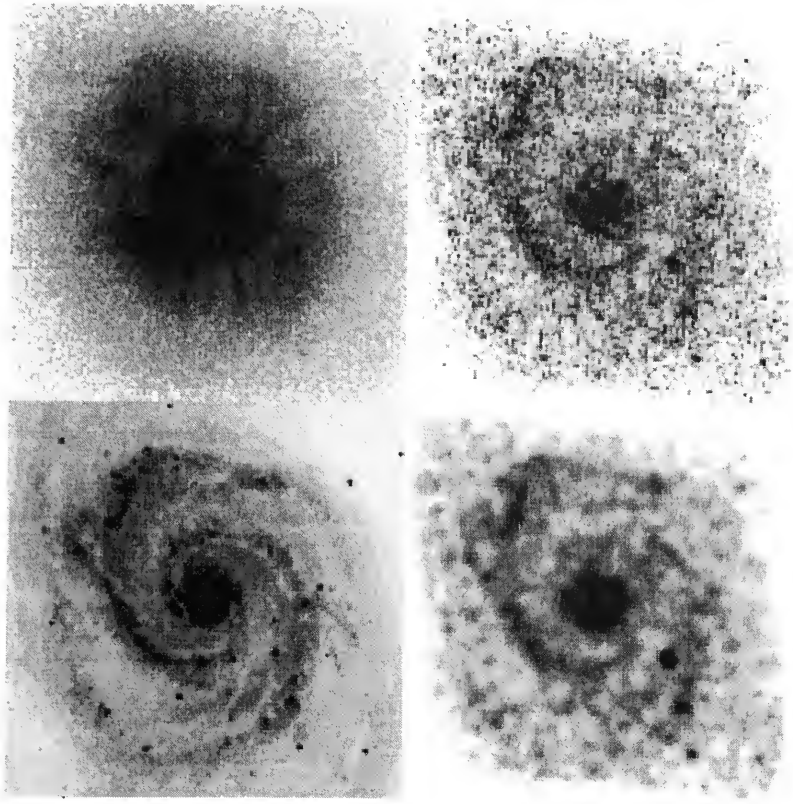


Figure 5. The results of applying the two-channel restoration technique to a simulated HST PC image of a QSO consisting of a spiral galaxy having a point source added at the nucleus which is 100 times as bright as the peak in the underlying galaxy. *Top left:* the original input image with appropriate noise. *Bottom left:* the true, noise free image of the QSO plus galaxy. *Top right:* the result of applying 30 accelerated standard Lucy-Richardson iterations to the image shown at top left. *Bottom right:* the result of 50 iterations of the two channel method using a floating prior with a smoothing kernel having $\sigma = 1$ pixel.

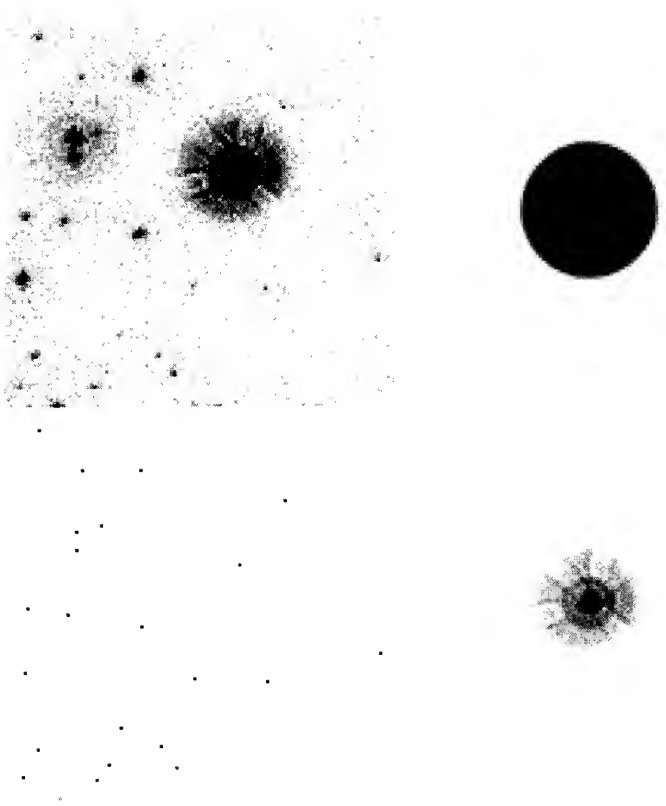


Figure 6. An example of "blind iterative restoration." *Top left:* a small section of a WFPC simulated star cluster image. *Top right:* an initial estimate for the PSF, a "top hat" with a rounded edge. *Bottom left:* the result of restoration using the two channel method with PSF modification switched on. *Bottom right:* the resultant PSF derived from the data.

Beyond Richardson-Lucy

The Restoration of HST Images and Spectra II
Space Telescope Science Institute, 1994
R. J. Hanisch and R. L. White, eds.

Generalized Cross-Validation as a Stopping Rule for the Richardson-Lucy Algorithm

Kevin M. Perry and Stanley J. Reeves

Department of Electrical Engineering, Auburn University, Auburn, AL 36849-5201

Abstract. This paper presents a criterion for stopping non-linear iterative algorithms, specifically the Richardson-Lucy algorithm that is widely used to restore images from the Hubble Space Telescope. The criterion is based on generalized cross-validation and is also computed iteratively. We will present examples displaying the power of the stopping rule, and will discuss the abilities and shortcomings of this method. We also present a caveat about the method.

1. Introduction

The restoration of the Hubble Space Telescope (HST) image data is currently being done using several different algorithms. One of these algorithms is the Richardson-Lucy (R-L) algorithm (Richardson 1972, Lucy 1974), which is a realization of the Expectation-Maximization (EM) algorithm (Dempster et al. 1977). The R-L method is iterative in nature, and converges to the maximum likelihood solution (Shepp and Vardi 1982). Unfortunately, due to the noise inherent in the problem, this maximum likelihood solution may not be the most visually pleasing solution. This noise amplification problem has been observed in other iterative algorithms, and the method of stopping the algorithm before it has reached convergence has been shown to be a valid method of regularizing the solution (Biemond et al. 1990). The reasoning behind stopping the iterations is as follows: the algorithm attempts to fit the solution to the observed data, but the observed data is degraded with noise. Thus, at some point in the process, the solution is being fit more to the noise than the image data. Therefore, the process should be stopped at the point where there is a balance between the fit to the image data and the amplification of the noise. The randomized generalized cross-validation (RGCV) criterion has been used as a stopping rule for linear iterative restoration algorithms (Reeves and Perry 1992). In this paper, we will show how the RGCV criterion can be applied to nonlinear algorithms, specifically the R-L method.

For our purposes, the observed image g shall be modeled as the result of a blur operation D upon the original image f with an additive term to model any noise from the physical system or the observation system. Therefore,

$$g = Hf + n \ ,$$

where the images and noise are ordered as vectors. The problem is then to determine f given the observation g and often the blur operation D. Some information about the statistical nature of the noise may also be available. The problem is generally ill-posed.

2. The Richardson-Lucy Algorithm

The algorithm credited to Richardson and Lucy has the following form:

$$\hat{f}_{n+1} = \hat{f} \left[H^T \left[\frac{g}{H\hat{f}_n} \right] \right] \ .$$

Note that in this form of operator notation, an operation between two vectors is done point-by-point. For example, the division by $H\hat{f}_n$ is point-by-point, and the multiplication by $\hat{f}_n$ is point-by-point.

3. The Randomized Generalized Cross Validation (RGCV) Criterion

Cross-validation divides the data into two sets. It uses one set to determine some characteristics of the data, and then uses the other set to verify the characteristics. Extending this concept, generalized cross-validation (GCV) uses each data point as the testing set, and then averages a weighted prediction error over the entire data set (Golub et al. 1979).

The GCV criterion can be stated as

$$V(n) = \frac{\frac{1}{N}\|(I - A_n)g\|^2}{\left[\frac{1}{N}\mathrm{tr}(I - A_n)\right]^2},$$

where N is the number of data points. The matrix A_n is called the influence matrix at the nth iteration. If we define $\hat{f}_n$ as the nth estimate for the original data as determined by some iterative algorithm, then A_n is defined as

$$A_n g = H\hat{f}_n .$$

The main drawback of the GCV criterion comes from the denominator of $V(n)$. Unlike the numerator, which can be easily computed at each iteration, the denominator term, needs the equivalent of N restoration computations to be computed explicitly. This computational expense is due to the trace term. Therefore, we estimate the trace term using the following lemma.

Lemma 1 *Let B be an $N \times N$ symmetric matrix and let $u = (u_1, ..., u_N)^T$ be a vector of N independent samples of a zero-mean random variable U with a variance of one. Then*

$$E[u^T B u] = \mathrm{tr}B .$$

Proof The proof is straightforward.

Several researchers independently observed that the denominator term can be estimated using this property (Girard 1989, Hutchinson 1990, Reeves and Mersereau 1991). So the denominator of V(n) can be computed as

$$\left[\frac{1}{N}u^T(I - A_n)u\right]^2 = \left[\frac{1}{N}(u^T u - u^T A_n u)\right]^2 ,$$

and $V(n)$ is now called the Randomized GCV (RGCV) criterion. Still, we do not want to compute A_n. So using the relationship $A_n g = H\hat{f}_n$, we define a new image $\hat{w}_n$ that satisfies

$$A_n u = H\hat{w}_n ,$$

which means that $\hat{w}_n$ is like a "restored" version of u at the nth iteration.

Therefore, instead of computing A_n explicitly, we can compute $\hat{w}_n$ by restoring u in the same manner as we restore g. Thus, $V(n)$ now has the form

$$V(n) = \frac{\frac{1}{N}||g - H\hat{f}_n||^2}{[\frac{1}{N}(u^T u - u^T H\hat{w}_n)]^2} .$$

The computation of the RGCV criterion simply boils down to the computation of $\hat{w}_n$. The factor we must deal with is "restoring" u in the same manner as g. Fortunately, the influence matrix A_n can also be defined as (Wahba 1983)

$$a_{n_{i,l}} = \frac{d\hat{g}_{n,i}}{dg_l},$$

where $a_{n_{i,l}}$ is the (i,l)th entry in the matrix A_n, and

$$\hat{g}_n = H\hat{f}_n \ .$$

Therefore, if one computes $\frac{d\hat{g}_{n,i}}{dg_l}$ for each iteration, and uses

$$H\hat{w}_{n,i} = \sum_l \frac{d\hat{g}_{n,i}}{dg_l} u_l$$

as the update for $\hat{w}_n$, then the "restoration" on u proceeds as the restoration on g. Equivalently, from $\hat{g}_n = H\hat{f}_n$,

$$\hat{w}_{n,i} = \sum_l \frac{d\hat{f}_{n,i}}{dg_l} \ .$$

Thus, if the derivative $\frac{d\hat{f}_{n,i}}{dg_l}$ can be computed analytically for all n, then the update for $\hat{w}_n$ can be established.

4. RGCV for Steepest Descent

In order to demonstrate the use of the derivative method for computing the RGCV criterion, we will derive the update for $\hat{w}_n$ for the steepest-descent algorithm. We have presented this method in a previous paper, but we did not present the vigorous justification at that time (Reeves and Perry 1992).

The basic form of the steepest-descent method is

$$\hat{f}_{n+1,i} = \hat{f}_{n,i} + \beta_n[H^T(g - H\hat{f}_n)]_i \ .$$

Note that β_n, the relaxation parameter, can be a fixed scalar, or it can be computed as a scalar function of the current estimate $\hat{f}_n$. In any case, it is considered to be a constant with respect to g_l.

The first step is to take the derivative of $\hat{f}_{n+1,i}$ with respect to g_l. The result is

$$\frac{d\hat{f}_{n+1,i}}{dg_l} = \frac{d\hat{f}_{n,i}}{dg_l} + \beta_n \left[h_{l-i+1} - \sum_j h_{j-i+1} \sum_k h_{j-k+1} \frac{d\hat{f}_{n,k}}{dg_l} \right] \ .$$

Then, multiplying both sides by u_l and taking the summation over l you have

$$\sum_l \frac{d\hat{f}_{n+1,i}}{dg_l} u_l = \sum_l \frac{d\hat{f}_{n,i}}{dg_l} u_l + \beta_n \left[\sum_l h_{l-i+1} u_l - \sum_j h_{j-i+1} \sum_k h_{j-k+1} \sum_l \frac{d\hat{f}_{n,k}}{dg_l} \right] \ .$$

Finally, making the substitution for the definition of $\hat{w}_n$, we get

$$\hat{w}_{n+1,i} = \hat{w}_{n,i} + \beta_n[H^T(u - H\hat{w}_n)]_i \ ,$$

which is the steepest-descent algorithm operating on u, but using the same β_n that was used for g.

5. RGCV for Richardson-Lucy

We can follow the same derivation for the Richardson-Lucy algorithm. The basic Richardson-Lucy iteration is

$$\hat{f}_{n+1,i} = \hat{f}_{n,i} \sum_j \frac{h_{j-i+1} g_j}{\sum_k h_{j-k+1} \hat{f}_{n,k}} .$$

Taking the derivative with respect to a particular pixel in the observed image, g_l, we have

$$\frac{d\hat{f}_{n+1,i}}{dg_l} = \frac{d\hat{f}_{n,i}}{dg_l} \sum_j \frac{h_{j-i+1} g_j}{\sum_k h_{j-k+1} \hat{f}_{n,k}} + \hat{f}_{n,i} \left[\frac{h_{l-i+1}}{\sum_k h_{l-k+1} \hat{f}_{n,k}} - \sum_j \frac{(h_{j-i+1} g_j) \left(\sum_m h_{j-m+1} \frac{d\hat{f}_{n,m}}{dg_l} \right)}{\left(\sum_k h_{j-k+1} \hat{f}_{n,k} \right)^2} \right] .$$

Then, multiplying by u_l and summing over l,

$$\sum_l \frac{d\hat{f}_{n+1,i}}{dg_l} u_l = \sum_l \frac{d\hat{f}_{n,i}}{dg_l} u_l \sum_j \frac{h_{j-i+1} g_j}{\sum_k h_{j-k+1} \hat{f}_{n,k}} + \hat{f}_{n,i} \left[\sum_l \frac{h_{l-i+1} u_l}{\sum_k h_{l-k+1} \hat{f}_{n,k}} - \sum_j \frac{(h_{j-i+1} g_j) \left(\sum_m h_{j-m+1} \sum_l \frac{d\hat{f}_{n,m}}{dg_l} u_l \right)}{\left(\sum_k h_{j-k+1} \hat{f}_{n,k} \right)^2} \right] .$$

Using the definition of $\hat{w}_n$, we get

$$\hat{w}_{n+1,i} = \hat{w}_{n,i} \left[H^T \left[\frac{g}{H \hat{f}_n} \right] \right]_i + \hat{f}_{n,i} \left[H^T \left[\frac{u}{H \hat{f}_n} \right] - H^T \left[\frac{[g][H \hat{w}_n]}{\left[H \hat{f}_n \right]^2} \right] \right]_i .$$

6. Using the RGCV Criterion as a Stopping Rule

The RGCV stopping rule canbe summarized by the following steps:

$n = 0$
do until $V(n) > V(n-1)$
{
    Increment n.
    Compute $\hat{f}_n$.
    Compute $\hat{w}_n$.
    Compute $V(n)$:
}

7. Experiments

Two experiments were performed to demonstrate the RCGV stopping rule. The first was done with artificial data, and the second with HST data.

For the first experiment, an image was created that simulates a scene that might be typical for the wide-field planetary camera (see Fig. 1). The image was then blurred using various "simple" blurs, and Poisson noise was added using an ADU gain of 8.0 e-/DN. Read-out noise was neglected

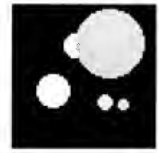


Figure 1. The simulated image used for the first experiment. (64 by 64 pixels)

to simplify the problem. The R-L algorithm was performed, and the RGCV stopping rule was computed. The mean-square error between the estimate and the original was also computed at each iteration. Finally, the χ–square measure was computed for further comparison. The results are summarized in Table 1.

Table 1. Summary of stopping rules for simulated image blurred with "simple" blurs.

		MMSE		χ–square		RGCV
Blur Type	n	MSE	n	MSE	n	MSE
7 point uniform horiz.	26	74.5	17	90.7	44	89.3
9 point uniform horiz.	31	87.8	22	100.3	35	89.0
3 by 3 point flat	30	67.4	13	97.7	24	69.4
11 by 11 point flat	156	558.5	81	599	147	558.9

The second group of experiments used the HST data for images of Jupiter, which are part of the simulation data available from the Space Telescope Science Institute database. The R-L algorithm was performed with the RGCV stopping rule. The results are shown in Fig. 2.

8. Conclusions

The application of the RGCV criterion as a stopping rule for non-linear iterative algorithms appears to have both analytic and experimental justification. The comparison to the χ–square criterion shows that the RGCV criterion more closely approximates the minimum mean-square error stopping point, and does so without any prior knowledge of the noise present in the system.

Unfortunately, there is one major drawback to the RGCV criterion, namely computation. The normal R-L iteration involves two convolutions. The computation of $\hat{w}_n$ uses previously computed convolutions and adds an additional three convolutions. Thus, by using the RGCV criterion, one increases the computation time by approximately a factor of 2.5. The advantage of the RGCV that may counter this disadvantage is the fact that the RGCV stopping rule is autonomous.

9. Caveat

The authors wish to warn the readers that while the analytical justification of the RGCV method appears to be valid, and several examples (including real HST data) demonstrate its ability, there have been cases in simulated data sets where the RGCV criterion failed to perform as a stopping rule. In other words, for some images, the RGCV criterion continues to decrease indefinitely, even after the noise amplification has completely destroyed the restoration. Other authors have noticed similar problems with other GCV based measurements (Thompson et al. 1989).

The authors have noticed this problem in several scenarios, and have tried to qualify the situations where it may occur. The factors that the authors have witnessed as contributing to the problem are a complex point-spread functions (PSF) and a low ratio of data points to PSF points.

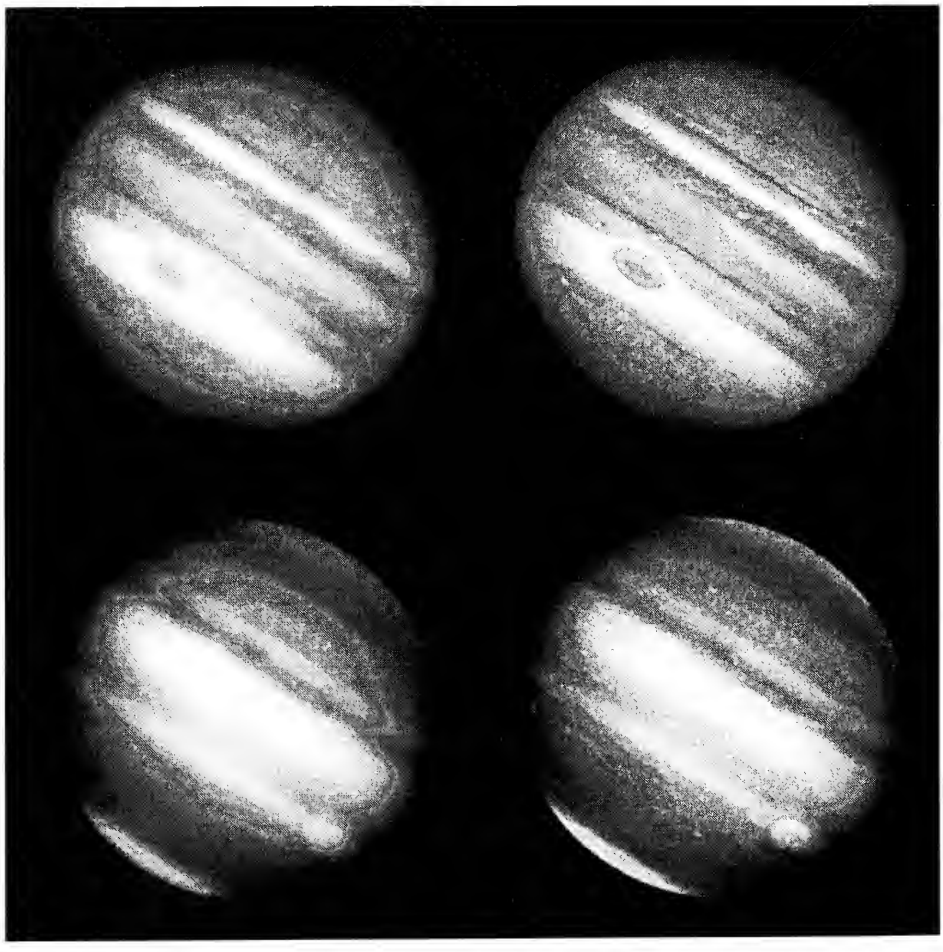


Figure 2. Summary of RGCV stopping rule on HST data. Upper left: The recorded image of Jupiter using the F413M filter (data set W0VI0207T), and with the bad data pixels adjusted. Upper right: The restored image after 42 iterations (as chosen by the RGCV criterion). Lower left: The recorded image of Jupiter using the F889N filter (data set W0VI0106T), and with the bad data pixels adjusted. Lower right: The restored image after 40 iterations (as chosen by the RGCV criterion).

References

Biemond, J., Lagendijk, R. L., & Mersereau, R. M. 1990, Proc. of the IEEE, 78, 856

Dempster, A. P., Laird, N. M., & Rubin, D. B. 1977, J. R. Stat. Soc. B, 39, 1

Girard, D. A. 1989, Numerische Mathematik, 31, 377

Golub, G. H., Heath, M., & Wahba, G. 1979, Technometrics, 21, 215

Hutchinson, M. F. 1990, Communications in Statistics — Simulation and Computation, 19 (2), 433

Lucy, L. B. 1974, AJ, 79, 745

Reeves, S. J., & Mersereau, R. M. 1991, SPIE Vol. 1452 — Image Processing Algorithms and Techniques II, 127

Reeves, S. J., & Perry, K. M. 1992, SPIE Vol. 1657 — Image Processing Algorithms and Techniques III, 192

Richardson, W. H. 1972, J. Opt. Soc. Am., 62, 55

Shepp, L. A., & Vardi, Y. 1982, IEEE Trans. Med. Imaging, MI-1, 113

Thompson, A. M., Kay, J. W., & Titterington, D. M. 1989, J. Stat. Comp. and Sim., Nov. 1989, 199

The Restoration of HST Images and Spectra II
Space Telescope Science Institute, 1994
R. J. Hanisch and R. L. White, eds.

Image Restoration Using the Damped Richardson-Lucy Method

Richard L. White

Space Telescope Science Institute, 3700 San Martin Drive, Baltimore, MD 21218

Abstract. A modification of the Richardson-Lucy iteration that reduces noise amplification in restored images is described.

1. Introduction

The Richardson-Lucy (R-L) algorithm (Richardson 1972, Lucy 1974) is the technique most widely used for restoring HST images. The standard R-L method has a number of characteristics that make it well-suited to HST data:

- The R-L iteration converges to the maximum likelihood solution for Poisson statistics in the data (Shepp and Vardi 1982), which is appropriate for optical data with noise from counting statistics.
- The R-L method forces the restored image to be non-negative and conserves flux both globally and locally at each iteration.
- The restored images are robust against small errors in the point-spread function (PSF).
- Typical R-L restorations require a manageable amount of computer time.

The R-L iteration can be derived very simply if we start with the imaging equation and the equation for Poisson statistics. The imaging equation tells how the true image is blurred by the PSF:

$$I(i) = \sum_j P(i|j)\, O(j)\,, \tag{1}$$

where O is the unblurred object, $P(i|j)$ is the PSF (the fraction of light coming from true location j that gets scattered into observed pixel i), and I is the noiseless blurry image. If the PSF does not vary with position in the image, then $P(i|j) = P(i-j)$ and the sum becomes a convolution. The probability of getting N counts in a pixel when the mean expected number of counts is $\bar{N}$ is given by the Poisson distribution:

$$\mathcal{P}(N|\bar{N}) = \frac{e^{-\bar{N}}\bar{N}^N}{N!} \tag{2}$$

and so the joint likelihood $\mathcal{L}$ of getting the observed counts $D(i)$ in each pixel given the expected counts $I(i)$ is

$$\ln \mathcal{L} = \sum_i D(i) \ln I(i) - I(i) - \ln D(i)!\,. \tag{3}$$

The maximum likelihood solution occurs where all partial derivatives of $\mathcal{L}$ with respect to $O(j)$ are zero:

$$\begin{aligned} \frac{\partial \ln \mathcal{L}}{\partial O(j)} &= 0 \\ &= \sum_i \left[\frac{D(i)}{I(i)} - 1\right] P(i|j)\,. \end{aligned} \tag{4}$$

The R-L iteration is simply

$$O_{new}(j) = O(j)\frac{\sum_i P(i|j)\frac{D(i)}{I(i)}}{\sum_i P(i|j)} \,. \tag{5}$$

It is clear from a comparison of Eqs. (4) and (5) that if the R-L iteration converges (as has been proven by Shepp & Vardi 1982), meaning that the correction factor approaches unity as the iterations proceed, then it must indeed converge to the maximum likelihood solution for Poisson statistics in the data.

2. Noise Amplification

Despite its advantages, the R-L method has some serious shortcomings. In particular, noise amplification can be a problem. This is a generic problem for all maximum likelihood techniques, which attempt to fit the data as closely as possible. If one performs many R-L iterations on an image containing an extended object such as a galaxy, the extended emission usually develops a "speckled" appearance (Fig. 1). The speckles are not representative of any real structure in the image, but are instead the result of fitting the noise in the data too closely. In order to reproduce a small noise bump in the data it is necessary for the unblurred image to have a very large noise spike; pixels near the bright spike must then be very black (near zero brightness) in order to conserve flux.

The only limit on the amount of noise amplification in the R-L method is the requirement that the image not become negative. Thus, once the compensating holes in the image are pushed down to zero flux, nearby spikes cannot grow any further and noise amplification ceases. The positivity constraint alone is sufficient to control noise amplification in images of star fields on a black background; in that case one can perform thousands of R-L iterations without generating an unacceptable amount of noise. However, for smooth objects observed at low signal-to-noise, even a modest number of R-L iterations (20–30) can produce objectionable noise.

The usual practical approach to limiting noise amplification is simply to stop the iteration when the restored image appears to become too noisy. However, the question of where to stop is a difficult one. The approach suggested by Lucy (1974) was to stop when the reduced χ^2 between the data and the blurred model is about 1 per degree of freedom. Unfortunately, one does not really know how many degrees of freedom have been used to fit the data. If one stops after a very few iterations then the model is still very smooth and the resulting χ^2 should be comparable to the number of pixels. If one performs many iterations, however, then the model image develops a great deal of structure and so the effective number of degrees of freedom used is large; in that case, the fit to the data ought to be considerably better. There is no criterion for the R-L method that tells how close the fit ought to be. Note that there is such a criterion built in to the MEMSYS 5 maximum entropy package (Gull and Skilling 1991), and the pixon approach of Piña and Puetter (1993) uses similar ideas.

Another problem is that the answer to the question of how many iterations to perform often is different for different parts of the image. It may require hundreds of iterations to get a good fit to the high signal-to-noise image of a bright star, while a smooth, extended object may be fitted well after only a few iterations. In Fig. 1, note how the images of both the central star and the bright star at the top center continue to improve as the number of iterations increases, while the noise amplification in the extended nebulosity is getting much worse. Thus, one would like to be able to slow or stop the iteration automatically in regions where a smooth model fits the data adequately, while continuing to iterate in regions where there are sharp features (edges or point sources).

Another approach to controlling noise amplification is to smooth the final restored image. This method has been developed and mathematically justified by Snyder and his co-workers (Snyder and Miller 1985, Snyder et al. 1987). Unfortunately, for HST images the amount of smoothing required to reduce the noise amplification is very large. Fig. 2 shows the effect of various amounts

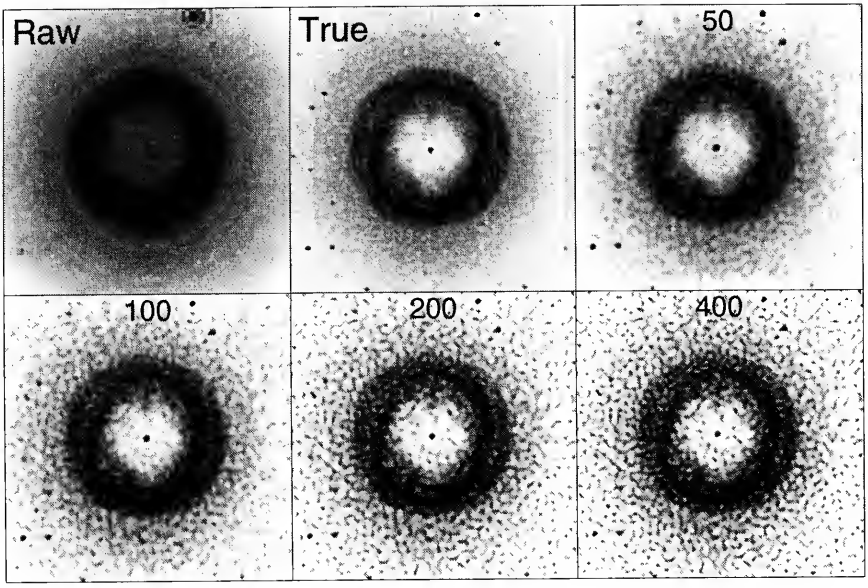


Figure 1. R-L restoration of simulated 85×85 pixel HST PC observation of a planetary nebula. Restored and true images are 256×256 (method of White 1990 was used to restore image with finer pixels than data.) As the number of iterations (shown at top of each image) increases, the images of bright stars improve, but noise is amplified unacceptably in the nebulosity.

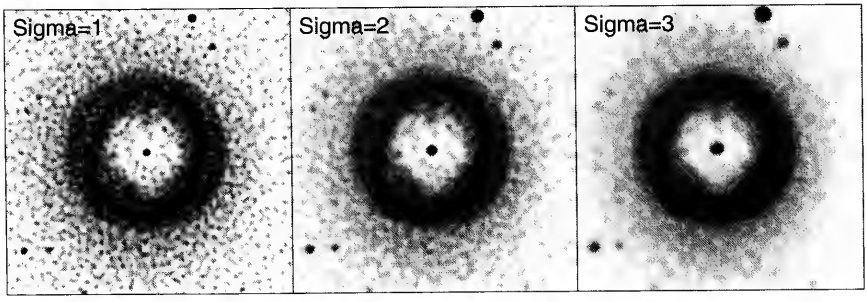


Figure 2. Results of smoothing 400-iteration R-L restoration from Fig. 1 with a Gaussian. The width of the Gaussian is σ = 1, 2, and 3 pixels for the three cases shown. Images of stars are grossly blurred when the R-L image is smoothed heavily enough to substantially reduce the noise in the nebulosity.

of smoothing on the restored planetary nebula image. By the time the noise amplification in the nebulosity has been reduced to a visually acceptable level, the images of stars have been grossly blurred. For most purposes, one must pay far too high a price to avoid noise amplification using this method.

3. The Damped R-L Iteration

An effective approach to such an adaptive stopping criterion is to modify the likelihood function of Eq. (3) so that it becomes flatter in the vicinity of a good fit. The approach I have taken is to use a likelihood function that is identical to Eq. (3) when the difference between the blurred model I and the data D is large compared with the noise, but that is essentially constant when the difference is smaller than the noise.

There are two important advantages to using a modified form of the likelihood function to control noise amplification:

- The end point of the iteration is well-defined, allowing properties of the final solution to be studied.
- Acceleration techniques may be used to reach the solution more quickly without changing the final answer.

It is appropriate to prevent noise amplification by recognizing its effects in the data domain rather than by applying regularizing constraints on the restored image. Residuals that are smaller than one expects based on the known noise properties of the data should be avoided by using a likelihood function that does not reward overfitting the data. Of course, it may be necessary to use constraints on the model image in order to select one of the many possible restored images that fit the data adequately well, but in my opinion it is important to first establish better goodness-of-fit criteria so that the set of possible restored images is as small as possible before applying regularization to select one of those images. The work of Piña & Puetter (1992) on the Maximum Residual Likelihood criterion for data fitting is a notable attempt in this direction.

The *damped R-L iteration* starts from the likelihood function

$$\ln \mathcal{L} = \sum_i f(U_i) \,, \tag{6}$$

where

$$U_i = -\frac{2}{T^2}\left[D(i)\ln\frac{I(i)}{D(i)} - I(i) + D(i)\right] \tag{7}$$

and

$$\begin{aligned} f(x) &= \tfrac{N-1}{N+1}\left(1 - x^{N+1}\right) + x^N \,, & x < 1 \\ &= x \,, & x \geq 1 \,. \end{aligned} \tag{8}$$

$f(x)$ is the "damping function". It is chosen to be a simple function that is linearly proportional to x for $x > 1$, is approximately constant for $x \sim 0$, and has continuous first and second derivatives at $x = 1$ (this allows acceleration techniques to be used to speed convergence of the method). The constant N determines how suddenly the function f becomes flat for $x < 1$. For $N = 0$, $f(x) = x$ and there is no flattening at all. The larger the value of N, the flatter is the function. The results reported in this paper use $N = 10$, but the value of N has little effect on the results as long as it is larger than a few.

U_i is a slightly modified version of the $\ln \mathcal{L}$ function of Eq. (3). The constants and multiplicative factors are chosen so that the expected value of U in the presence of Poisson noise is unity if the threshold $T = 1$. The threshold T then determines at what level the damping turns on: if $T = 1$ the damping occurs at 1σ, if $T = 2$ at 2σ, etc.

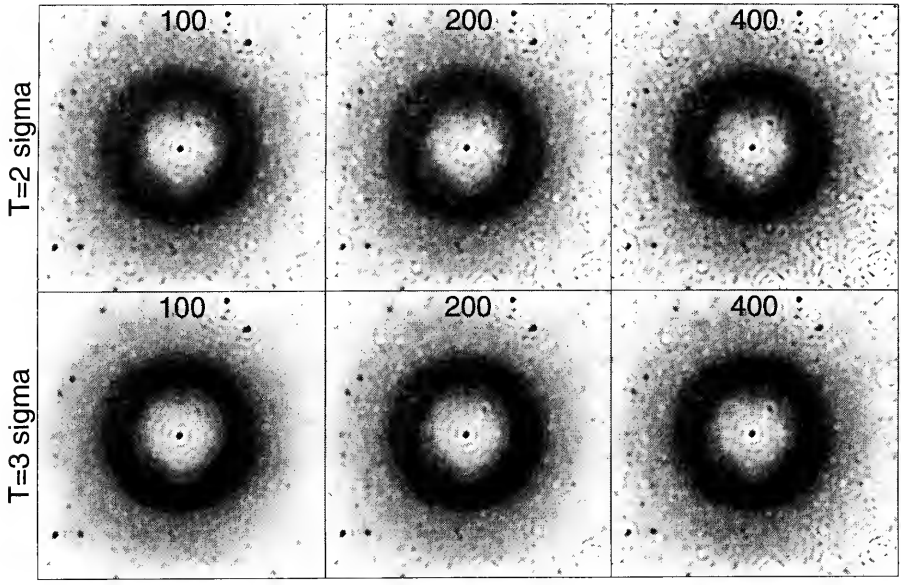


Figure 3. Restoration of data from Fig. 1 using the damped iteration with noise thresholds of 2 and 3σ. As number of iterations increases, images of bright stars continue to improve, but noise amplification is much better controlled than for standard R-L iteration.

With this new likelihood function, we can follow the steps outlined above to derive the damped R-L iteration:

$$O_{new}(j) = O(j)\frac{\sum_i P(i|j)\left[1 + \tilde{U}_i^{N-1}[N-(N-1)\tilde{U}_i]\dfrac{D(i)-I(i)}{I(i)}\right]}{\sum_i P(i|j)}\,, \tag{9}$$

where

$$\tilde{U}_i = \min(U_i, 1)\,. \tag{10}$$

Note that in regions where the data and model do *not* agree, $\tilde{U} = 1$ and so this iteration is exactly the same as the standard R-L iteration. In regions where the data and model do agree, however, the second term gets multiplied by a factor which is less than 1, and the ratio of the numerator and denominator approaches unity. This has exactly the desired character: it damps changes in the model in regions where the differences are small compared with the noise.

Fig. 3 shows the results of applying the damped iteration to the planetary nebula data of Fig. 1. Note that the noise amplification in the nebulosity has been greatly reduced, but that the stars are still very sharp. Fig. 4 shows the result of applying both the standard R-L method and the damped iteration to HST observations of Saturn; again, the noise in the restored image is greatly reduced without significantly blurring the sharp edges of the planet and its rings. This is perhaps more easily seen in Fig. 5, which shows the brightness of the restored image along a cut through the planet and rings. Sharp features such as narrow gaps in the rings are essentially identical in the R-L and damped images, but the disk of the planet is much smoother in the damped image. Note that the

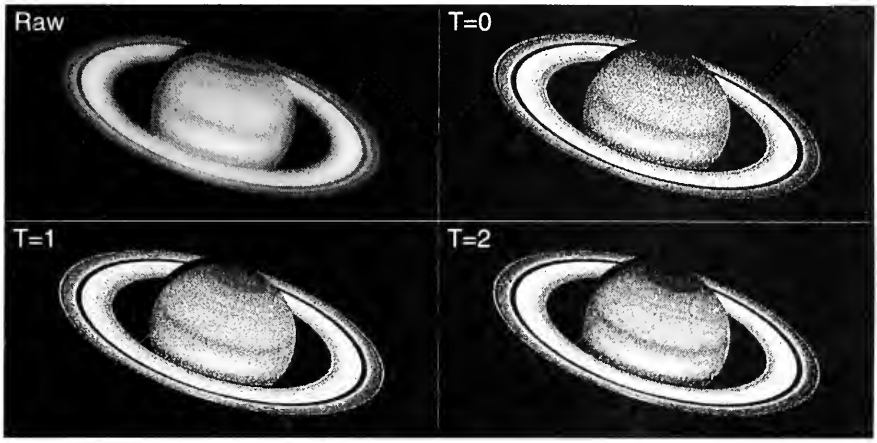


Figure 4. HST observations of Saturn and restorations using the R-L iteration ($T = 0$) and the damped iteration with noise thresholds of 1 and 2σ. 100 accelerated iterations were used in each case.

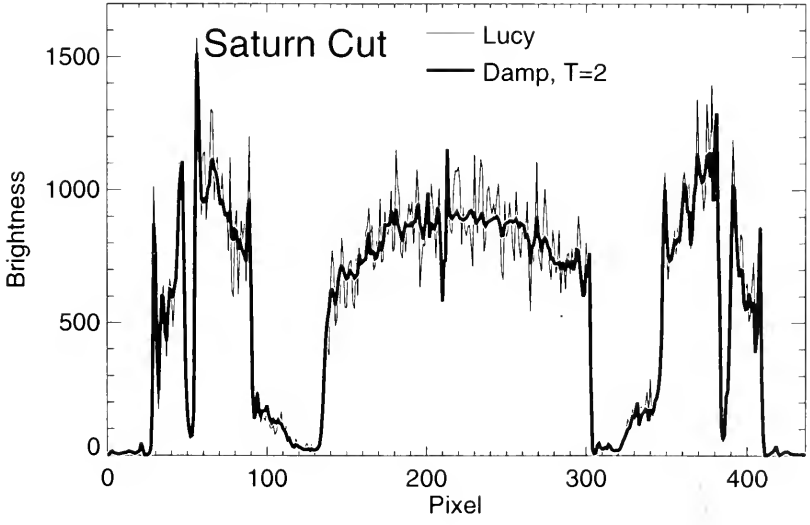


Figure 5. Cut through the rings and disk from upper right to lower left for the restored Saturn images in Fig. 4.

mean brightness in the disk of the planet is the same in the damped and R-L images, indicating good photometric linearity.

4. Summary

This paper describes an iterative image restoration technique that closely resembles the R-L iteration but that avoids the amplification of noise that occurs in the R-L iteration. I call this the "damped" R-L iteration because the modification appears as a damping factor that slows changes in regions of the model image that fit the data well while allowing the model to continue to improve in regions where it fits the data less well.

The approach used is to modify the likelihood function so that it is "flatter" in the vicinity of a good fit. It is unlikely that this ad hoc approach represents the best solution to the noise amplification problem, but it does have the advantages that it is easy to implement and robust, and it appears to produce better results than the R-L method in many cases. The method produces restored images that have good photometric linearity and little bias.

References

Gull, S. F., & Skilling, J. 1991, *Quantified Maximum Entropy MemSys5 User's Manual*, Maximum Entropy Data Consultants, Ltd., Royston

Lucy, L. B. 1974, AJ, 79, 745

Piña, R. K., & Puetter, R. C. 1992, PASP, 104, 1096

Piña, R. K., & Puetter, R. C. 1993, PASP, 105, 630

Richardson, W. H. 1972, J. Opt. Soc. Am., 62, 55

Shepp, L. A., & Vardi, Y. 1982, IEEE Trans. Med. Imaging, MI–1, 113

Snyder, D. L. & Miller, M. I. 1985, IEEE Trans. Nuclear Science, NS–32, 3864

Snyder, D. L., Miller, M. I., Thomas, L. J., Jr., & Politte, D. G. 1987, IEEE Trans. Med. Imaging, MI–6, 228

White, R. L. 1990, in *The Restoration of HST Images and Spectra*, R. L. White & R. J. Allen, eds., Space Telescope Science Institute, Baltimore, 139

The Restoration of HST Images and Spectra II
Space Telescope Science Institute, 1994
R. J. Hanisch and R. L. White, eds.

Image Restoration with Denoising Using Multi-Resolution

Jean-Luc Starck

Cisi-Ingenierie, 1360 route des Dolines 06560 Valbonne, and Observatoire de la Côte d'Azur, B.P. 229 F-06304 Nice Cedex 4, France

Fionn Murtagh[1]

Space Telescope – European Coordinating Facility, European Southern Observatory, Karl-Schwarzschild-Strasse 2, D-85748 Garching bei München, Germany

Albert Bijaoui

Observatoire de la Côte d'Azur, B.P. 229 F-06304 Nice Cedex 4, France

Abstract. This paper shows how an effective noise suppression strategy can be incorporated into algorithms for the solution of the inverse problem. The residual in the fit of the restored image, at each iteration, is analyzed using a wavelet transform. In order to suppress noise, only significant wavelet coefficients are retained. The effectiveness of this procedure is demonstrated.

1. Introduction

Consider an image characterized by its intensity distribution $I(x, y)$, corresponding to the observation of an object $O(x, y)$ through an optical system. If the imaging system is linear and shift-invariant, the relation between the object and the image in the same coordinate frame is a convolution:

$$I(x, y) = O(x, y) * P(x, y) + N(x, y) \,. \tag{1}$$

$P(x, y)$ is the point spread function (PSF) of the imaging system, and $N(x, y)$ is additive noise. We want to determine $O(x, y)$ knowing $I(x, y)$ and $P(x, y)$. This inverse problem has led to a large amount of work, the main difficulties being the existence of: (i) a cut-off frequency of the PSF, and (ii) an intensity noise (see for example Cornwell 1988). Eq. 1 is always an ill-posed problem. This means that there is no unique least-squares solution of minimal norm $\| I(x, y) - P(x, y) * O(x, y) \|^2$, and a regularization is necessary.

The best restoration algorithms are generally iterative (Katsaggelos 1991). Van Cittert (1931) proposed the following iteration:

$$O^{(n+1)}(x, y) = O^{(n)}(x, y) + \alpha(I(x, y) - P(x, y) * O^{(n)}(x, y)) \tag{2}$$

where α is a convergence parameter generally taken as 1. In this equation, the object distribution is modified by adding a term proportional to the residual. Another iterative algorithm is provided by the minimization of the norm $\| I(x, y) - P(x, y) * O(x, y) \|^2$ (Landweber 1951) and leads to:

$$O^{(n+1)}(x, y) = O^{(n)}(x, y) + \alpha P_s(x, y) * [I(x, y) - P(x, y) * O^{(n)}(x, y)] \tag{3}$$

[1] Affiliated to Astrophysics Division, Space Science Department, European Space Agency

where $P_s(x,y) = P(-x,-y)$. The Richardson-Lucy method (Lucy 1974, Richardson 1972) uses an iterative approach to compute a maximum likelihood estimate:

$$O^{(n+1)} = O^{(n)} \left[\frac{I}{I^{(n)}} * P^* \right] \tag{4}$$

and

$$I^{(n)} = P * O^{(n)} \tag{5}$$

where P^* is the transpose of the PSF.

2. The Wavelet Transform

Extensive literature exists on the wavelet transform and its applications (Meyer 1991, Daubechies 1988, Chui 1992, Meyer 1990, Meyer 1992, Ruskai et al. 1992). In this paper, our discrete approach of the wavelet transform is implemented with a version of the so-called *à trous* (with holes) algorithm (Holschneider and Tchamitchian 1990, Shenasa 1992, Starck and Bijaoui 1993). One assumes that the sampled data $\{c_0(k)\}$ are the scalar products at pixels k of the function $f(x)$ with a scaling function $\phi(x)$ which corresponds to a low pass filter:

$$c_0(k) =< f(x), \phi(x-k) > \ . \tag{6}$$

The scaling function satisfies the dilation equation,

$$\frac{1}{2}\phi(\frac{x}{2}) = \sum_l h(l)\phi(x-l)\ , \tag{7}$$

with h being a discrete low pass filter. Consequently the scalar product $c_i(k) = \frac{1}{2^i} < f(x), \phi(\frac{x-k}{2^i}) >$ is obtained by:

$$c_i(k) = \sum_l h(l) c_{i-1}(k + 2^{i-1} l) \tag{8}$$

and the signal difference w_i by

$$w_i(k) = c_{i-1}(k) - c_i(k) \tag{9}$$

or

$$w_i(k) = \frac{1}{2^i} < f(x), \psi(\frac{x-k}{2^i}) > \tag{10}$$

when the wavelet function ψ is defined by

$$\frac{1}{2}\psi(\frac{x}{2}) = \phi(x) - \frac{1}{2}\phi(\frac{x}{2})\ . \tag{11}$$

The B-spline of degree 3 was used in our calculations. A reconstruction algorithm uses the last smoothed array $c_{n_p}(x,y)$ added to all the differences w_i:

$$c_0(k) = c_{n_p} + \sum_{i=1}^{n_p} w_i(k)\ . \tag{12}$$

The above *à trous* algorithm is easily extendible to two-dimensional space. This leads to a convolution with a mask of 5×5 pixels for the wavelet related to a B_3-spline. At each scale i we obtain a set $\{w_i(k,l)\}$ (we will call this a "wavelet plane" throughout the following discussion), which has the same number of pixels as the image.

2.1. Regularization using Significant Structures

Gaussian Noise If we use an iterative deconvolution algorithm such as Van Cittert or Richardson-Lucy, we define $R^{(n)}(x, y)$, the error at iteration n:

$$R^{(n)}(x, y) = I(x, y) - P(x, y) * O^{(n)}(x, y) . \tag{13}$$

By using the *à trous* wavelet transform algorithm, $R^{(n)}$ can be defined by the sum of its n_p wavelet planes and the last smooth plane (see Eq. 12):

$$R^{(n)}(x, y) = c_{n_p}(x, y) + \sum_{j=1}^{n_p} w_j(x, y) . \tag{14}$$

The wavelet coefficients provide a mechanism to extract from the residuals at each iteration only the significant structures. A large part of these residuals are generally statistically insignificant. The significant residual is

$$\bar{R}^{(n)}(x, y) = c_{n_p}(x, y) + \sum_{j=1}^{n_p} \alpha(w_j(x, y), N_j)\, w_i(x, y) . \tag{15}$$

N_j is the standard deviation of the noise at scale j, and α is a function which is defined by

$$\alpha(a, \sigma) = \begin{cases} 1 & \text{if } |a| \geq k\sigma \\ 0 & \text{if } |a| < k\sigma \end{cases} . \tag{16}$$

The standard deviation of the noise N_j is estimated from the standard deviation of the noise in the image. This is done from the study of noise variation in the wavelet space, with the hypothesis of a white Gaussian noise.

Poissonian Noise If the noise in the data I is Poissonian, the transform

$$T(I(x, y)) = 2\sqrt{I(x, y) + \frac{3}{8}} \tag{17}$$

acts as if the data arose from the Gaussian white noise model (Donoho 1992) with $\sigma = 1$. The noise in $S^{(n)} = T(I) - T(I^{(n)})$ (with $I^{(n)} = P * O^{(n)}$) can be suppressed in the same way as previously. $S^{(n)}$ is decomposed into wavelet coefficients, and only the significant coefficients are kept. The reconstruction gives $\bar{S}^{(n)}$. The residual $R^{(n)}$ can be expressed as a function of $S^{(n)}$ (see Appendix) by

$$R^{(n)}(x, y) = S^{(n)}(x, y)\left[\frac{S^{(n)}(x, y)}{4}\right) + \sqrt{I^{(n)}(x, y) + \frac{3}{8}}\Big] \tag{18}$$

and $\bar{R}^{(n)}$ is obtained by

$$\bar{R}^{(n)}(x, y) = \bar{S}^{(n)}(x, y)\left[\frac{\bar{S}^{(n)}(x, y)}{4} + \sqrt{I_n(x, y) + \frac{3}{8}}\right] . \tag{19}$$

We show now how the iterative deconvolution algorithms can be modified in order to take into account only the significant structure at each scale.

Regularization of Van Cittert's Algorithm Van Cittert's (1931) iteration is

$$O^{(n+1)}(x,y) = O^{(n)}(x,y) + \alpha R^{(n)}(x,y) \tag{20}$$

with $R^{(n)}(x,y) = I(x,y) - P(x,y) * O^{(n)}(x,y)$. Regularization using significant structures leads to:

$$O^{(n+1)}(x,y) = O^{(n)}(x,y) + \alpha \bar{R}^{(n)}(x,y)\,. \tag{21}$$

The basic idea of our method consists of detecting, at each scale, structures of a given size in the residual $R^{(n)}(x,y)$ and putting them in the restored image $O^{(n)}(x,y)$. The process finishes when no more structures are detected. Then, we have separated the image $I(x,y)$ into two images $\tilde{O}(x,y)$ and $R(x,y)$. $\tilde{O}$ is the restored image, which does not contain any noise, and $R(x,y)$ is the final residual which does not contain any structure. R is our estimation of the noise $N(x,y)$.

Regularization of the One-Step Gradient Method The one-step gradient iteration is

$$O^{(n+1)}(x,y) = O^{(n)}(x,y) + P(-x,-y) * R^{(n)}(x,y) \tag{22}$$

with $R^{(n)}(x,y) = I(x,y) - P(x,y) * O^{(n)}(x,y)$. Regularization by significant structures leads to

$$O^{(n+1)}(x,y) = O^{(n)}(x,y) + P(-x,-y) * \bar{R}^{(n)}(x,y)\,. \tag{23}$$

Regularization of the Richardson-Lucy Algorithm Define $I^{(n)}(x,y) = P(x,y) * O^{(n)}(x,y)$. Then $R^{(n)}(x,y) = I(x,y) - I^{(n)}(x,y)$, and hence $I(x,y) = I^{(n)}(x,y) + R^{(n)}(x,y)$. The Richardson-Lucy equation is

$$O^{(n+1)}(x,y) = O^{(n)}(x,y)\left[\frac{I^{(n)}(x,y) + R^{(n)}(x,y)}{I^{(n)}(x,y)} * P(-x,-y)\right] \tag{24}$$

and regularization leads to

$$O^{(n+1)}(x,y) = O^{(n)}(x,y)\left[\frac{I^{(n)}(x,y) + \bar{R}^{(n)}(x,y)}{I^{(n)}(x,y)} * P(-x,-y)\right]\,. \tag{25}$$

Convergence The standard deviation of the residual decreases until no more significant structures are found. Convergence can be estimated from the residual. The algorithm stops when

$$\frac{\sigma_{R^{(n-1)}} - \sigma_{R^{(n)}}}{\sigma_{R^{(n)}}} < \epsilon\,. \tag{26}$$

3. Example

As an example of typical results, Fig. 1 shows a simulated HST Wide Field and Planetary Camera (WF/PC I) image of a distant cluster of galaxies. Its construction was described in Freudling and Caulet (1993). The upper left is the "ground truth" image, the upper right the image degraded through convolving with a spherically aberrated PSF and adding noise, the lower left is a restoration with noise removal, and the lower right is a Richardson-Lucy restoration. These restorations used, respectively, 28 and 40 unaccelerated iterations using a Gaussian noise model. Various tests were carried out on the object detection potential of the two restored images, with a conclusion pointing narrowly in favor of the noise-suppressed restoration. Further details are to be found in Starck and Murtagh (1993).

Another area where the strategy described in this paper works well is when the image background varies. Richardson-Lucy, without intervention in wavelet space, must first model the background and remove it. It appears that the wavelet-based approach handles such a situation well, and without such removal of the background, due to the fact that it detects significant structures in differenced (smoothed) images. An example of this may be found in Starck and Murtagh (1993).

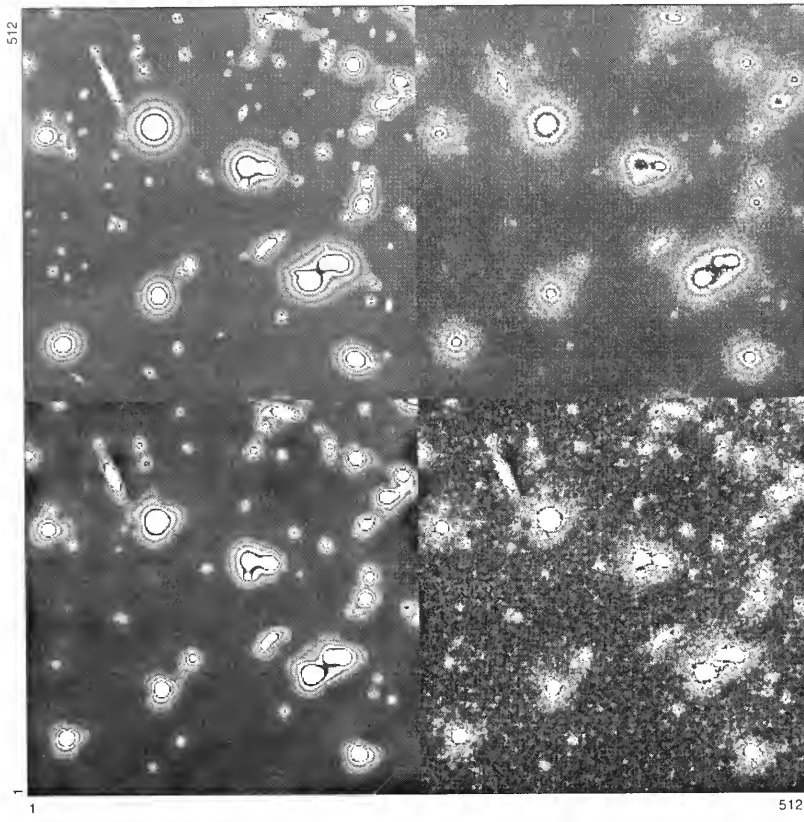


Figure 1. Upper left: "ground truth" image; upper right: the image degraded through convolving with a spherically aberrated PSF and adding noise; lower left: restoration with noise removal; lower right: Richardson-Lucy restoration.

4. Conclusion

An adaptation of widely-used image restoration methods has been presented, which uses a wavelet transform to suppress noise. The examples carried out on a range of images show the powerfulness of this approach. Unlike in various other approaches, there is no regularization coefficient to be estimated; and the problem of appropriate result is addressed directly in terms of noise and not in terms of smoothness or other such constraints. Photometry tests, using the approach described here, did not necessarily show an improvement over Richardson-Lucy restorations. Hence we propose, as work to be carried out in the future, an assessment of the usefulness of "soft" or adaptive alternatives to the noise suppression strategy adopted here, which was based on a "hard" threshold.

Appendix

We have the following relations:

$$\begin{aligned} S^{(n)}(x,y) &= T(I(x,y)) - T(I^{(n)}(x,y)) \\ R^{(n)}(x,y) &= I(x,y) - I^{(n)}(x,y) \\ I^{(n)}(x,y) &= P(x,y) * O^{(n)}(x,y) \end{aligned}$$

and the transform T is defined by

$$T(I(x,y)) = 2\sqrt{I(x,y)+\frac{3}{8}}\,.$$

Then we have

$$\begin{aligned} [T(I(x,y))]^2 &= [S^{(n)}(x,y) + T(I^{(n)}(x,y))]^2 \\ &= (S^{(n)}(x,y))^2 + 4(I^{(n)}(x,y)+\frac{3}{8}) + 4S^{(n)}(x,y)\sqrt{I^{(n)}(x,y)+\frac{3}{8}} \end{aligned}$$

and

$$[T(I(x,y))]^2 = 4(I(x,y)+\frac{3}{8})\,.$$

We deduce from the two last equations that $I(x,y)$ can be expressed by

$$I(x,y) = \frac{1}{4}[(S^{(n)}(x,y))^2 + 4(I^{(n)}(x,y)+\frac{3}{8}) + 4S^{(n)}(x,y)\sqrt{I^{(n)}(x,y)+\frac{3}{8}}] - \frac{3}{8}\,.$$

By replacing I by its expression in $R^{(n)}(x,y) = I(x,y) - I^{(n)}(x,y)$, we have

$$R^{(n)}(x,y) = S^{(n)}(x,y)[\frac{S^{(n)}(x,y)}{4} + \sqrt{I^{(n)}(x,y)+\frac{3}{8}}]\,.$$

References

Chui, C. H. 1992, *Wavelet Analysis and its Application*, Academic Press, New York

Cornwell, T. J. 1988, in *Diffraction-Limited Imaging with Very Large Telescopes*, Proceedings, NATO Advanced Study Institute, Cargèse, Corsica, 273

Daubechies, I. 1988, Comm. Pure Appl. Math., 41, 909

Donoho, D. L. 1992, in *Progress in Wavelet Analysis and Applications*, éditions Frontières, Gif sur Yvette

Freudling, W. & Caulet, A. 1993, in *5th ESO/ST-ECF Data Analysis Workshop*, P. J. Grosbøl & R. C. E. de Ruijsscher, eds., European Southern Observatory, Garching, 63

Frieden, B. R. 1975, in *Topics in Applied Physics*, Springer-Verlag, Berlin, 177

Galatsanos, N. P., & Katsaggelos, A. K. 1992, IEEE Trans. Image Proc., July

Golub, G. H., Heath, M. & Wahba, G. 1979, Technometrics, 21, 215

Holschneider, M., & Tchamitchian, P. 1990, in *Les Ondelettes en 1989*, P. G. Lemarié, ed., Springer-Verlag, Berlin, 102

Katsaggelos, A. K., ed. 1991, *Digital Image Restoration*, Springer-Verlag, Berlin

Landweber, L. 1951, Am. J. Math., 73, 615

Lucy, L. B. 1974, AJ, 79, 745

Meyer, Y. 1990, *Ondelettes*, Hermann, ed., Paris

Meyer, Y. 1991, *Ondelettes et Paquets d'Ondes*, INRIA, Rocquencourt

Meyer, Y. 1992, *Ondelettes: Algorithmes et Applications*, Collection Acquis Avancés de l'Informatique

Richardson, W. H. 1972, J. Opt. Soc. Am., 62, 55

Ruskai, M. B., Beylkin, G., Coifman, R., Daubechies, I., Mallat, S., Meyer, Y., & Raphael, L. 1992, Wavelets and Their Applications, Jones and Barlett

Shensa, M. J. 1992, Proc. IEEE Trans. Sig. Proc., 40, 2464

Starck, J.-L., & Bijaoui, A. 1993, Sig. Proc., in press

Starck, J.-L., & Murtagh, F. 1993, A&A, in press

Van Cittert, P. H. 1931, Z. Physik, 69, 298

The Restoration of HST Images and Spectra II
Space Telescope Science Institute, 1994
R. J. Hanisch and R. L. White, eds.

Prior Models and the Richardson-Lucy Restoration Method

R. Molina, J. Mateos, and J. Abad

Department of Computer Science, Faculty of Science, University of Granada, 18071 Granada, Spain

Abstract. Following the Bayesian paradigm for image restoration we show how smoothness constraints can be incorporated into the R-L method. We also examine different noise models and study their approximation by Gaussian models that are robust to detector errors.

1. Introduction

Image restoration refers to the problem of recovering an image, $\mathbf{f}$, from its blurred and noisy observation, $\mathbf{g}$, for the purpose of improving its quality or obtaining some type of information that is not readily available from the degraded image. Blur is frequently due to the relative motion between the subject and the camera, atmospheric turbulence, out of focus lenses, and/or the image sensor. Film grain, electronic noise, and quantization are the major sources of noise in a digital image.

In this paper we describe how the Bayesian approach to image restoration can be used to incorporate prior information, in the form of smoothness constraints, to the R-L restoration method (Lucy 1974). We also examine the noise models and study how to remove detector errors.

The work is divided as follows. In §2 we describe briefly the Bayesian paradigm. In §3 we study the observation process and how to remove detector errors. §4 is devoted to the study of prior models that take into account smoothness constraints and also to the study of the prior model used in the R-L restoration method. The next step is to calculate the estimate of the real underlying image, $\mathbf{f}$; the algorithm implementing this task and its relationship to the R-L restoration method are described in §5. Finally, a test example is shown in §6.

2. Bayesian Paradigm

The philosophy within statistics known as Bayesian inference has a very long history. It is distinguished from the perhaps more familiar classical statistical ideas by using prior information about the images being studied.

Bayesian methods start with a prior distribution, a probability distribution over images $\mathbf{f}$, $p(\mathbf{f})$, (it is here that we incorporate information on the expected structure within an image), it is also necessary to specify the probability distribution $p(\mathbf{g}|\mathbf{f})$, of observed images $\mathbf{g}$ if $\mathbf{f}$ were the true image. The Bayesian paradigm dictates that inference about the true $\mathbf{f}$ should be based on $p(\mathbf{f}|\mathbf{g})$ given by

$$p(\mathbf{f}|\mathbf{g}) \propto p(\mathbf{f})p(\mathbf{g}|\mathbf{f}) \, .$$

To show just one restoration, it is common to choose the mode of this posterior distribution, that is to display the image $\hat{\mathbf{f}}$ which satisfies

$$\hat{\mathbf{f}} \text{ maximizes } p(\mathbf{f})p(\mathbf{g}|\mathbf{f}) \, .$$

This is known as the maximum a posteriori (MAP) estimate of $\mathbf{f}$.

Equivalently, we can choose $\hat{\mathbf{f}}$ to minimize

$$-\log p(\mathbf{g}|\mathbf{f}) - \log p(\mathbf{f}) \, . \tag{1}$$

The first term in (1) is the familiar log likelihood of $\mathbf{f}$. The second term can be thought of as a roughness penalty, as images $\mathbf{f}$ which do not correspond to our prior conceptions will be assigned a small $p(\mathbf{f})$ and hence a large penalty.

In statistical physics it is common to define probabilities by the energy U of a system, so that

$$p(\mathbf{f}) \propto \exp[-\beta U(\mathbf{f})] \tag{2}$$

where β is $1/kT$, T being the temperature and k Boltzmann's constant. If we adopt this notation, $\hat{\mathbf{f}}$ minimizes

$$-\log(\text{likelihood}) + \beta U(\mathbf{f}) \; . \tag{3}$$

We can recognize this as a Lagrangian form, so its solution is equivalent to solving

$$\max(\text{likelihood subject to energy}) \leq \text{constraint} \; ,$$

and to

$$\min(\text{energy subject to likelihood}) \geq \text{constraint} \; ,$$

which correspond to the regularization approach to image restoration.

Many other deconvolution principles fit into one of these forms; in particular — as we will see later — the R-L restoration method. Maximum entropy methods also fit into this framework (Molina et al. 1992a).

Having described the Bayesian paradigm let us move on to examine the two ingredients of this paradigm, the observation process, $p(\mathbf{g}|\mathbf{f})$, and the prior model or image model, $p(\mathbf{f})$.

3. Observation Process

The observed image $\mathbf{g}$ differs from the true brightness distribution $\mathbf{f}$ in having been blurred and encountering statistical noise in the recording process.

If $p = m \times n$ is the size of the image, the blurring process is described by $\mathbf{Df}$ where $\mathbf{f}$ is a $p \times 1$ vector and $\mathbf{D}$ is the $p \times p$ point spread matrix defining the systematic blur and assumed to be known.

Let us describe the noise models. For each component i, $i = 1, 2, \ldots, m \times n$ of the observed vector $\mathbf{g}$ we could use a Poissonian model, thus obtaining $\mathbf{g}_i \sim \mathcal{P}((\mathbf{Df})_i)$. This model can be approximated, at least for high brightness values, by the Gaussian distribution $\mathcal{N}((\mathbf{Df})_i, (\mathbf{Df})_i)$.

An alternative model would be to assume $\mathbf{g}_i = \lambda \mathbf{z}_i$ where $\mathbf{z}_i \sim \mathcal{P}(\mathbf{Df})_i$. For this model we could use the following Gaussian approximation:

$$\mathbf{g}_i \sim \mathcal{N}((\mathbf{D}(\lambda \mathbf{f}))_i, \lambda(\mathbf{D}(\lambda \mathbf{f}))_i) \; .$$

Substituting $\lambda \mathbf{f}$ by $\mathbf{f}$, we would have

$$\mathbf{g}_i \sim \mathcal{N}((\mathbf{Df})_i, \lambda(\mathbf{Df})_i) \; .$$

Finally, a model like

$$\mathbf{g}_i = \mathcal{P}((\mathbf{Df})_i) + \mathcal{N}(0, \sigma^2)$$

and the corresponding $\lambda \mathbf{z}_i + \mathcal{N}(0, \sigma^2)$ can be approximated by normal distributions having the form $\mathcal{N}((\mathbf{Df})_i, a + b(\mathbf{Df})_i)$ for appropriate constants a and b.

If we use the Gaussian approximation we have

$$-2 \log p(\mathbf{g}|\mathbf{f}) = \text{const} + \sum_i D_i^2 \tag{4}$$

with $D_i = [(\mathbf{g}_i - (\mathbf{Df}_i))/\sigma(\mathbf{Df}_i)]$, $\sigma((\mathbf{Df})_i)$ being the standard deviation, $\sqrt{a + b(\mathbf{Df})_i}$, in the noise model just described.

This Gaussian approximation allows the easy incorporation of robust statistics concepts to deal with detector errors (Molina and Ripley 1989). The idea is to down-weight observations which are far away from their means. Such values are given too much weight in (4). The squared term in D_i represents the number of standard deviations that $\mathbf{g}_i$ is away from its mean. In robust statistics $\sum_i D_i^2$ is replaced by $\rho(D_i)$ for a function ρ which penalizes extreme values less severely. A typical function ρ is the Huber's 'Proposal 2' function defined by

$$\rho(x) = \begin{cases} x^2 & \text{for} \quad |x| \leq c\,, \\ 2c|x| - c^2 & \text{for} \quad |x| > c\,. \end{cases}$$

This is quadratic in the center, but penalizes large deviations linearly rather than quadratically. Equivalently, observations D_i are down-weighted if $|D_i|$ exceeds 2. In practice c is chosen at about 2, which down-weights only those observations more than two standard deviations away from their means.

The noise model we will use in this paper is Poissonian; see Molina and Ripley (1989) for the use of robust Gaussian noise models.

4. Prior Models

Consider an image with no stars but only with regions of smoothly varying luminosity. We then expect $\mathbf{f}_i \geq 0$ and $\mathbf{f}$ to be spatially smooth. Probably the simplest probability models that can be used to model smoothness are spatial autoregressions (Ripley 1981).

The conditional autoregression (CAR) model is defined by

$$p(\mathbf{f}) \propto \exp\{-\frac{1}{2}\alpha\, \mathbf{f}^t(I - \phi N)\mathbf{f}\}$$

where α is the unknown hyper-parameter, matrix N is such that $N_{ij} = 1$ if cells i and j are spatial neighbors (pixels at distance one) and zero otherwise, and scalar ϕ is just less than 0.25. The term $\mathbf{f}^t(I - \phi N)\mathbf{f}$ represents in matrix notation the sum of squares of the values $\mathbf{f}_i$ minus ϕ times the sum of $\mathbf{f}_i\mathbf{f}_j$ for neighboring pixels i and j.

The parameters can be interpreted by the following expressions describing the conditional distribution

$$\begin{aligned} E(\mathbf{f}_i \mid \mathbf{f}_j,\ j \neq i) &= \phi \sum_{j \text{ nhbr } i} \mathbf{f}_j \\ \text{var}(\mathbf{f}_i \mid \mathbf{f}_j,\ j \neq i) &= \alpha^{-1} \end{aligned}$$

where the suffix 'j nhbr i' denotes the four neighbor pixels at distance one from pixel i. The parameter α measures the smoothness of the 'true' image.

Assuming a toroidal edge correction, the eigenvalues of the matrix $I - \phi N$ are $\lambda_{ij} = 1 - 2\phi(\cos(2\pi i/m) + \cos(2\pi j/n)), i = 0, 1, 2, \ldots, m-1, j = 0, 1, 2, \ldots, n-1$. So, the density of $\mathbf{f}$ has the form

$$p(\mathbf{f}) = \frac{1}{Z_{\text{prior}}(\alpha)} \exp\{-\frac{1}{2}\alpha\, \mathbf{f}^t(I - C)\mathbf{f}\}$$

where $Z_{\text{prior}}(\alpha) = (\prod_{i,j} \lambda_{ij})^{-1/2}(2\pi/\alpha)^{p/2}$, $C = \phi N$.

This prior model can be easily modified to work at log scale, which is the right scale for the deconvolution of galaxies (Molina et al. 1992a). Furthermore, this model can also take into account the existence of different objects in the image (Molina et al. 1992b).

Let us now examine the prior model used in the R-L restoration method. This method aims at maximizing $p(\mathbf{g}|\mathbf{f})$ when this conditional distribution is Poissonian. Under the Bayesian framework this is the same as maximizing the posterior distribution for the prior model $p(\mathbf{f}) = \text{const}$, together with the Poissonian noise model. The meaning of this prior is simple; all possible restorations have the same probability.

5. Algorithms

Having defined the prior and observational models, let us move on to estimate the MAP.

Following the R-L method, which, as we have said, corresponds to maximum a posterior estimation with a uniform image prior, we seek to find

$$\mathbf{f}_{\mathrm{RL}} = \arg\max_{\mathbf{f}} \{ \prod_{i=1}^{p} \exp[-(\mathbf{D}\mathbf{f})_i][(\mathbf{D}\mathbf{f})_i]^{g_i}/g_i! \} \,. \tag{5}$$

Applying logarithms and differentiating with respect to $\mathbf{f}$ we obtain the following equation:

$$\mathbf{D}^t(\mathbf{g}/\mathbf{D}\mathbf{f}_{\mathrm{RL}}) = \mathbf{1} \tag{6}$$

where we have assumed that $\mathbf{D}^t\mathbf{1} = \mathbf{1}$. To solve (6) Lucy uses the iterative scheme

$$\mathbf{f}_j^i = \mathbf{f}_j^i[\mathbf{D}^t(\mathbf{g}/\mathbf{D}\mathbf{f}^i)]_j$$

where i denotes iteration and j component of the vector. This iterative scheme is justified as an iterative scheme derived from EM principles. However, it can also be obtained by multiplying both sides of (6) by $\mathbf{f}$.

Let us now assume that we want to impose smoothness constraints on the solution by using a CAR prior model, we have

$$p(\mathbf{f}|\mathbf{g}) \propto \exp[-\frac{\alpha}{2}\mathbf{f}^t(I - C)\mathbf{f}] \prod_{i=1}^{p} \exp[-(H\mathbf{f})_i][(H\mathbf{f})_i]^{g_i}/g_i! \,. \tag{7}$$

Differentiating $-\log p(\mathbf{f}|\mathbf{g})$ with respect to $\mathbf{f}$ we obtain

$$\alpha(I - C)\mathbf{f} + \mathbf{1} - H^t(\mathbf{g}/H\mathbf{f}) = 0$$

or

$$\mathbf{f} + \alpha^{-1}\mathbf{1} = C\mathbf{f} + \alpha H^t(\mathbf{g}/H\mathbf{f}) \,. \tag{8}$$

Multiplying both sides of (8) by $\mathbf{f}$ we obtain the following iterative scheme:

$$\mathbf{f}_j^{i+1} = (\mathbf{f}_j^i[C\mathbf{f}^i]_j + \alpha^{-1}\mathbf{f}_j^i H^t(\mathbf{g}/H\mathbf{f}^i)]_j)/(\mathbf{f}_j^i + \alpha^{-1}) \tag{9}$$

where i denotes iteration and j component of the vector. This equation can be rewritten as

$$\mathbf{f}_j^{i+1} = \mu_j^i[C\mathbf{f}^i]_j + (1 - \mu_j^i)\mathbf{f}_j^i H^t(\mathbf{g}/H\mathbf{f}^i)]_j$$

where $\mu_j^i = \mathbf{f}_j^i/(\mathbf{f}_j^i + \alpha^{-1})$, $\mu_j^i = 0$ corresponds to the classical R-L restoration method.

Before examining an example, let us briefly comment on the problem of estimating α. Although it is possible to estimate α on a trial and error basis we are currently working on the use of the hierarchical Bayesian approach to image restoration (Molina 1993). The idea is to use the joint distribution defined as

$$p(\alpha, \mathbf{f}, \mathbf{g}) = p(\alpha)p(\mathbf{f}|\alpha)p(\mathbf{g}|\mathbf{f}) \,, \tag{10}$$

integrate (10) on α to obtain $p(\mathbf{f}, \mathbf{g})$, and then find $\underline{\mathbf{f}}$ which satisfies

$$\underline{\mathbf{f}} \text{ maximizes } p(\mathbf{f}, \mathbf{g}) \,.$$

For the problem we have at hand we can use improper noninformative priors, $p(\alpha) \propto$ const, and also gamma distributions. The relationship between this hierarchical model and the method developed by Katsaggelos, Kang, and Banham (1994) are under investigation.

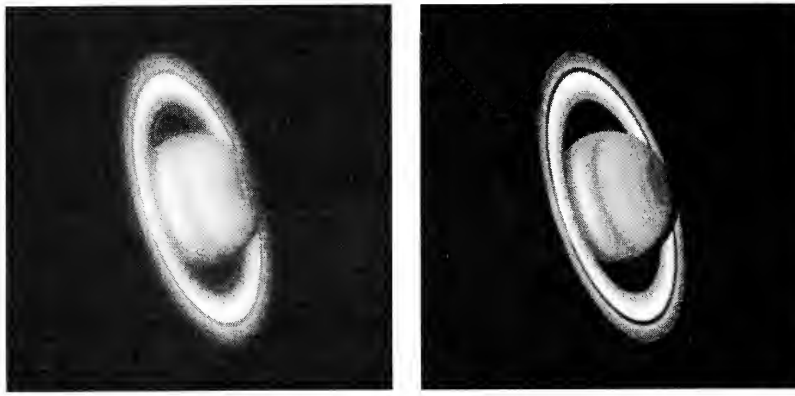


Figure 1. (a) Observed Saturn image. (b) Restoration with $\alpha^{-1} = 106,000$ after 50 iterations.

6. Examples

We shall apply the algorithm just developed to the restoration of a WF/PC image of Saturn. The distorted and restoration images obtained by our method after 50 iterations, using (9), with $\alpha^{-1} = 106,000$ are shown in Fig. 1.

Acknowledgments. This work has been supported by the "Comisión Nacional de Ciencia y Tecnología" under contract TIC91-0660.

References

Lucy, L. 1974, AJ, 79, 745

Katsaggelos, A. K., Kang, M. G., & Banham, M. R. 1994, this volume

Molina, R., & Ripley, B. D. 1989, J. Appl. Stat., 16, 193

Molina, R., Olmo, A., Perea, J., & Ripley B. D. 1992a, AJ, 103, 666

Molina, R., Ripley B. D., Molina, A., Moreno, F., & Ortiz, J. L. 1992b, AJ, 104, 1662

Molina, R. 1993, submitted to PAMI

Ripley, B. D. 1981, *Spatial Statistics*, Wiley, New York

The Restoration of HST Images and Spectra II
Space Telescope Science Institute, 1994
R. J. Hanisch and R. L. White, eds.

HST Image Restoration with Variable Resolution

Jorge Núñez

Departament d'Astronomia i Meteorologia, Universitat de Barcelona, Av. Diagonal 647, E-08028 Barcelona, and Observatorio Fabra, Barcelona, Spain

Jorge Llacer

Engineering Division, Lawrence Berkeley Laboratory, University of California, Berkeley, CA 94720

Abstract. Bayesian and maximum entropy algorithms that use a constant balancing parameter present large residuals in the low or the high S/N regions depending on the value chosen for the parameter. We present the development of a variant of the Bayesian algorithm with entropy prior (called FMAPE) that uses a balancing parameter that is variant in space. This variation of the balancing parameter implies a variable resolution in the restoration, allowing high resolution with low bias in high S/N regions and a lower resolution with lower artifacts in poor S/N regions. We have applied the algorithm to the FOC and WF/PC cameras of the HST.

1. Introduction

Since the discovery of the spherical aberration problem of the Hubble Space Telescope in 1990, a substantial amount of work has been done in image restoration directed towards optical astronomy. During the past few years, we have been developing statistically based algorithms for image restoration (Núñez and Llacer 1991, Núñez and Llacer 1993, Llacer and Núñez 1990, Llacer et al. 1993). In particular, we have developed Bayesian algorithms with entropy prior (FMAPE), and methods based on feasibility and cross-validation in order to compute the balancing parameter between the entropy prior and the likelihood term. The balancing parameter defines the degree of smoothness of the solution.

Bayesian and maximum entropy solutions are increasingly being used for the restoration of images from noisy and incomplete data. The Bayesian strategy seeks the image of highest probability given the data. The Bayesian target function and the likelihood are related through Bayes' Rule, which includes the probability distribution of the image, also known as image prior. We use entropy to define this prior probability.

2. Notation and Model

The notation that will be used in this paper can be summarized as follows:

$p_j, j = 1, \ldots D$	projection (measured) data
$a_i, i = 1, \ldots B$	emission density in the image (parameters to be estimated)
f_{ji}	Point Spread Function (PSF) or probability that an emission in pixel i in the source be detected at detector j
$b_j, j = 1, \ldots D$	background in the data
$n_j, j = 1, \ldots D$	read-out noise in the data
$C_j, j = 1, \ldots D$	detector gain corrections (flat field)

$f'_{ji} = f_{ji}/C_j$	corrected PSF
$q_i = \sum_{j=1}^{D} f'_{ji}$	total detection probability for an emission from pixel i
$h_j = \sum_{i=1}^{B} f'_{ji} a_i + b_j$	forward projection or blurring operation
Q_i	prior distribution density or default image (if any)
$\mathbf{p}, \mathbf{a}, \mathbf{f}, \mathbf{b}, \mathbf{n}, \mathbf{C}, \mathbf{f}', \mathbf{q}, \mathbf{h}, \mathbf{Q}$	corresponding arrays

We shall work with the following model: an object emits light with intensity given by a spatial distribution **a**. The light is focused by the optical system over a detector array consisting of individual discrete and independent detectors. Each detector has a different quantum efficiency characterized by a gain correction distribution **C**. A certain background radiation **b**, coming mainly from the sky but also from sources internal to the detector, is detected along with the spatial distribution **a**. We assume that the detection process is Poisson distributed. Finally, the detector is read by an electronic process which adds a Gaussian read-out noise **n** with zero mean and known standard deviation σ. The imaging equation corresponding to this model is:

$$\mathbf{f}' \mathbf{a} + \mathbf{b} + \mathbf{n} = \mathbf{p} \ . \tag{1}$$

Equation (1) in discrete form becomes:

$$\sum_{i=1}^{B} \frac{f_{ji}}{C_j} a_i + b_j + n_j = p_j \quad j = 1, \ldots, D \ .$$

Most of the imaging systems used in astronomy, such as charge coupled device (CCD) cameras and image pulse counting systems (IPCS), are described by Eq. (1).

3. The FMAPE Algorithm

We use the Bayesian strategy to obtain an iterative algorithm for image restoration. The application of Bayes' theorem to the image restoration problem gives

$$\mathbf{P}(\mathbf{a}|\mathbf{p}) = \frac{\mathbf{P}(\mathbf{p}|\mathbf{a})\mathbf{P}(\mathbf{a})}{\mathbf{P}(\mathbf{p})} \ . \tag{2}$$

The most probable image **a**, given data **p**, is obtained by maximizing $\mathbf{P}(\mathbf{a}|\mathbf{p})$ in Eq. (2) or the product $\mathbf{P}(\mathbf{p}|\mathbf{a})\mathbf{P}(\mathbf{a})$ since $\mathbf{P}(\mathbf{p})$ is constant.

The conditional probability $\mathbf{P}(\mathbf{p}|\mathbf{a})$ describes the noise in the data and its possible object dependence. It is fully specified in the problem by the likelihood function. As indicated above, we have two processes: the first to form the image on the detector array and the second to read the detector. Taking the whole process into account, the compound likelihood is (Núñez and Llacer 1993)

$$\mathbf{L} = \mathbf{P}(\mathbf{p}|\mathbf{a}) = \prod_{j=1}^{D} \sum_{k=0}^{\infty} \frac{1}{\sqrt{2\pi}\sigma} \ e^{-\frac{(k-p_j)^2}{2\sigma^2}} e^{-h_j} \ \frac{(h_j)^k}{k!}$$

and its logarithm is

$$\log \mathbf{L} = \sum_{j=1}^{D} \left[-\log(\sqrt{2\pi}\sigma) - h_j + \log \sum_{k=0}^{\infty} \left(e^{-\frac{(k-p_j)^2}{2\sigma^2}} \ \frac{(h_j)^k}{k!} \right) \right] . \tag{3}$$

The authors first introduced this compound likelihood for restoration of Poisson data in the presence of read-out noise (Llacer and Núñez 1990). Snyder et al. (1993) use the same likelihood form for

CCD cameras. If the process were pure Poisson (no read-out noise), the logarithm of the likelihood would be the classical expression

$$\log \mathbf{L} = \sum_{j=1}^{D} \left[-h_j + p_j \log h_j - \log(p_j!) \right] . \tag{4}$$

We use entropy to define the prior probability $\mathbf{P}(\mathbf{a})$ in a generalization of the concepts originally described by Frieden (1972). Let N be the total energy (usually the number of counts or photons) in the object. Assume that there is an intensity increment Δa describing an intensity jump to which we can assign some appropriate physical meaning.

Assume, in addition, that we have prior information regarding the statistical makeup of the object obtained for example, from previous observations or from observations at other wavelengths. If $Q_i \;\; i = 1, \ldots, B$ is the prior energy distribution, the prior probability of a photon going to pixel i is given by

$$g_i = Q_i / \sum_{l=1}^{B} Q_l .$$

In that case, the number of ways that the object can occur is

$$W(\mathbf{a}) = \frac{(N/\Delta a)!}{\prod_{i=1}^{B} (a_i/\Delta a)!} \prod_{i=1}^{B} (g_i)^{a_i/\Delta a} . \tag{5}$$

We take the prior probability $\mathbf{P}(\mathbf{a})$ of the object to be proportional to its multiplicity, as given by Eq. (5). Using Stirling's approximation and taking into account that both $\sum a_i$ and $\sum Q_i$ are constants, it is easy to show that the logarithm of the probability $\mathbf{P}(\mathbf{a})$ is

$$\log \mathbf{P}(\mathbf{a}) = -\sum_{i=1}^{B} \frac{a_i}{\Delta a} \log \frac{a_i}{Q_i} + \text{constant terms} .$$

This expression is the Shannon form of entropy with the inclusion of the parameter Δa and the spatial prior information $\mathbf{Q}$. This form of the entropy is also called cross-entropy.

The Bayesian function to be maximized is

$$BY = -\sum_{i=1}^{B} \frac{a_i}{\Delta a} \log \frac{a_i}{Q_i} + \sum_{j=1}^{D} \left[-\log(\sqrt{2\pi}\sigma) - h_j + \log \sum_{k=0}^{\infty} \left(e^{-\frac{(k-p_j)^2}{2\sigma^2}} \frac{(h_j)^k}{k!} \right) \right] -$$

$$- \mu \left(\sum_{i=1}^{B} q_i a_i - \sum_{j=1}^{D} p_j + \sum_{j=1}^{D} b_j \right) , \tag{6}$$

where h_j is given by

$$h_j = \sum_{i=1}^{B} f'_{ji} a_i + b_j$$

and μ is a Lagrange multiplier for the conservation of counts. Note that the relative weight of the two terms in the Bayesian function Eq. (6) is controlled by Δa.

To obtain the maximum of Eq. (6), we set $\partial BY / \partial a_i = 0$ and apply the method of successive substitutions, which affords us greater flexibility than other methods and results in rapidly converging algorithms. The Bayesian maximum a posteriori algorithm with entropy prior (FMAPE) is given by the iterative formula

$$a_i^{(k+1)} = K a_i^{(k)} \left[\frac{1}{q_i} \sum_{j=1}^{D} \frac{f_{ji} p'_j}{\sum_{l=1}^{B} f_{jl} a_l^{(k)} + C_j b_j} - \frac{1}{q_i \Delta a} \left(\log \frac{a_i^{(k)}}{Q_i} + 1 \right) + C \right]^n \quad i = 1, \ldots, B , \tag{7}$$

where

$$p_j' = \frac{\sum_{k=0}^{\infty}\left(k\, e^{-\frac{(k-p_j)^2}{2\sigma^2}}\,\frac{(h_j)^k}{k!}\right)}{\sum_{k=0}^{\infty}\left(e^{-\frac{(k-p_j)^2}{2\sigma^2}}\,\frac{(h_j)^k}{k!}\right)} \quad . \tag{8}$$

In Eq. (7) k is the index of the iteration, K is a constant to preserve the energy in the form $\sum_{i=1}^{B} q_i a_i = \sum_{j=1}^{D} p_j$, computed at the end of each iteration; C is an arbitrary constant (usually $C = 1$) to secure positivity in the solution; n is a constant to accelerate convergence up to approximately three times ($n = 3$). Constants C and n do not affect the point to which the algorithm converges. The parameter Δa determines the point of convergence of the solution between the trivial solution and the maximum likelihood solution. Δa is thus the balancing parameter of the Bayesian approach and it is the inverse of the weight of the entropy versus likelihood.

The iterative algorithm in Eqs. (7) and (8) has a number of desirable characteristics: (1) it solves the cases of both pure Poisson data and Poisson data with Gaussian read-out noise, (2) it maintains positivity of the solution, (3) it is easy to implement, (4) it includes case-specific prior information (default map) and flat field corrections, and (5) it removes background and can be accelerated to be faster than the Richardson-Lucy algorithm (Lucy 1974). The main loop (projection and back-projection) of the algorithm is similar in nature to the Expectation Maximization algorithm. The algorithm can be applied to a large number of imaging situations, including CCD and pulse counting cameras both in the presence and in the absence of background.

We can obtain a maximum likelihood algorithm from the general FMAPE by taking the limit when $\Delta a \to \infty$. The algorithm for the maximum likelihood (MLE) case in the presence of read-out noise and background is

$$a_i^{(k+1)} = K a_i^{(k)} \left[\frac{1}{q_i}\sum_{j=1}^{D}\frac{f_{ji}p_j'}{\sum_{l=1}^{B} f_{jl}a_l^{(k)} + C_j b_j}\right]^n \quad i = 1, \ldots, B \,, \tag{9}$$

where p_j' is given by Eq. (8).

In the case of no background and no read-out noise, Eq. (9) becomes

$$a_i^{(k+1)} = a_i^{(k)} \left[\frac{1}{q_i}\sum_{j=1}^{D}\frac{f_{ji}p_j}{\sum_{l=1}^{B} f_{jl}a_l^{(k)}}\right]^n \quad i = 1, \ldots, B \,. \tag{10}$$

For $n = 1$ and disregarding the gain (flat field) corrections ($q_i = 1$), Eq. (10) is identical to the Richardson-Lucy algorithm.

4. New Approach: Space Variant Hyperparameter Δa

The hyperparameter Δa is a key parameter in the Bayesian framework. It determines the relative weight between the prior information and the likelihood in the Bayesian function and therefore determines the chosen solution between the default image and the maximum likelihood solution. Ultimately, it defines the degree of smoothness of the solution. Up to now, we have considered Δa to be an adjustable parameter belonging to object space that is constant for a given image. The value of Δa can be adjusted by cross-validation (Núñez and Llacer 1991; Núñez 1993).

The use of a single Δa in Bayesian restoration produces a good global fit of the data, but usually the part of the image in which the light level is usually noisy (over-restored) while the bright objects (stars) are too smooth (under-restored). This effect is also present in the Richardson-Lucy algorithm at the stopping point. Furthermore, Bayesian methods with entropy and MEM present photometric linearity problems.

The effect can be observed in the image of residuals:

$$R_j = \frac{(\sum_{i=1}^{B} f_{ji}a_i - p_j)^2}{\sum_{i=1}^{B} f_{ji}a_i} \quad j = 1, \ldots, D$$

If the data are Poisson, the R_j values should be randomly distributed around unity. However, for low values of the hyperparameter the residuals are strongly correlated with the data. For a very high value of the hyperparameter, the residuals are not correlated with the data but they are very near zero. In this case the data are overfitted and the resulting image very is noisy with high peaks and low valleys. In general, after obtaining the optimum value of Δa by cross-validation, the residuals for the background part of the image are random but too low, resulting in a noisy restoration. On the other hand, the residuals at the bright stars are large and still correlated with the data, giving a too smooth restoration for the bright objects. As stated above, this effect is also present in the Richardson-Lucy algorithm for a given stopping point.

To solve this problem several approaches have been presented in this conference (Katsaggelos 1994, Lucy 1994, Puetter 1994, Stark et al. 1994, White 1994). Our approach consists in the introduction of a space variant Δa hyperparameter in the Bayesian framework. A variable Δa allows different degrees of smoothing in different regions, ranging from very smooth (or very similar to the prior) regions ($\Delta a \to 0$) to MLE regions ($\Delta a \to \infty$). We have used this variable hyperparameter in connection with known default images for medical tomography (Llacer, Veklerov, and Núñez 1991).

Assume that the intensity increment Δa described in §3 is space variant in the object space. Let $\Delta a_i \quad i = 1, \ldots, B$ be the variable increment. Now, in place of $N/\Delta a$ units we distribute $\sum_{i=1}^{B} a_i/\Delta a_i$ units. The number of ways in which an object can occur is then

$$W(\mathbf{a}) = \frac{(\sum_{i=1}^{B} a_i/\Delta a_i)!}{\prod_{i=1}^{B}(a_i/\Delta a_i)!} \prod_{i=1}^{B} (g_i)^{a_i/\Delta a_i} .$$

A natural choice for the prior energy distribution is

$$g_i = \frac{Q_i/\Delta a_i}{\sum_{i=1}^{B} Q_i/\Delta a_i} .$$

Using the Stirling approximation and taking the prior probability to be proportional to the multiplicity, its logarithm is

$$\log \mathbf{P}(\mathbf{a}) = (\sum_{i=1}^{B} \frac{a_i}{\Delta a_i}) \log(\sum_{i=1}^{B} \frac{a_i}{\Delta a_i}) - \sum_{i=1}^{B} \frac{a_i}{\Delta a_i} \log \frac{a_i Q'}{Q_i} + \text{constant terms} ,$$

where

$$Q' = \sum_{i=1}^{B} \frac{Q_i}{\Delta a_i} .$$

With the modifications described, the algorithm **FMAPEVAR** is

$$a_i^{(k+1)} = K a_i^{(k)} \left[\frac{1}{q_i} \sum_{j=1}^{D} \frac{f_{ji} p_j'}{\sum_{l=1}^{B} f_{jl} a_l^{(k)} + C_j b_j} - \frac{1}{q_i \Delta a_i} \left(\log \frac{a_i^{(k)} Q'}{Q_i \sum_{l=1}^{B} \frac{a_l^{(k)}}{\Delta a_l}} \right) + C \right]^n \quad i = 1, \ldots, B .$$

The modified algorithm has the same desirable characteristics as the FMAPE algorithm, but now it is possible to change the values of Δa_i in order to determine the degree of smoothing in different regions or, in connection with known default images, to adjust the proximity between the restoration and the default image $\mathbf{Q}$.

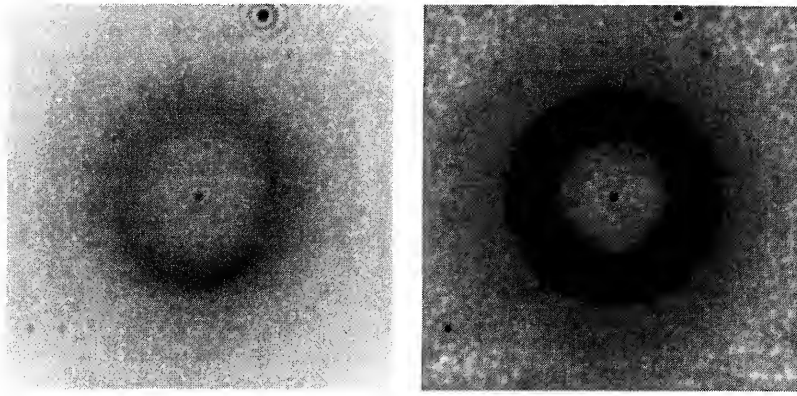


Figure 1. Left: Raw image for the planetary nebula simulation. Right: Restoration of the planetary nebula using the FMAPEVAR algorithm with three channels.

To compute the Δa_i for the different regions, we can use segmentation techniques in conjunction with local cross-validation to define several levels of data fitting, depending on the signal-to-noise ratio. We are currently studying the suitability of the different segmentation algorithms to this problem.

5. Results

To illustrate the use of the algorithm to control noise amplification, we have reconstructed a 256×256 pixels planetary nebula simulation for the PC camera prepared by R. Hanisch at ST ScI. The PSF has a similar radius to the ring of the planetary nebula. The images combine background, a diffuse object, and several bright stars. Fig. 1 (left) shows the raw image.

We have reconstructed the raw image using the FMAPEVAR algorithm with a space variant Δa parameter. We set $\Delta a = 200$ for the background, $\Delta a = 500$ for the diffuse planetary nebula and a high value for the bright stars of $\Delta a = 10^6$ allowing the algorithm to reach maximum likelihood in the stars.

Fig. 1 (right) shows the result of this three-channel restoration. We obtained a smooth background, a well reconstructed nebula, and sharp images of the stars. We suppressed the noisy amplification in the background and in the nebula, while fully developing the images of the stars.

In order to compare the new algorithm with the constant Δa approach and to study the photometric properties of the restoration, we made two restorations of the raw image using a constant value of $\Delta a = 200$ and a constant $\Delta a = 500$ respectively. Since this is a simulated image, we can quantitatively study the restorations with respect to the true unblurred image. For the restorations, we computed the percentage of energy recovered in a 7×7 box located in the background, the ring, the central star and the brightest star.

Fig. 2 shows the results. The restoration with constant $\Delta a = 200$ only has correct photometry for the background. The restoration with constant $\Delta a = 500$ performs well in the background and in the ring (although the background is too noisy in the restoration) but the photometry is poor in the stars because the stars are not properly developed. However, the restoration with variable Δa in three channels performs much better in the stars, while maintaining the quality the photometry in the ring and the background. We obtained an accuracy better than 95% (or 0.05 mag.) for the whole image. We adjusted the values of the variable Δa approximately without any local cross-

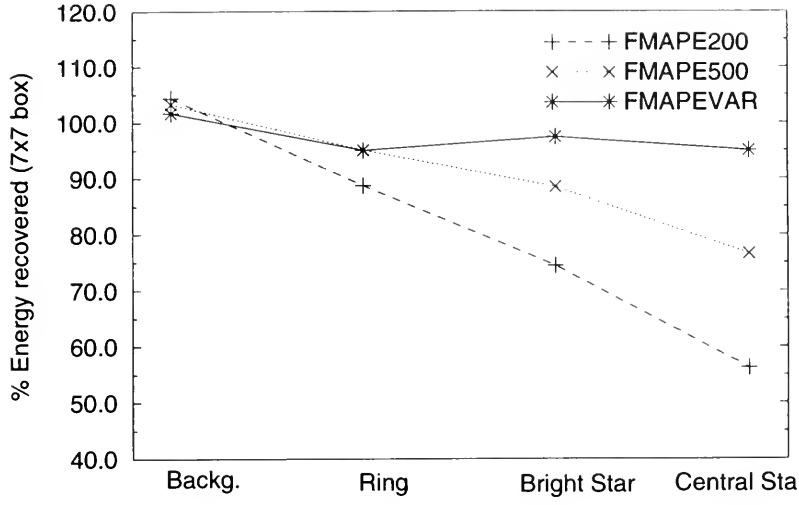


Figure 2. Photometric behavior of the FMAPEVAR algorithm.

validation or segmentation. Thus, by using an optimum adjustment it is possible to obtain even better photometry.

6. Conclusions

In this paper we have developed a variant of the Bayesian algorithm with entropy prior (FMAPE) that uses a balancing parameter that is variant in space. To the desirable characteristics of the FMAPE algorithm, this development adds the capability of different degrees of smoothing in different regions of the image. This implies a variation in resolution in the restoration, allowing high resolution in high S/N regions and lower resolution with lower artifacts in poor S/N regions

The algorithm can prevent amplification of noise while maintaining the photometric accuracy in the whole image (background, diffuse objects and stars).

We have applied the algorithm to the planetary nebula simulation using three different values for the hyperparameter corresponding to the background, nebula, and stars respectively, with excellent results both in the general aspect of the restoration and in the photometry.

More work is still needed to determine the different hyperparameters to use in the different regions. We are studying image segmentation techniques to optimize the process.

7. Acknowledgments

The authors would like to thank Robert J. Hanisch of the ST ScI for supplying part of the data used in this paper and for the support and computing facilities at that institute. This work was supported in part by the DGICYT Ministerio de Educación y Ciencia (Spain) under grant No. BP90–0478.

Partial support was also obtained from the D. G. U. Generalitat de Catalunya. J. Llacer's work is supported by the U. S. Department of Energy under contract No. DE–AC03–76SF00098.

References

Frieden, B. R. 1972, J. Opt. Soc. Amer., 62, 511

Katsaggelos, A., 1994, this volume

Llacer, J., & Núñez, J. 1990, in *The Restoration of HST Images and Spectra*, R. L. White & R. J. Allen, eds., Space Telescope Science Institute, Baltimore, 62

Llacer, J., Veklerov, E., & Núñez, J. 1991, in *Information Processing in Medical Imaging XII* (IPMI), Lecture Notes in Computer Science, 511, 81

Llacer, J., Veklerov, E., Coakley, K.J., Hoffman, E.J., & Núñez, J. 1993, IEEE Trans. Med. Imaging, 12 (2), 215

Lucy, L. B. 1974, AJ, 79, 745

Lucy, L. B., 1994, this volume

Núñez, J. 1993, *Image Restoration Newsletter*, R. J. Hanisch, ed., Space Telescope Science Institute, Baltimore, 1, 24

Núñez, J., & Llacer, J. 1991, in *Astronomical Data Analysis Software and Systems I*, ASP Conference Series 25, D. M. Worrall, C. Biemesderfer, & J. Barnes, eds., 210

Núñez, J., & Llacer, J. 1993, PASP, 105, 1192

Puetter, R. 1994, this volume

Snyder, D. L., Hammoud, A. M., & White, R. L. 1993, J. Opt. Soc. Amer., 10, 1014

Starck, J. L., Murtagh, F., & Bijaoui, A. 1993, this volume

White, R. 1993, this volume

The Restoration of HST Images and Spectra II
Space Telescope Science Institute, 1994
R. J. Hanisch and R. L. White, eds.

Regularization and Smoothing for the Restoration of Hubble Space Telescope Images

P. Benvenuti

ST-ECF, Karl-Schwarzschild-Str. 2, D-85748 Garching, Germany

F. Maggio

CRS4, Via N. Sauro 10, I-09123 Cagliari, Italy

S. Seatzu

Dipartimento di Matematica, Universitá di Cagliari, Via Merello 92, I-09123 Cagliari, Italy

Abstract. Two numerical methods for the restoration of noisy images from the Hubble Space Telescope are presented. The first one stems from a B-spline expansion followed by the Tikhonov regularization method. The second one is based on a coarse smoothing coupled with the Tikhonov regularization. The effectiveness of these methods is illustrated by the restoration of some images already proposed in the literature.

1. Introduction

From the mathematical point of view the problem of the restoration of the images generated by the Hubble Space Telescope (HST) consists of the solution in $L^2(\mathbb{R}^2)$ of the convolution equation

$$(k * f)(\underline{x}) = g(\underline{x}) \Longleftrightarrow \int k(\underline{x} - \underline{y}) f(\underline{y}) d\underline{y} = g(\underline{x}) \tag{1}$$

where the kernel k represents the Point Spread Function (PSF), the symbol $\underline{z} \equiv (z_1, z_2)$ denotes a vector of $\mathbb{R}^2$ and f is the unknown function, that must be non-negative. As the image g is contaminated by noise, the deconvolution of equation (1) by the usual Fourier methods, that is the Fourier inversion of the solution of the equation

$$\widehat{k}(\underline{\omega})\widehat{f}(\underline{\omega}) = \widehat{g}(\underline{\omega}) \ , \tag{2}$$

where the notation $\widehat{u}(\underline{\omega})$ denotes the Fourier transform of u, is not effective. Many methods exist, both statistical and deterministic, linear and non-linear, for solving Eq. (1). They are essentially regularization methods, in the sense that the quality of the results, that is the value of the signal-to-noise ratio of the restored images, strongly depends on a non-negative parameter. In the case of iterative methods this parameter is the number of iterations. A different approach, based on the usual Fourier techniques applied to images previously smoothed, could also be adopted. Unfortunately the smoothing techniques are really effective only if the noise is white, but this hypothesis is not acceptable in our case.

In this work we use two different approaches: the first one is based on both the finite dimensional approximation of the unknown function and on the so-called Tikhonov regularization method; the second stems from a coarse smoothing and also exploits the above regularization method.

2. Finite Dimensional Approximation and Tikhonov Regularization Method

The Tikhonov regularized solution of the convolution equation (1) is the minimizer in H of the regularization functional (Tikhonov & Arsenin 1977, Morozov 1984, Groetsch 1984):

$$J_\lambda(f) = \| k * f - g \|^2 + \lambda \| f \|^2 \ , \tag{3}$$

where $\lambda \in I\!R^+$ is the regularization parameter to be chosen depending on the noise level. This functional can be expressed by the Plancherel theorem in the frequency space as follows:

$$\tilde{J}_\lambda \left(\hat{f}\right) = \| \hat{k}\hat{f} - \hat{g} \|^2 + \lambda \| \hat{f} \|^2 \ . \tag{4}$$

Hence it is straightforward to prove, by variational calculus, that the minimizer $\hat{f}_\lambda$ of $\tilde{J}_\lambda \left(\hat{f}\right)$ is the solution of the equation

$$\left(|\hat{k}(\underline{\omega})|^2 + \lambda\right) \hat{f}(\underline{\omega}) = \hat{k}^*(\underline{\omega})\hat{g}(\underline{\omega}) \ , \tag{5}$$

where α^* denotes the complex conjugate of α.

Since any image is numerically represented as a non-negative function with finite support R (in other words: $f(\underline{x}) = 0$ outside R), we restrict our considerations to the rectangle $R = [0, 2L] \times [0, 2L]$. The regularized solution f_λ, in a discrete set of equally spaced points of R, is then the inverse discrete Fourier transform of

$$\hat{f}_\lambda(\underline{\omega}) = \frac{\hat{k}^*(\underline{\omega})}{|\hat{k}(\underline{\omega})|^2 + \lambda} \hat{g}(\underline{\omega}) \ , \tag{6}$$

which can be evaluated easily by the Fast Fourier Transform (FFT) method.

Let $V = \{y_{rs} = (rh, sh) : r, s = 0, 1, ..., N-1,\ h = 2L/N\}$ be the set of knots of a regular mesh of the space domain R and $B_{m,l}$ be the B-spline of order m relative to the $m+1$ equally spaced knots $\{lh, (l+1)h, ..., (l+m)h\}$ of the extended interval $[0, 2L + (m-1)h]$. Then, the unknown function f can be approximated by the $m-th$ order spline

$$s_m(y_1, y_2) = \sum_{r=0}^{N-1} \sum_{s=0}^{N-1} c_{rs} B_{m,r}(y_1) B_{m,s}(y_2) \ , \tag{7}$$

whose coefficients, as a direct consequence of (6), may be obtained solving in a regular mesh of points the system

$$\sum_{r=0}^{N-1} \sum_{s=0}^{N-1} c_{rs} \hat{B}_{m,r}(\omega_1) \hat{B}_{m,s}(\omega_2) = \frac{\hat{k}^*(\omega_1, \omega_2)}{|\hat{k}(\omega_1, \omega_2)|^2 + \lambda} \hat{g}(\omega_1, \omega_2) \ , \tag{8}$$

where

$$\hat{B}_{m,l}(\omega) = h e^{-ilh\omega} \left(e^{-ih\omega/2} \mathrm{sinc}(h\omega/2)\right)^m \ , \ \mathrm{sinc}(x) = \frac{\sin x}{x} \ , \ i = \sqrt{-1}. \tag{9}$$

Let $\Omega = \{\omega_{rs} = (\frac{r\pi}{L}, \frac{s\pi}{L}) : r, s = 0, 1, ..., N-1\}$ be the set of knots of a uniform decomposition of the frequency domain $[0, \pi N/L] \times [0, \pi N/L]$. The matrix $(A)_{rs} = e^{-irh\omega_s}$ is an orthogonal matrix since $A^*A = NI$, where A^* is the adjoint of A and I is the identity matrix of order N; then we can solve the system (8) using the inverse discrete Fourier transform. Thus we obtain:

$$c_{rs} = \frac{1}{4L^2} \sum_{j=0}^{N-1} \sum_{k=0}^{N-1} e^{2\pi ijr/N} e^{2\pi iks/N} H_\lambda^m(j, k; h) \ , \tag{10}$$

where

$$H_\lambda^m(j,k;h) = \left(\frac{e^{ih\omega_j/2}}{\mathrm{sinc}(h\omega_j/2)}\right)^m \left(\frac{e^{ih\omega_k/2}}{\mathrm{sinc}(h\omega_k/2)}\right)^m \hat{f}_\lambda(\omega_{jk}) \ ,$$

$\omega_l = l\pi/L$ and $\hat{f}_\lambda(\omega_{jk})$ is the Tikhonov regularized solution in the frequency space.

Since the approximation properties of splines are only asymptotically satisfied, that is they hold only if h converges to 0, we replace the above expression of c_{rs} by its limit for $h \to 0$. In this way $H_\lambda^m(j,k;h)$ converges to $\hat{f}_\lambda(\omega_{jk})$ and then, excep for edge effects, $c_{rs} \to f_\lambda(y_{rs})$. In other words, the coefficients of B-splines converge, for each value of m, to the values of the Tikhonov regularized solution obtained by the discrete Fourier transform applied to $\hat{f}_\lambda(\omega_{jk})$.

This regularization technique, based on approximation by splines, can then be seen as a method, with a very low computational complexity, for recovering the image in all the points (except for the edges) of its domain and not only in the knots of its mesh. It is remarkable that these limits can also been obtained by expressing f as a linear combination of sinc functions and then applying the Tikhonov regularization method in the frequency space. Such an expansion is usual for band limited functions, as a direct consequence of the sampling theorem which states that, if $\hat{f}(\omega) = 0$ outside the domain $\Omega_0 = [-\omega_0, \omega_0] \times [-\omega_0, \omega_0]$, then the values of f at the sampling points $\{(r\pi/\omega_0, s\pi/\omega_0) \ , \forall \text{ integer } r, s\}$ can be interpolated by a bi-infinite series of sinc functions to reproduce f. In our numerical computations we approximate this series by the finite sum

$$f(y_1, y_2) \simeq \sum_{r=0}^{N-1} \sum_{s=0}^{N-1} f(y_{rs}) \mathrm{sinc}(\omega_0 y_1 - r\pi) \mathrm{sinc}(\omega_0 y_2 - s\pi) \ , \tag{11}$$

where the coefficients $f(y_{rs})$ are the values of f at the sampling points. We recall, for the reader's convenience, that the Fourier transform of a sinc function is a square pulse. More precisely

$$\widehat{\mathrm{sinc}}(\omega) = \pi \chi_{[-1,1]} \quad \text{where} \quad \chi_{[-1,1]} = \begin{cases} 1 & , \ |\omega| \leq 1 \ , \\ 0 & , \ |\omega| > 1 \ . \end{cases}$$

Therefore the Tikhonov regularized solution of equation (1) sought in a finite-dimensional span of sinc functions, is the solution on a discrete set of $\underline{\omega}$ values of the problem

$$\sum_{r=0}^{N-1} \sum_{s=0}^{N-1} e^{-i\pi r\omega_1/\omega_0} e^{-i\pi s\omega_2/\omega_0} f_\lambda(y_{rs}) = \left(\frac{\omega_0}{\pi}\right)^2 \frac{\hat{k}^*(\underline{\omega})}{|\hat{k}(\underline{\omega})|^2 + \lambda} \hat{g}(\underline{\omega}) \ . \tag{12}$$

For each value of $\lambda > 0$ we have a regularized solution f_λ whose expansion by sinc is uniquely characterized by the coefficients $f_\lambda(y_{rs})$ which can be easily obtained applying the FFT to relation (12). These coefficients, as already remarked, are the limits (for $h \to 0$) of the coefficients of the same regularized solution spanned by B-splines. This fact is not completely surprising because the interpolation in the sampling points by splines of order m, for $m \to \infty$ converges to the interpolation in the same points by sinc functions (Schoenberg 1973).

3. Smoothing and Regularization

The recovery of HST images has also been obtained by a method based on the following two steps:

1. The blurred image is partially smoothed using the public domain package WAVE II of S. Mallat;

2. The convolution equation with the data function partially smoothed is then solved by the Tikhonov regularization technique.

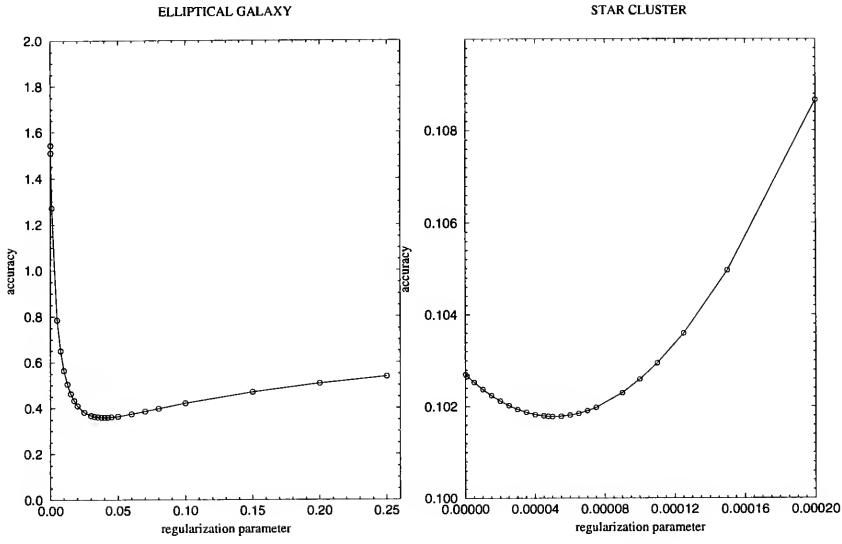


Figure 1. Accuracy for the Elliptical Galaxy (left) and the Star Cluster (right).

Although the mentioned package contains a special coding algorithm for smoothing images, we did not use it, because it is really effective only when the noise is white and this is not true in our case. For this reason we prefer to use the double layer coding algorithm of the same package that separates the edge from the texture information (Mallat and Zhong 1992).

By so doing, multiscale edges can be detected and characterized with a wavelet transform, while the location of the edges in the images is provided by the local maxima of the wavelet transform modulus. In this way a first image approximation, from the important edges that are selected by the coding algorithm, is then reconstructed. The edge coding algorithm restores the main image features but removes all the textures as well as fine image details. An error image is then computed by subtracting the coded image from the original one and, in order to recover some of the texture information, the error image is coded within a wavelet orthonormal basis. This orthogonal basis decomposes the error image into details appearing at different resolutions and within different spatial orientations. The error image to be coded is then chosen both from the image fluctuations detected at different resolution scales and the number of pixels of the original image. The smoothed image is finally obtained by adding up the coded error to the coded edge image.

This image is only roughly smoothed. For this reason, in the second step, instead of solving Eq. (1) with this coded image, we solve it in a discrete set of equally spaced points of R, by the Tikhonov regularization method of § 2.

4. Numerical Results

In our numerical experiments we mainly considered two sets of simulated data provided by R. Hanisch, namely the Star Cluster and the Elliptical Galaxy 2 (Hanisch 1993).

Fig. 1 shows the relative error (the inverse of the signal-to-noise ratio) achieved using the Tikhonov method in the restoration of the above mentioned images. These two cases show that much better accuracy is achieved for the Star Cluster than for the Elliptical Galaxy image. We

mainly impute this fact to the strong difference existing between the signal-to-noise ratio associated with the Star Cluster and with the Elliptical Galaxy. Indeed their ratio is approximately 3.6.

We recall that any linear system $A\underline{x} = \underline{b}$, where $\| \delta\underline{b} \| / \| \underline{b} \|$ is the inverse of the signal-to-noise ratio of the noisy data and $\| \delta\underline{x} \| / \| \underline{x} \|$ is the corresponding ratio for the solution, provides the following inequality (Stoer and Bulirsch 1980):

$$\frac{\| \delta\underline{x} \|}{\| \underline{x} \|} \leq \mathrm{cond}(A) \frac{\| \delta\underline{b} \|}{\| \underline{b} \|} \quad . \tag{13}$$

Therefore, as the condition numbers of the Fourier transforms of the two PSFs provided by R. Hanisch have approximately the same value, we must expect a better signal-to-noise ratio in the restoration of Star Cluster images. For this reason we feel that, for solving Eq. (1), the use of a regularization method is recommended in the case of the Star Cluster and it is mandatory for the Elliptical Galaxy. Furthermore we believe that, as a consequence of its worse signal-to-noise ratio, the latter case provides a more critical but significant test to assess the effectiveness of the numerical methods used for the restoration.

Fig. 2 shows the comparison among the true image (top left), the simulated data (top right) and the Tikhonov regularized solution (bottom left). The image at the bottom right of Fig. 2 will be called into play later on. The most evident effect is that the regularization method is only partially effective in cleaning up the image background contaminated from the simulated noise (compare the true image with the simulated data). The problem is not of particular relevance in this case but it could be if we treat images including low-intensity sources distributed in the background. It can be noticed that while the inside structure of the Galaxy is satisfactorily well restored in the center, where the image function takes its highest values, the restoration is more critical near to the edge, where the Tikhonov method is not fully effective in discriminating between the low values of the true image and the noisy background.

As an example of numerical computation based on relations (11)–(12) we used J413, a real image of Jupiter observed on 4 March 1992 with the WFPC. This is an extended image (without sharp cut-offs) which is essentially band limited. Tikhonov regularization coupled with interpolation by sinc functions allows to restore images on an additional "shifted" grid $\{\xi_{hk}\}$ by the following relation:

$$f_\lambda(\xi_{hk}) = S_{hn} f_\lambda(y_{nm}) S_{km}^T \quad \text{where} \quad S_{hk} = \frac{\sin(\omega_0(\xi_h - y_k))}{\omega_0(\xi_h - y_k)} \quad .$$

The overall computational complexity is $O(N^3)$, which must be compared with $O(N^2 \log_2 N)$ for the Tikhonov restoration. The use of effective BLAS 3 routines for matrix-matrix product (ESSL 2, 1992) makes acceptable the CPU-time increase. As a result, Tikhonov regularization takes about 7 sec on the initial 512×512 mesh and 24 sec on an extended 1024×1024 mesh (IBM RISC 6000/560 with ESSL 2). Fig. 3 shows two Tikhonov restorations of J413 on the above mentioned grids. For validating the method described in § 3 we use again the already shown Elliptical Galaxy image. We partially restored the simulated image using WAVE II and got a whole set of smoothed images corresponding to different scale edges. Taking the smoothed image characterized by the finest scale and treating it by the Tikhonov method, we obtained a significant improvement of the accuracy, as shown in Fig. 4. It can be appreciated that the new minimum of the accuracy takes the value $\eta = 0.29$ instead $\eta = 0.36$ provided by the Tikhonov method. It is worthwhile to note that the improvement is appreciable for low values of the regularization parameter. The restored image obtained by this method is shown in Fig. 2 (bottom right).

5. Conclusions

Our numerical experiments suggest that the approximation by sinc functions coupled with the Tikhonov regularization technique is an effective restoration method. For images satisfying the band limited hypothesis underlying the sampling theorem restoration on finer grids can be obtained without an appreciable overload.

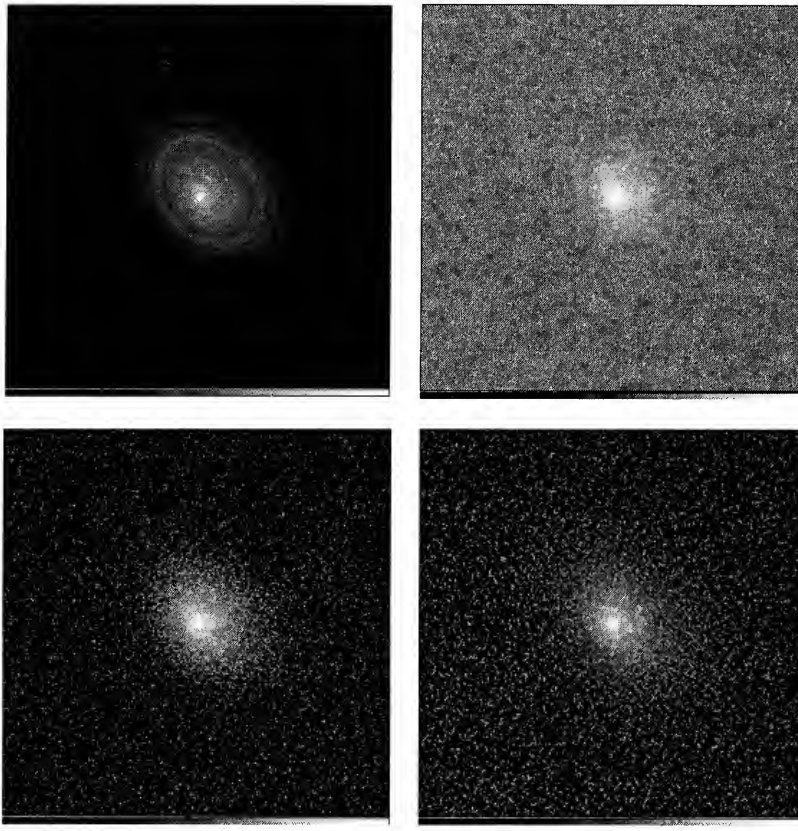


Figure 2. Comparisons of the true image of the Elliptical Galaxy (top left), the simulated image (top right) and two restorations. Bottom left: Tikhonov restoration with $\lambda = 0.04$. Bottom right: restoration obtained by multiresolution smoothing and Tikhonov regularization with $\lambda = 0.01$.

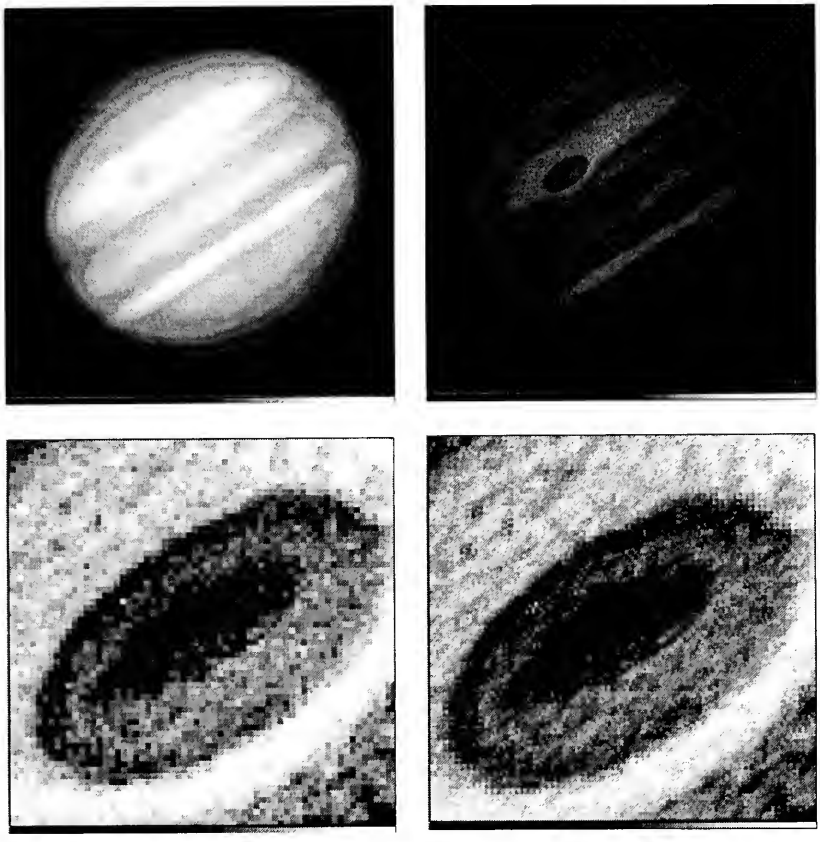


Figure 3. Restorations of Jupiter (J413). Upper left: observed data. Upper right: Tikhonov restoration with $\lambda = 0.01$ on a 512×512 mesh. Lower left: detail of the Red Spot on a 512×512 mesh. Lower right: Red Spot restoration on a 1024×1024 mesh. A square root scaling has been used.

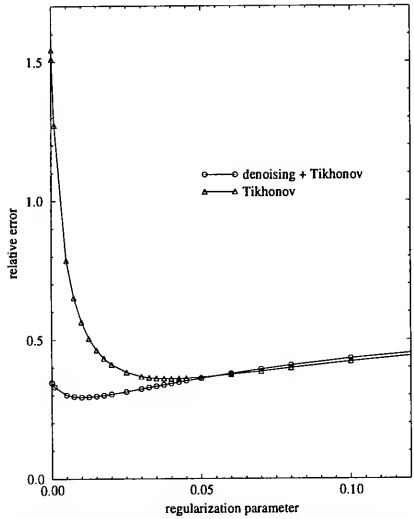


Figure 4. Accuracy improvement of the Tikhonov restored images provided by the multiresolution analysis. The figure refers to the Elliptical Galaxy.

The multiresolution analysis coupled with the Tikhonov regularization brings a remarkable improvement of the accuracy. However it is not clear how this method performs in smoothing images affected by Poisson noise. We feel it will be possible to refine this smoothing method in order to obtain the same accuracy of the most popular restoration techniques (mainly the Richardson-Lucy one) but characterized by a much lower computational effort.

Acknowledgments. We are grateful to A. Innocenti (CRS4) for helping us in using WAVE II. The work of the second author (F. Maggio) was supported by Sardinian Regional Authorities.

References

Engineering and Scientific Subroutine Library (ESSL), V. 2, IBM

Groetsch, C. W. 1984, *The Theory of the Tikhonov Regularization for Fredholm Equations of the First Kind*, Pitman, Boston

Hanisch, R. 1993, in *Restoration Newsletter*, R. J. Hanisch, ed., Space Telescope Science Institute, Baltimore, 1, 76

Mallat, S., & Zhong, S. 1992, IEEE Trans. Patt. Anal. and Mach. Int., 14, 7, 710

Morozov, V. A. 1984, *Methods for Solving Incorrectly Posed Problems*, Springer-Verlag, New York

Schoenberg, I.J. 1973, *Cardinal Spline Interpolation*, CBMS-NSF Series in Applied Math., SIAM Publ., Philadelphia, 54

Stoer, J., & Bulirsch, R. 1980, *Introduction to Numerical Analysis*, Springer-Verlag, Berlin

Tikhonov, A. N., & Arsenin, V. I. 1977, *Solutions of Ill-Posed Problems*, Winston & Sons, Washington

The Restoration of HST Images and Spectra II
Space Telescope Science Institute, 1994
R. J. Hanisch and R. L. White, eds.

Compensation for Read-Out Noise in HST Image Restoration

Donald L. Snyder

Washington University, Department of Electrical Engineering, Box 1127, St. Louis, MO 63130

Carl W. Helstrom

University of California, San Diego CA 92037

Aaron D. Lanterman and Mohammad Faisal

Washington University, St. Louis, MO 63130

Richard L. White

Space Telescope Science Institute, 3700 San Martin Drive, Baltimore, MD 21218

Abstract. Data acquired with the charge coupled device camera on the HST are modeled as an additive Poisson-Gaussian mixture, with the Poisson component representing cumulative counts of object-dependent photoelectrons, object-independent photoelectrons, bias electrons and thermoelectrons, and the Gaussian component representing read-out noise. Two methods are examined for compensating for read-out noise. One method is based upon approximating the Gaussian read-out noise by a Poisson noise and then using the expectation-maximization (modified Richardson-Lucy) algorithm for Poisson distributed data to effect the compensation. This method has been used for restoring HST images. The second method directly uses the expectation-maximization algorithm derived for the Poisson-Gaussian mixture data. This requires the determination of the conditional-mean estimate of the Poisson component of the mixture, which is accomplished by the evaluation of a nonlinear function of the data. The second method requires more computation than the first, but modest improvements in the quality of the restorations are realized, particularly for fainter objects.

1. Introduction

Various methods are being used to compensate for the resolution loss that is present in image data acquired by the Hubble Space Telescope. The common goal of these methods is to invert the excess point spreading caused by spherical aberration in the primary mirror so as to restore the images to their original design resolution. Inverse problems of this type are notoriously ill-posed. Not only can noise amplification cause serious degradation as improved resolution is sought for such problems but, also, the details of the implementation of any restoration method can have a significant impact on the quality of the restoration, with seemingly insignificant design choices having the potential to cause large changes in the result.

There are a variety of noise sources present in HST image data acquired with a CCD camera. The photo-conversion process by which object light is converted into photoelectrons introduces object-dependent noise characterized statistically as a Poisson random process. Nonideal effects introduce extraneous electrons that are indistinguishable from object-dependent photoelectrons. Examples of this excess "noise" include object-independent photoelectrons, bias or "fat zero" electrons, and thermo-electrons. We use the term *background counts* for the cumulative effect of

these extraneous electrons. Background counts are Poisson distributed. Read-out noise further contributes to the degradation of images acquired with the charge coupled device camera aboard the HST. This noise is characterized as a Gaussian random process.

We use the following mathematical model due to Snyder, Hammoud, and White (1993) for describing CCD image data acquired by the HST:

$$r(j) = n_{\text{obj}}(j) + n_0(j) + g(j), \quad j = 0, 1, \cdots, J-1 \tag{1}$$

where $r(j)$ are the data acquired by reading out pixel j of the CCD-camera array, $n_{\text{obj}}(j)$ is the number of object-dependent photoelectrons, $n_0(j)$ is the number of background electrons, $g(j)$ is readout noise, and J is the number of pixels in the CCD camera array. The statistical description of these quantities is as follows. The random variables $n_{\text{obj}}(j)$, $n_0(j)$, and $g(j)$ are statistically independent of each other and of $n_{\text{obj}}(k)$, $n_0(k)$, and $g(k)$ for $j \neq k$. Object-dependent counts $\{n_{\text{obj}}(j), j = 0, 1, \cdots, J-1\}$ form a Poisson process with mean-value function $\{\mu_{\text{obj}}(j), j = 0, 1, \cdots, J-1\}$, where

$$\mu_{\text{obj}}(j) = \beta(j) \sum_{i=0}^{I-1} p(j|i)\lambda(i)\,, \tag{2}$$

where $\{\beta(j), j = 0, 1, \cdots, J-1\}$ accounts for nonuniform flat-field response, detector efficiency, and the spatial extent of the detector array,

$$\{p(j|i), \quad i = 0, 1, \cdots, I-1, \quad j = 0, 1, \cdots, J-1\}$$

is the HST point spread function (PSF), $\{\lambda(i), i = 0, 1, \cdots, I-1\}$ is the object's intensity function. The flat-field response function $\beta(\cdot)$ is assumed to be known, for example, through a flat-field calibration measurement. The PSF $p(\cdot|\cdot)$ is also assumed to be known, for example, from a theoretical model of HST optics or through observations of an unresolved star. Background counts $\{n_0(j), j = 0, 1, \cdots, J-1\}$ form a Poisson process with mean-value function $\{\mu_0(j), j = 0, 1, \cdots, J-1\}$; this mean-value function is assumed to be known, for example, through a dark-field calibration measurement or from HST images taken near but not including the object of interest. Read-out noise $\{g(j), j = 0, 1, \cdots, J-1\}$ forms a sequence of independent, identically distributed, Gaussian random variables with mean m and standard deviation σ; the constants m and σ are assumed to be known.

One approach for compensating for blurring while recognizing the statistical properties of noise encountered in charge coupled device cameras is based on the method of maximum-likelihood estimation, as discussed by Snyder, Hammoud, and White (1993). This leads to the following iteration for producing a maximum-likelihood estimate of the object's intensity function:

$$\hat{\lambda}^{\text{new}}(i) = \hat{\lambda}^{\text{old}}(i) \frac{1}{\bar{\beta}(i)} \sum_{j=0}^{J-1} \left[\frac{\beta(j)p(j|i)}{\hat{\mu}^{\text{old}}(j)} \right] F\left[r(j), \hat{\mu}^{\text{old}}(j), \sigma \right] , \tag{3}$$

where

$$\bar{\beta}(i) = \sum_{j=0}^{J-1} p(j|i)\beta(j)\,, \tag{4}$$

$$\hat{\mu}^{\text{old}}(j) = \beta(j) \sum_{i=0}^{I-1} p(j|i)\hat{\lambda}^{\text{old}}(i) + \mu_0(j)\,, \tag{5}$$

and where

$$F\left[r(j), \mu, \sigma \right] = E\left[n_{\text{obj}}(j) + n_0(j) | r, \hat{\lambda}^{\text{old}} \right] = E\left[r(j) - g(j) | r, \hat{\lambda}^{\text{old}} \right] \tag{6}$$

is the conditional-mean estimate of $n_{\text{obj}}(j) + n_0(j)$ in terms of the data r and the intensity estimate $\hat{\lambda}^{\text{old}}$ to be updated. Evaluation of the conditional mean (6) yields

$$F[r, \mu, \sigma] = \mu \frac{p(r - m - 1 : \mu, \sigma)}{p(r - m : \mu, \sigma)} , \tag{7}$$

where $p(x : \mu, \sigma)$ is the Poisson–Gaussian mixture probability density given by

$$p(x : \mu, \sigma) = \sum_{n=0}^{\infty} \frac{1}{n!} \mu^n e^{-\mu} \frac{1}{\sqrt{2\pi}\sigma} \exp\left[-\frac{1}{2\sigma^2}(x - n)^2\right] . \tag{8}$$

This iteration is used as follows:

STEP 0. Initialization: select $\hat{\lambda}^{\text{old}}(\cdot)$

STEP 1. Compute $\hat{\lambda}^{\text{new}}(\cdot)$ using (3)

STEP 2. If done, display $\hat{\lambda}^{\text{new}}(\cdot)$ else $\hat{\lambda}^{\text{old}}(\cdot) \leftarrow \hat{\lambda}^{\text{new}}(\cdot)$ and perform STEP 1

Special cases of the iteration (3) have appeared previously in the literature. If $\beta(j) = 1$ and $\mu_0(j) = 0$ for all j, then (3) is the same as that given by Llacer and Núñez (1990). As σ tends to zero, the Gaussian exponentials in Eq. (7) become very concentrated about $n = r - m$, and $F[r, \mu, \sigma]$ tends towards $r - m$, in which case the iteration (3) is that of Politte and Snyder (1991) for image recovery from Poisson-distributed data in the presence of nonuniform detector response and background counts. If, in addition to $\sigma = 0$, there hold $\beta(j) = 1$ and $\mu_0(j) = 0$ for all j, then (3) becomes the Richardson-Lucy iteration (Richardson 1972, Lucy 1974) and the Shepp-Vardi (1982) expectation-maximization algorithm for Poisson-distributed data.

2. Implementation

The computational burden of implementations of the iteration (3) has motivated the use of approximations for the function (7). However, as we have commented, there is a potential for errors introduced by approximations to be magnified when realizing solutions to ill-posed stochastic inverse-problems. For this reason, we have studied two implementations of (3). The first, which is already in use for restoring HST images, relies on an approximation to $F[r, \mu, \sigma]$ obtained by substituting a Poisson variate for the read-out noise $g(j)$ in Eq. (1). The second relies on an evaluation of $F[r, \mu, \sigma]$ via a saddlepoint approximation.

We describe these two approaches in the remainder of this section. Results we have obtained using them are presented and compared in the next section.

2.1. Implementation via Poisson Approximation

For this approximation, we follow the discussion given by Snyder, Hammoud, and White (1993, p. 1017). The read-out noise $g(j)$ from the jth pixel is a Gaussian-distributed random variable with mean m and standard deviation σ. If the data $r(j)$ are modified to form $r(j) - m + \sigma^2$, thereby converting the read-out noise to $g(j) - m + \sigma^2$, which is Gaussian distributed with mean and variance equal to σ^2, and if σ^2 is large, then according to Feller (1968), $g(j) - m + \sigma^2$ is approximately equal in distribution to a Poisson-distributed random variable with mean σ^2. With this approximation, the CCD data (1) become

$$r(j) - m + \sigma^2 = n_{\text{obj}}(j) + n_0(j) + n_{\text{read-out}}(j), \quad j = 0, 1, \cdots, J - 1 \tag{9}$$

where $n_{\text{read-out}}(j)$ is Poisson distributed with mean σ^2. Under this approximation, the data $\{r(j) - m + \sigma^2, j = 0, 1, \cdots, J - 1\}$ form a Poisson process with mean-value function

$$\left\{\mu_{\text{obj}}(j) + \mu_0(j) + \sigma^2, \quad j = 0, 1, \cdots, J - 1\right\} .$$

As there is now no Gaussian read-out noise, the iteration (3) becomes

$$\hat{\lambda}^{\text{new}}(i) = \hat{\lambda}^{\text{old}}(i)\frac{1}{\bar{\beta}(i)}\sum_{j=0}^{J-1}\left[\frac{\beta(j)p(j|i)}{\beta(j)\sum_{i=0}^{I-1}p(j|i)\hat{\lambda}^{\text{old}}(i)+\mu_0(j)+\sigma^2}\right]\left[r(j)-m+\sigma^2\right], \quad (10)$$

which is the Richardson-Lucy iteration modified to accommodate nonuniformity of flat-field detector-response and background noise. This is equivalent to making the following approximation to $F[r,\mu,\sigma]$ in Eq. (7) when the standard deviation of the read-out noise is large:

$$F[r,\mu,\sigma] \approx \frac{\mu}{\mu+\sigma^2}\left(r-m+\sigma^2\right). \quad (11)$$

We note also from Eq. (7) that as σ tends to zero, $F[r,\mu,\delta]$ tends towards $r-m$, suggesting that the approximation (11) for large values of σ will be a good choice for small values of σ as well.

2.2. Implementation via Saddlepoint Approximation

An alternative to the Poisson approximation is to evaluate the function (7) as needed for use in Eq. (3). However, this function depends on r, μ, and σ in a complicated way, so some numerical method of evaluation is necessary. We have used the saddlepoint integration and saddlepoint approximation methods introduced by Helstrom (1978) for calculating tail probabilities.

The Laplace transform of the Poisson-Gaussian-mixture density (8) is

$$L[p(\cdot:\mu,\sigma)](s) = \int_{-\infty}^{\infty} p(x:\mu,\sigma)e^{-sx}dx = \exp[\mu(e^{-s}-1)+\frac{1}{2}\sigma^2 s^2], \quad (12)$$

and, by the inverse Laplace transform,

$$p(x:\mu,\sigma) = \frac{1}{2\pi j}\int_{c-j\infty}^{c+j\infty}\exp[\Phi(s)]ds, \quad (13)$$

where

$$\Phi(s) = xs+\frac{1}{2}\sigma^2 s^2+\mu(e^{-s}-1). \quad (14)$$

The saddlepoint integration and saddlepoint approximation methods can be used with (13) to produce accurate evaluations of the function $F[r,\mu,\sigma]$ in (7). The saddlepoint $\tilde{s}$, which is real, is the root of

$$\Phi'(\tilde{s}) = \frac{d\Phi(s)}{ds}|_{s=\tilde{s}} = x+\sigma^2\tilde{s}-\mu e^{-\tilde{s}} = 0. \quad (15)$$

This root can be determined numerically using a Newton iteration,

$$\tilde{s}^{\text{new}} = \tilde{s}^{\text{old}} - \frac{\Phi'(\tilde{s}^{\text{old}})}{\Phi''(\tilde{s}^{\text{old}})}, \quad (16)$$

where $\Phi''(\tilde{s}) = \sigma^2+\mu e^{-\tilde{s}}$. We iterate Eq. (16), from starting values described below, until $\left|(\tilde{s}^{\text{new}}-\tilde{s}^{\text{old}})/\tilde{s}^{\text{new}}\right| < \epsilon$, where we have used $\epsilon = 10^{-6}$ as the tolerance parameter. The saddlepoint approximation to Eq. (13) is

$$p(x:\mu,\sigma) \approx \frac{1}{\sqrt{2\pi\Phi''(\tilde{s})}}e^{\Phi(\tilde{s})}. \quad (17)$$

Starting Values Useful starting values for the Newton iteration (16) can be obtained by making approximations to the exponential in Eq. (15). For example, replacing the exponential by two terms of its Taylor series,

$$e^{-\tilde{s}} \approx 1 - \tilde{s} \,, \tag{18}$$

and then solving $x + \sigma^2\tilde{s} - \mu(1 - \tilde{s}) = 0$ suggests using

$$\tilde{s}^{\text{old}} = \frac{\mu - x}{\mu + \sigma^2} \tag{19}$$

as a starting value. Although this is adequate for some values of x and μ, for low values of x combined with large values of μ, this resulted in starting values so far from the saddlepoint that hundreds of Newton iterations were required for convergence (see Table 1). We have developed two alternative approaches for establishing more efficient starting values. One of these is obtained via the Padé approximation (Newton, Gould, and Kaiser 1961):

$$e^{-\tilde{s}} \approx \frac{1 - \frac{1}{2}\tilde{s}}{1 + \frac{1}{2}\tilde{s}} \,. \tag{20}$$

Solving this for $\tilde{s}$ in terms of $e^{-\tilde{s}}$ and substituting the result into Eq. (15) suggests using

$$\tilde{s}^{\text{old}} = -\ln\left[\frac{(x - 2\sigma^2 - \mu) + \sqrt{(x - 2\sigma^2 - \mu)^2 + 4\mu\,(x + 2\sigma^2)}}{2\mu}\right] \tag{21}$$

as the starting value, where the sign of the radical is selected to make the argument of the logarithm positive. Although the Padé approximation (20) is inaccurate for $|\tilde{s}| > 2$, we have found that the Newton iterations (16) started with (21) converge, to within a tolerance parameter $\epsilon = 10^{-6}$, in no more than five iterations for x and μ ranging from 10^{-1} to 10^5 for both $\sigma = 3$ and $\sigma = 13$ (see Table 1). For this reason, we have generally used starting values based on this approximation in all of our studies for a read-out noise level of $\sigma = 13$ electrons. Another alternative is to use a small number of Picard iterations

$$\tilde{s}^{\text{new}} = \ln\left[\frac{\mu}{x + \sigma^2\tilde{s}^{\text{old}}}\right] \tag{22}$$

to solve Eq. (15). We start these iterations using (19) and only perform a small number of them, typically three, before converting to Newton iterations. Also, if the argument of the logarithm in (22) becomes negative during the Picard iterations, we revert to using (19) to start the Newton iterations. With this approach, no more than five Newton iterations were required for convergence to within a tolerance parameter of $\epsilon = 10^{-6}$ in evaluating either the numerator or denominator of (7), and often one or two sufficed.

Shown in Table 1 are the starting values and the numbers of Newton iterations required for convergence to a tolerance parameter of $\epsilon = 10^{-6}$ for each of the Taylor, Padé, and Picard methods for determining starting values for $\sigma = 13$ and a wide range of values of μ and $r - m$. While a variety of starting values and numbers of Newton iterations are seen, all three methods converged to the same saddlepoints to the degree of numerical precision shown in the table. Both the Padé and Picard methods required no more than 5 Newton iterations, while the Taylor method often required many more.

Accommodation of Wide Dynamic-Range Because of the wide dynamic range encountered in HST images, with values of μ ranging over several orders of magnitude, it is necessary to modify Eq. (7) to avoid overflow and underflow, writing it as

$$F[r, \mu, \sigma] = \mu \exp[\ln p(r - m - 1 : \mu, \sigma) - \ln p(r - m : \mu, \sigma)] \,, \tag{23}$$

Table 1. Comparison of starting values for Newton iterations.

$log_{10}\mu$	$log_{10}(r-m)$	Taylor: Starting	Newtons	Padé: Starting	Newtons	Picard: Starting	Newtons	sad-point
0	0	0.000e+00	1	-0.000e+00	1	0.000e+00	1	0.000e+00
0	1	-5.294e-02	2	-5.295e-02	2	-5.294e-02	2	-5.293e-02
0	2	-5.824e-01	2	-5.982e-01	3	-5.824e-01	2	-5.811e-01
0	3	-5.876e+00	5	-6.497e+00	5	-5.876e+00	5	-5.021e+00
0	4	-5.882e+01	54	-9.176e+00	4	-9.043e+00	2	-9.044e+00
0	5	-5.882e+02	581	-1.151e+01	3	-1.149e+01	2	-1.149e+01
1	0	5.028e-02	2	5.036e-02	2	5.396e-01	4	5.035e-02
1	1	0.000e+00	1	-0.000e+00	1	0.000e+00	1	0.000e+00
1	2	-5.028e-01	3	-5.042e-01	3	-5.028e-01	3	-4.947e-01
1	3	-5.531e+00	6	-4.208e+00	5	-3.354e+00	4	-3.647e+00
1	4	-5.581e+01	54	-6.873e+00	3	-6.785e+00	2	-6.786e+00
1	5	-5.586e+02	554	-9.207e+00	3	-9.195e+00	1	-9.195e+00
2	0	3.680e-01	3	3.970e-01	3	8.725e-01	4	3.934e-01
2	1	3.346e-01	3	3.582e-01	3	6.895e-01	4	3.555e-01
2	2	0.000e+00	1	-0.000e+00	1	0.000e+00	1	0.000e+00
2	3	-3.346e+00	6	-2.005e+00	4	-1.886e+00	3	-1.912e+00
2	4	-3.680e+01	37	-4.572e+00	3	-4.525e+00	2	-4.526e+00
2	5	-3.714e+02	369	-6.904e+00	3	-6.896e+00	1	-6.896e+00
3	0	8.546e-01	5	1.523e+00	4	1.661e+00	4	1.422e+00
3	1	8.469e-01	5	1.495e+00	4	1.615e+00	4	1.400e+00
3	2	7.699e-01	4	1.248e+00	3	1.279e+00	4	1.197e+00
3	3	0.000e+00	1	-0.000e+00	1	0.000e+00	1	0.000e+00
3	4	-7.699e+00	10	-2.275e+00	3	-2.264e+00	2	-2.264e+00
3	5	-8.469e+01	85	-4.602e+00	3	-4.597e+00	1	-4.597e+00
4	0	9.833e-01	6	3.448e+00	4	3.096e+00	4	2.985e+00
4	1	9.824e-01	6	3.422e+00	4	3.077e+00	4	2.972e+00
4	2	9.735e-01	6	3.191e+00	4	2.908e+00	3	2.846e+00
4	3	8.850e-01	6	2.070e+00	3	2.013e+00	3	2.010e+00
4	4	0.000e+00	1	-0.000e+00	1	0.000e+00	1	0.000e+00
4	5	-8.850e+00	11	-2.300e+00	2	-2.299e+00	1	-2.299e+00
5	0	9.983e-01	8	5.694e+00	5	4.871e+00	3	4.811e+00
5	1	9.982e-01	8	5.667e+00	5	4.860e+00	3	4.802e+00
5	2	9.973e-01	8	5.437e+00	5	4.754e+00	3	4.714e+00
5	3	9.883e-01	8	4.321e+00	4	4.084e+00	3	4.081e+00
5	4	8.985e-01	6	2.275e+00	3	2.265e+00	2	2.265e+00
5	5	0.000e+00	1	-0.000e+00	1	0.000e+00	1	0.000e+00

where in the saddlepoint approximation

$$\ln p(x:\mu,\sigma) \approx \Phi(\tilde{s}) - \frac{1}{2}\ln[2\pi\Phi''(\tilde{s})]\,, \tag{24}$$

Since the factors of 2π cancel in determining (7), we omit these in our computations.

Accuracy of the Approximations The saddlepoint integration of Eq. (13) is performed by integrating along the vertical line $c = \tilde{s} + jy$, $-\infty < y < \infty$, through the saddlepoint to obtain

$$p(x:\mu,\sigma) = \frac{1}{\pi}e^{-\mu}\int_0^\infty Re\left\{\exp\left[x(\tilde{s}+jy) + \frac{1}{2}(\tilde{s}+jy)^2 + \mu e^{-(\tilde{s}+jy)}\right]\right\} dy\,. \tag{25}$$

The trapezoidal rule was used, taking the step size as $\Delta y = [\Phi''(\tilde{s})]^{-1/2}$ and stopping the summation when the absolute value of the integrand fell below $\delta\Delta y$ times the accumulated sum, with $\delta = 10^{-5}$. Smaller step-sizes did not noticeably alter the results, which were also checked by evaluating Eq. (7) with the summation in Eq. (8); the results agreed to at least six significant figures. The discrepancies may be attributable to round-off error in the summations in Eq. (8).

Shown in Table 2 are values of $F[r:\mu,\sigma]$ obtained for the methods of saddlepoint integration (Eq. 25), saddlepoint approximation (Eq. 17), and Poisson approximation (Eq. 11) for $\sigma^2 = 200$ and several values of $r-m$ and μ. One purpose of this table is to compare "exact" values obtained with saddlepoint integration to values obtained with the saddle-point and Poisson approximations.

Starting values for the Newton iteration were determined by the Picard method. The saddlepoint approximation is seen to agree closely with the results of saddlepoint integration. If we take the values of $F[r : \mu, \sigma]$ obtained in Table 2 by saddlepoint integration as "exact," then the sum of absolute errors for the saddlepoint and Poisson approximations differ by several orders of magnitude, being 0.0004 and 19.81, respectively.

Table 2. Evaluations of $F(\cdot)$.

σ^2	μ	$r-m$	sp-int	sp-approx	P-approx
200	20	-55.0	14.122700	14.122702	13.181818
200	20	-25.0	16.236131	16.236133	15.909091
200	20	5.0	18.638897	18.638900	18.636364
200	20	35.0	21.362958	21.362960	21.363636
200	20	65.0	24.441810	24.441813	24.090909
200	20	95.0	27.910075	27.910078	26.818182
200	200	100.0	153.126879	153.126854	150.000000
200	200	140.0	171.028212	171.028184	170.000000
200	200	180.0	190.001082	190.001050	190.000000
200	200	220.0	209.998996	210.998963	210.000000
200	200	260.0	230.972066	230.972031	230.000000
200	200	300.0	252.868626	252.868590	250.000000
200	500	350.0	396.313041	396.313010	392.857143
200	500	400.0	430.003993	430.003964	428.571429
200	500	450.0	464.555108	464.555081	464.285714
200	500	500.0	499.897950	499.897926	500.000000
200	500	550.0	535.970081	535.970060	535.714286
200	500	600.0	572.714680	572.714662	571.428571
200	500	650.0	610.080121	610.080105	607.142857

Shown in Table 3 are values of $F[r : \mu, \sigma]$ obtained for the methods of saddlepoint approximation, using Eqs. (23) and (24), and Poisson approximation (Eq. 11) with $\sigma = 13$, corresponding to the level of read-out noise present with the HSTs CCD camera, and $\sigma = 3$, corresponding to possible values of read-out noise present with improved CCD cameras, for a wide range of values of r and μ. The purpose of this table is to compare values obtained with the saddlepoint and Poisson approximations over a dynamic range encountered in HST image data. In this table, $|\Delta F|$ is the magnitude of the value of the difference in values of F determined by the saddlepoint and Poisson approximations.

The saddlepoint and Poisson approximations are seen from these tables to produce practically identical results for $\mu \approx r - m$. Elsewhere, the Poisson result is consistently less than the saddlepoint result, with the difference worsening as $r - m$ strays further and further from μ. For $r - m < \mu$, the error tends to level off to some value, whereas for $r - m > \mu$, the error rises rapidly and somewhat proportionately to $r - m$. The percentage error can be quite large, especially when $r - m$ is much larger than μ. For instance, for $\mu = 1$, $\sigma = 13$ and $r - m = 10^5$, $|\Delta F|/$(sp-approx)=99.4%.

In performing Newton iterations to evaluate $F(r, \mu, \sigma)$ in Eq. (7), the saddlepoint for the denominator was evaluated first, and the saddlepoint then determined was used as the starting value for Newton iterations in evaluating the numerator. Except when x was close to μ, the number of Newton iterations for the numerator was then no larger than for the denominator and was usually smaller.

2.3. Computer Implementation

Serial Implementation A C subroutine to compute $F[r, \mu, \sigma]$ using the saddlepoint approximation with a Newton iteration to determine the saddlepoint is given in the Appendix for starting values obtained with either the Padé or Picard methods. With $\mu = 10^2$ and $r - m = 10^3$, which are typical values, 2^{16} calls of these subroutines take approximately 16 seconds on a SUN Sparcstation IPC and 8 seconds on a SUN Sparcstation IPX. These times can probably be lowered substantially while maintaining nearly the same accuracy by using a larger value for the tolerance parameter ϵ.

Table 3. More Evaluations of $F(\cdot)$.

$log_{10}\mu$	$log_{10}(r-m)$	sp-approx ($\sigma = 13$)	P-approx ($\sigma = 13$)	$\lvert\Delta F\rvert$ ($\sigma = 13$)	sp-approx ($\sigma = 3$)	P-approx ($\sigma = 3$)	$\lvert\Delta F\rvert$ ($\sigma = 3$)
0	0	9.97e-01	1.00e+00	2.92e-03	9.56e-01	1.00e+00	4.43e-02
0	1	1.05e+00	1.05e+00	1.66e-03	2.26e+00	1.90e+00	3.60e-01
0	2	1.78e+00	1.58e+00	2.01e-01	6.27e+01	1.09e+01	5.18e+01
0	3	1.51e+02	6.88e+00	1.45e+02	9.38e+02	1.01e+02	8.37e+02
0	4	8.47e+03	5.98e+01	8.41e+03	9.92e+03	1.00e+03	8.92e+03
0	5	9.81e+04	5.89e+02	9.75e+04	9.99e+04	1.00e+04	8.99e+04
1	0	9.48e+00	9.50e+00	1.34e-02	5.72e+00	5.26e+00	4.57e-01
1	1	9.97e+00	1.00e+01	2.63e-02	9.87e+00	1.00e+01	1.26e-01
1	2	1.64e+01	1.50e+01	1.33e+00	8.11e+01	5.74e+01	2.37e+01
1	3	3.84e+02	6.53e+01	3.18e+02	9.59e+02	5.31e+02	4.28e+02
1	4	8.85e+03	5.68e+02	8.29e+03	9.94e+03	5.27e+03	4.67e+03
1	5	9.84e+04	5.60e+03	9.29e+04	9.99e+04	5.26e+04	4.73e+04
2	0	6.74e+01	6.32e+01	4.18e+00	1.69e+01	9.17e+00	7.68e+00
2	1	7.00e+01	6.65e+01	3.44e+00	2.31e+01	1.74e+01	5.63e+00
2	2	9.99e+01	1.00e+02	1.17e-01	1.00e+02	1.00e+02	3.75e-02
2	3	6.77e+02	4.35e+02	2.42e+02	9.79e+02	9.26e+02	5.38e+01
2	4	9.24e+03	3.78e+03	5.45e+03	9.96e+03	9.18e+03	7.76e+02
2	5	9.88e+04	3.72e+04	6.16e+04	9.99e+04	9.18e+04	8.19e+03
3	0	2.41e+02	1.45e+02	9.57e+01	3.19e+01	9.91e+00	2.20e+01
3	1	2.46e+02	1.53e+02	9.34e+01	3.91e+01	1.88e+01	2.03e+01
3	2	3.02e+02	2.30e+02	7.20e+01	1.19e+02	1.08e+02	1.11e+01
3	3	1.00e+03	1.00e+03	6.18e-02	1.00e+03	1.00e+03	4.33e-03
3	4	9.62e+03	8.70e+03	9.19e+02	9.98e+03	9.92e+03	5.96e+01
3	5	9.92e+04	8.57e+04	1.35e+04	1.00e+05	9.91e+04	8.42e+02
4	0	5.05e+02	1.67e+02	3.38e+02	4.88e+01	9.99e+00	3.88e+01
4	1	5.12e+02	1.76e+02	3.36e+02	5.65e+01	1.90e+01	3.75e+01
4	2	5.81e+02	2.65e+02	3.16e+02	1.38e+02	1.09e+02	2.96e+01
4	3	1.34e+03	1.15e+03	1.90e+02	1.02e+03	1.01e+03	1.24e+01
4	4	1.00e+04	1.00e+04	7.81e-03	1.00e+04	1.00e+04	0.00e+00
4	5	9.96e+04	9.85e+04	1.11e+03	1.00e+05	9.99e+04	6.02e+01
5	0	8.14e+02	1.70e+02	6.44e+02	6.67e+01	1.00e+01	5.67e+01
5	1	8.21e+02	1.79e+02	6.43e+02	7.47e+01	1.90e+01	5.57e+01
5	2	8.97e+02	2.69e+02	6.28e+02	1.58e+02	1.09e+02	4.90e+01
5	3	1.69e+03	1.17e+03	5.23e+02	1.04e+03	1.01e+03	3.22e+01
5	4	1.04e+04	1.02e+04	2.31e+02	1.00e+04	1.00e+04	1.26e+01
5	5	1.00e+05	1.00e+05	0.00e+00	1.00e+05	1.00e+05	0.00e+00

Parallel Implementation We have implemented the EM iteration (Eq. 3) on a DECmpp 12000Sx / Model 200 single-instruction multiple-data (SIMD) computer with a 64×64 processor array. Ideally, images would be mapped directly with one pixel per processor. Since, in these simulations, we perform computation on images of dimension 256×256 pixels, zero padded to 512×512, 64 pixels are stored per processor. Convolutions are computed with FFT algorithms tailored to take advantage of the SIMD architecture. The additions, multiplications, and division in (3) are performed for 64^2 pixels in parallel. Both the Poisson and saddlepoint approximations (using Padé's approximation) have been implemented. Each processor computes values of F for its pixels independently of the other processors, which is ideal for SIMD architectures. For a 512×512 image (the data must be zero-padded so that the point spread does not wrap around), we can compute 50 EM-Poisson iterations in approximately 61 seconds and 50 EM saddlepoint iterations in approximately 81 seconds. The exact times for the EM saddlepoint iterations depends upon the number of Newton iterations required, which depends upon the data. If machine precision or other factors cause μ to be zero for some pixels, the denominator in Eq. (21) will be zero. In that case, we take the ratio F/μ in the EM algorithm (3) to be zero.

3. Comparisons of Implementations

In this section, we describe the results of experiments that we have performed to examine the effect on restored images of the Poisson and saddlepoint approximations to the function $F[r, \mu, \sigma]$.

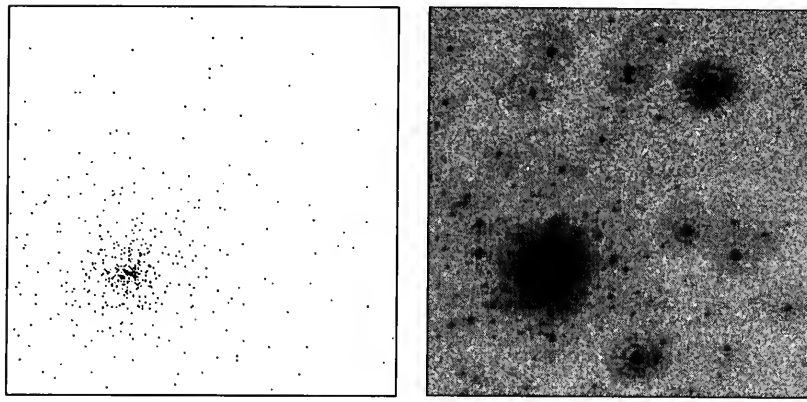


Figure 1. Simulated star cluster (left) and simulated HST image of star cluster (right).

3.1. Experiment 1: Restoration of Star Cluster Data Using 50 EM Iterations

We performed restorations for the simulated star cluster (file "truth.fit") provided by the Space Telescope Science Institute. The star cluster, shown in the left panel of Fig. 1, consists of a field of unresolved stars at random locations and having randomly selected brightnesses. We simulated an HST image of the star cluster using a space-invariant PSF (file "mpsf12.fit" provided by ST ScI) representative of the HSTs PSF, Poisson-distributed photo-counts, and a read-out noise level of $\sigma = 13$ electrons r.m.s. This is shown in the right panel of Fig. 1. In these and the following images, gray scale values are shown in a logarithmic scale so that low level noise and distortions can be seen.

The restorations obtained from the simulated data by performing 50 EM iterations (Eq. 3) based on the Poisson (Eq. 11) and saddlepoint (Eq. 23) approximations are shown in Fig. 2. It is difficult to compare these two restorations directly, so we have summarized one feature of them in Fig. 3 where shown are the ratios of estimated to true brightness versus true brightness for the restorations in Fig. 2. These scatter diagrams were obtained by placing a point at location (tb, eb/tb) for each star in the star cluster for restorations based on the Poisson (left) and saddlepoint (right) approximations. The solid lines were obtained by averaging the logarithmic values in the scatter diagrams with a rectangular window of width 0.1 moved in steps of 0.05.

The two solid lines in Fig. 3 are redrawn together in Fig. 4 so they can be compared more readily. It is evident from this graph that, for 50 EM iterations, the saddlepoint approximation tends to yield more accurate brightness estimates, especially for fainter stars. The difference does not appear to be enormous, but it is present, with the saddlepoint approximation yielding brightness estimates closer to truth than with the Poisson approximation. Recognizing that these are logarithmic plots, there is seen to be an improvement in the accuracy of the brightness estimate by a factor of about 1.5 for fainter objects, which can be significant when accurate photometry is sought.

3.2. Experiment 2: Averages for 25 Repeated Star Cluster Simulations, 50 Iterations Each

For this experiment, we simulated 25 independent realizations of HST images that would be acquired by the HST for the star cluster field. These images were produced in the same manner as the image in the right-hand panel of Fig. 1. We then restored each of the 25 images using 50 EM iterations for both the Poisson and saddlepoint approximations. Shown in Fig. 5 is the average ratio of estimated brightness to brightness versus brightness for the star cluster data; this graph was obtained by averaging the 25 versions of Fig. 4 obtained in the multiple simulations.

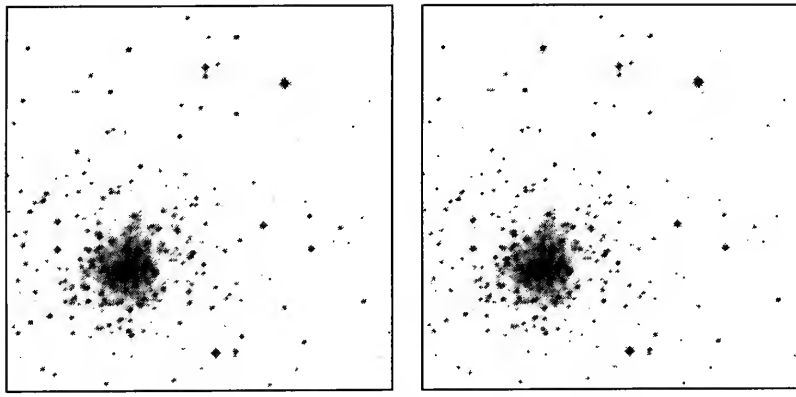


Figure 2. Restorations of simulated HST star cluster data using the Poisson (left) and saddlepoint (right) approximations and 50 EM iterations.

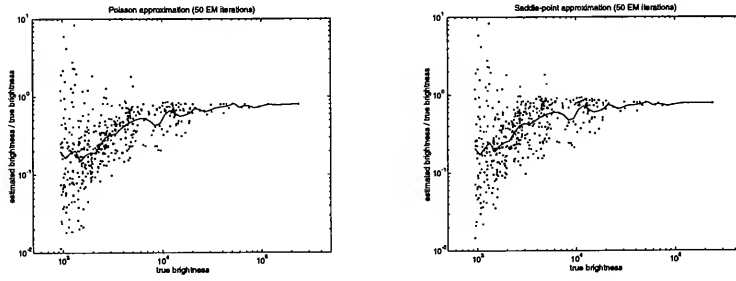


Figure 3. Ratio of estimated to true brightness vs. true brightness for simulated star cluster data. Obtained using 50 EM iterations and the Poisson (left) and saddlepoint (right) approximations.

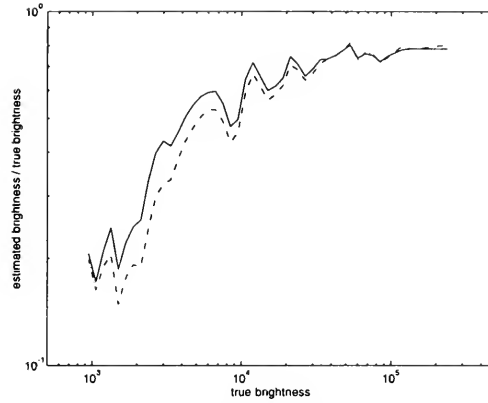


Figure 4. Smoothed ratio of estimated to true brightness vs. true brightness for star cluster data. Obtained from one restoration of a simulated HST star cluster image using 50 EM iterations and the Poisson (dashed line) and saddlepoint (continuous line) approximations.

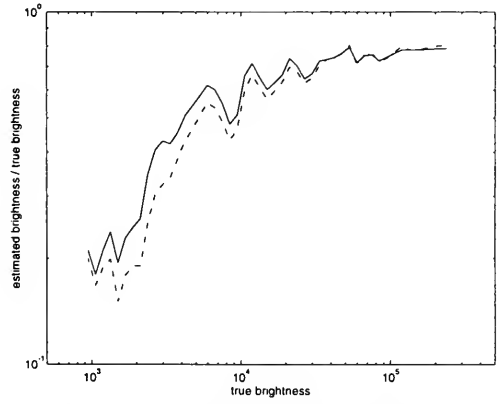


Figure 5. Smoothed ratio of estimated to true brightness vs. true brightness for star cluster data. Obtained by averaging 25 restorations of simulated HST star cluster images using 50 EM iterations each and the Poisson (dashed line) and saddlepoint (continuous line) approximations.

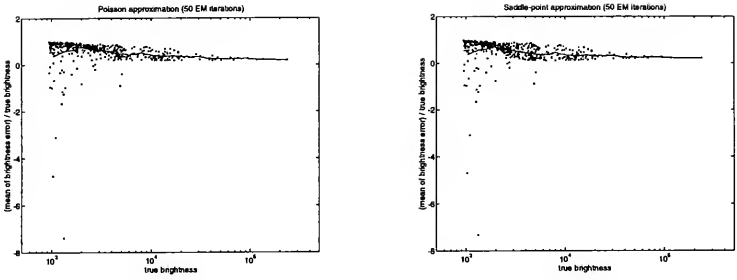


Figure 6. Relative bias in estimating brightness with the Poisson and saddlepoint approximations for a simulated star cluster field. Obtained by averaging errors in 25 simulations with restorations performed using 50 EM iterations each.

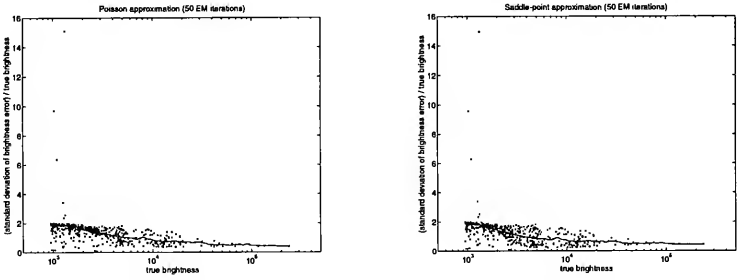


Figure 7. Relative standard deviation of error in estimating brightness with the Poisson and saddlepoint approximations for a simulated star cluster field. Obtained by averaging errors in 25 simulations with restorations performed using 50 EM iterations each.

Shown in Fig. 6 is the ratio of the mean error to brightness versus brightness for the two approximations. These scatter plots were obtained by averaging the error (estimated brightness minus true brightness) across the 25 restorations, dividing by the true brightness, and creating a point at (tb, error/tb) for each star in the star cluster. Shown in Fig. 7 is a corresponding scatter plot showing the ratio of standard deviation to brightness versus brightness. The solid lines in Figs. 6 and 7 were obtained by averaging the values in the scatter diagrams with a rectangular window of width 0.2 moved in steps of 0.01. These are replotted in Fig. 8 to facilitate comparisons.

3.3. Experiment 3: Restoration of Star Cluster Data Using 500 EM Iterations

Shown in Fig. 9 are the results for the same conditions as for Fig. 3 but with 500 EM iterations having been performed. When these extended numbers of iterations are performed, we find less scatter for fainter stars compared to restorations performed with 50 EM iterations, and there is less bias for all brightness levels. We also find at 500 EM iterations that the Poisson and saddlepoint approximations yield virtually identical results so far as these scatter diagrams are concerned. However, these scatter diagrams only reflect the restored images at the true star locations. Restored images of the star cluster for 500 iterations are in Fig. 10, which is the result of performing 500 EM iterations on one simulated HST star cluster image. These images are shown on a logarithmic scale, so low level noise is also evident, but this is noise is small, typically at a level of about 50 compared to stars that are at a level several orders of magnitude larger (about 10^4). The low level

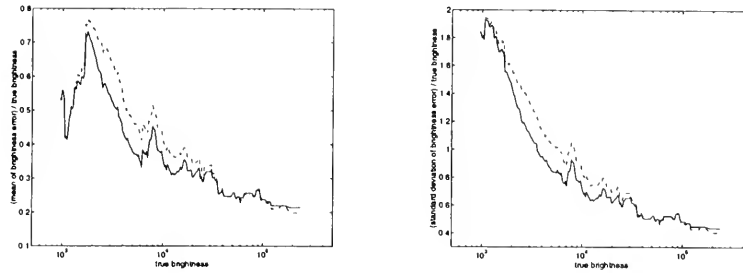


Figure 8. Mean and standard deviation of error for the saddlepoint (solid line) and Poisson (dashed line) approximations. Obtained by averaging restorations of 25 simulated HST images of the star cluster, each restored with 50 EM iterations.

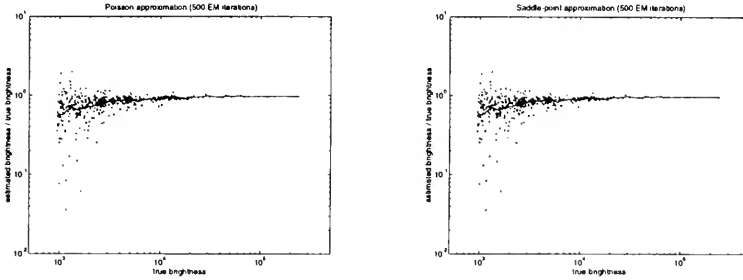


Figure 9. Ratio of estimated to true brightness vs. true brightness for simulated star cluster data. Obtained using 500 EM iterations and the Poisson (left) and saddlepoint (right) approximations.

noise is reduced substantially in Fig. 11, which is the average of 25 restorations of HST star cluster images, each being the result of performing 500 EM iterations. The average results shown for the Poisson and saddlepoint approximations are virtually identical at 500 EM iterations and low level noise is greatly reduced.

4. Conclusions

Use of the Poisson approximation of the function $F(r, \mu, \sigma)$ that accounts for read-out noise does not introduce significant errors in the restoration of HST images. However, some improvement in photometric performance can be obtained at the expense of computation time by a more accurate evaluation of this function. The improvement is seen most for faint objects when 50 iterations of the expectation-maximization algorithm are used to perform the restorations. For star cluster fields, performing more iterations of the expectation-maximization algorithm yields significantly improved restorations, and the Poisson approximation is not improved noticeably by a more accurate approximation of the function.

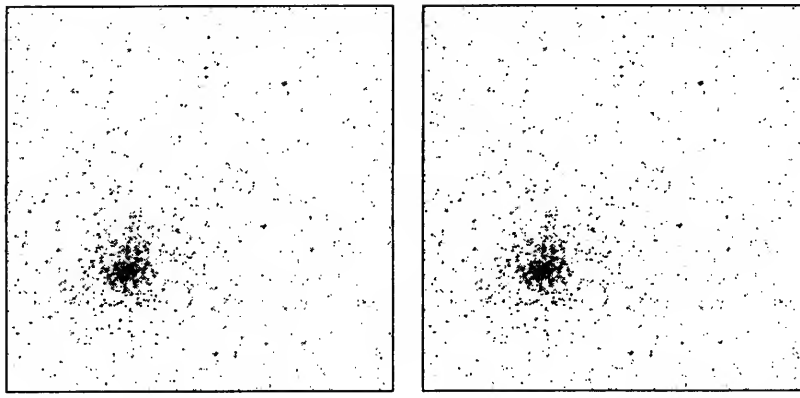


Figure 10. Restorations of one simulated HST star cluster image obtained using 500 EM iterations and the Poisson (left) and saddlepoint (right) approximations.

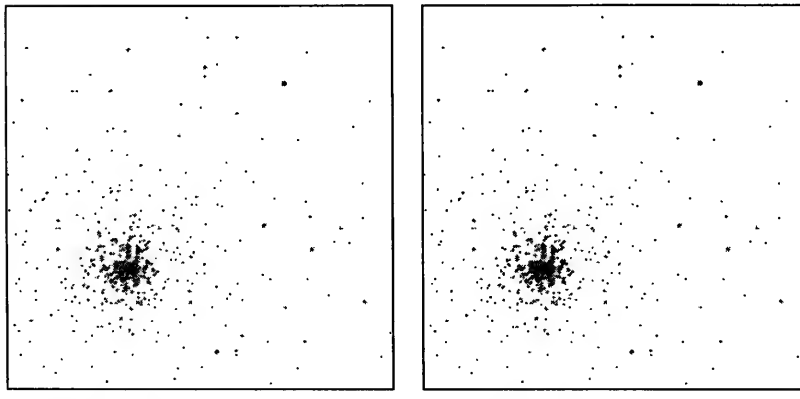


Figure 11. Average of restorations of 25 simulated HST star cluster images obtained using 500 EM iterations each and the Poisson (left) and saddlepoint (right) approximations.

Appendix: C Subroutine for Saddlepoint Method

```
#include <stdio.h>
#include <math.h>

#define EPSILON 0.000001

/* Number of Picard iterations to use */
#define NUM_PICARD 3

float the_function();
double sp_approx(), newton_iteration(), pade(), picard();

/* As listed below, the_function uses the Pade approximation.  To use */
/* the Picard iteration, change the "pade" to "picard" in the code    */
/* for the_function                                                   */

/* Function: the_function                                          */
/* Evaluates F given v, mu, and sigmasquared, where v = r - m */
/* Returns evaluation of F                                         */
/* Remember the normalizing 2*PI factor has been left out     */

float the_function(v, mu, sigmasquared)
float v, mu, sigmasquared;
{
        double log_q0, log_q1, saddlepoint;
        float v_minus_1;

        v_minus_1 = v - 1.0;
        saddlepoint = pade(v,mu,sigmasquared);
        saddlepoint = newton_iteration(v,mu,sigmasquared,saddlepoint);
        log_q0 = sp_approx(v,mu,sigmasquared,saddlepoint);
        saddlepoint = newton_iteration(v_minus_1,mu,sigmasquared,saddlepoint);
        log_q1 = sp_approx(v_minus_1,mu,sigmasquared,saddlepoint);
        return((float) (mu * exp(log_q1 - log_q0)));
}  /* end the_function */

/* Function: sp_approx                                             */
/* Returns the saddlepoint approximation to the the log of    */
/* p(x,mu,sigmasquared) given the saddlepoint found by the    */
/* Newton iteration.  Remember 2*PI factor has been left out. */

double sp_approx(x,mu,sigmasquared,saddlepoint)
float x, mu, sigmasquared;
double saddlepoint;
{
double mu_exp_minus_s, phi2;

mu_exp_minus_s = mu * exp (-saddlepoint);
phi2 = sigmasquared + mu_exp_minus_s;
return(-mu + (saddlepoint * (x + 0.5 * sigmasquared * saddlepoint))
                + mu_exp_minus_s - 0.5 * log(phi2));
}  /* end sp_approx */

/* Function: newton_iteration                                      */
/* Returns the saddlepoint found by Newton iteration for a given */
/* x, mu, sigmasquared and an initial esimate of the saddlepoint */
/* (found with either the Pade or Picard approaches               */

double newton_iteration(x, mu, sigmasquared, initial_saddlepoint)
float x, mu, sigmasquared;
double initial_saddlepoint;
{
double mu_exp_minus_s, saddlepoint, change;

saddlepoint = initial_saddlepoint;
do {
        mu_exp_minus_s = mu * exp(-saddlepoint);
        change = (x + sigmasquared * saddlepoint - mu_exp_minus_s)
                / (sigmasquared + mu_exp_minus_s);
        saddlepoint -= change;
} while((fabs(change) > EPSILON * fabs(saddlepoint)));
return(saddlepoint);
}  /* end newton_iteration */
```

```
/* Function: pade                                                    */
/* Returns the initial saddlepoint estimated by the Pade approx. */

double pade(x, mu, sigmasquared)
float x, mu, sigmasquared;
{
double bterm;

bterm = x - 2*sigmasquared - mu;
return(-log(0.5 * (bterm
        + sqrt(bterm*bterm + 4*mu*(2*sigmasquared + x))) / mu));
}  /* end pade */

/* Function: picard                                                  */
/* Returns the initial saddlepoint estimated by the Picard iter. */

double picard(x, mu, sigmasquared)
float x, mu, sigmasquared;
{
int i;
double argument_to_log, saddlepoint, taylor;

/* Use Taylor approx. to get starting point for Picard iteration. */
saddlepoint = taylor = (mu - x) / (mu + sigmasquared);
for (i = 0; i < NUM_PICARD; i++)
{
        argument_to_log = mu / (x + sigmasquared * saddlepoint);
        if (argument_to_log <= 0.0)       /* Break out of loop if */
                return(taylor);           /* arg. to log goes neg.*/
        else saddlepoint = log(argument_to_log);
}
return(saddlepoint);
}  /* end picard */
```

Acknowledgments. This work was supported in part by the National Science Foundation under grant number MIP–9101991.

References

Feller, W. 1968, *Introduction to Probability Theory and Its Applications*, Wiley, New York, pp. 190, 245

Helstrom, C. 1978, IEEE Trans. Aero. Elec. Sys., AES–14, 630

Llacer, J. & Núñez 1990, in *The Restoration of Hubble Space Telescope Images*, R. L. White & R. J. Allen, eds., Space Telescope Science Institute, Baltimore, 62

Lucy, L. 1974, AJ, 79, 745

Newton, G. C., Gould, L. A., & Kaiser, J. F. 1961, *Analytical Design of Linear Feedback Controls*, Wiley, New York, 239

Politte, D. G. & Snyder, D. L. 1991, IEEE Trans. Med. Imaging, 10, 82

Richardson, W. H. 1972, J. Opt. Soc. Am. A, 62, 55

Shepp, L. A. & Vardi, Y. 1982, IEEE Trans. Med. Imaging, MI–1, 113

Snyder, D. L., Hammoud, A. M., & White, R. L. 1993, J. Opt. Soc. Am. A, 10, 1014

PSFs, Phase Retrieval, and Blind Deconvolution

The Restoration of HST Images and Spectra II
Space Telescope Science Institute, 1994
R. J. Hanisch and R. L. White, eds.

Effect of OTA Breathing on Hubble Space Telescope Images

H. Hasan and P. Y. Bely[1]

Space Telescope Science Institute, 3700 San Martin Dr., Baltimore, MD 21218

Abstract. The effect of short term focus changes on Hubble Space Telescope images after the refurbishment mission is examined. Based on simulations we conclude that image quality will not be degraded significantly for the cameras if focus variations are within ± 5 microns in the blue and within ± 2 microns in the UV. The tolerance for the spectrographs is wider, about ± 10 microns.

1. Introduction

Short term focus changes caused by small ($\sim$ 5 microns) motions of the secondary mirror of the Optical Telescope Assembly (OTA) of the Hubble Space Telescope (HST) are referred to as OTA "breathing". Though the cause of the secondary mirror movement is not fully understood, it is now apparent that the motion is thermally induced (Hasan and Bely 1993). Secondary mirror variations, ΔSM, are related to the mean temperature of four sensors in the light shield around the secondary mirror, denoted LS, and the mean of LS over the previous orbit, MLS, by the equation

$$\Delta \mathrm{SM} = 0.7(\mathrm{LS} - \mathrm{MLS}) + K \qquad (1)$$

where K is an arbitrary constant.

OTA breathing may be only a minor nuisance for now, but could become more serious once the spherical aberration has been corrected with COSTAR and WFPC II. These focus changes can affect the quality of the science data in many ways. The most obvious effect is loss of sensitivity resulting from image blur. However, there are many other ramifications such as the difficulty in performing image enhancement with an unstable point spread function (PSF), or the impossibility of doing photometry with the cameras or the spectrographs. There are also secondary effects due to focal plane shifts induced by the fine guidance sensors (FGS) reacting to focus variation (when guiding in fine lock).

These effects can be very subtle and certainly depend on the type of observation (limiting detection, crowded fields, etc.), and complex simulations may be required in order to pinpoint the level at which a focus change becomes intolerable. In order to make a first order estimate of this "pain level" we have determined the PSF at various wavelengths and focus values using the following criteria:

- Imaging: Strehl ratio not degraded by more than 20%.
- Spectroscopy with small aperture: not more than 10% energy loss.
- Spectroscopy with large aperture: not more than 1% energy loss (in order to maintain adequate photometric calibration).

[1] Affiliated to Astrophysics Division, Space Science Department, European Space Agency

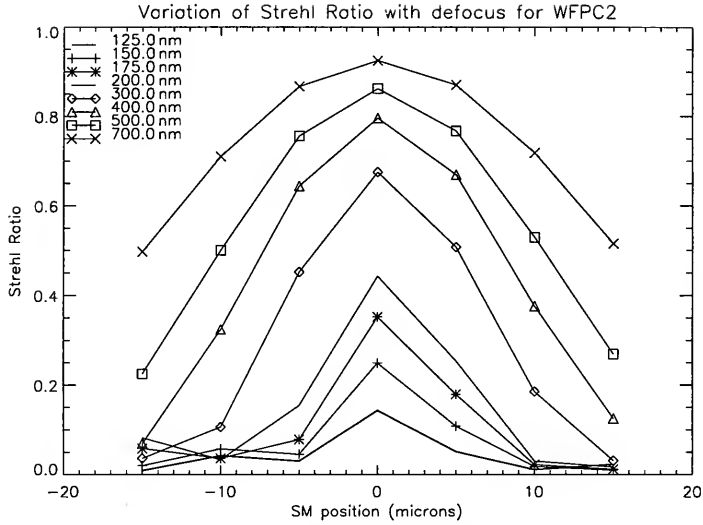


Figure 1. Variation of strehl ratio with defocus with WFPC2. Negative values of the secondary mirror (SM) position correspond to a movement towards the primary.

2. Simulations and Results

The point spread function (PSF) has been determined using the Telecope and Image Modelling (TIM) software (Hasan and Burrows 1994) with mirror wavefront errors (excluding spherical aberration) as measured pre-launch and a $\lambda/13$ wavefront error at 632.8 nm for the rest of the optics (secondary mirror residual misalignment, COSTAR and FOC/FOS/HRS or WFPC II internal wavefront errors). For the FOS an astigmatism term has been included to account for the off-axis position of the instrument. The PSF has been finely sampled on 5 milliarcseconds pixels. Although the simulations have been performed for WFPC II and FOS, the results are expected to apply to the FOC and GHRS, respectively.

The results are shown in Figs. 1 to 7. The variation of Strehl ratio with focus for eight wavelengths varying from 125 nm to 700 nm for the cameras is plotted in Fig. 1. For the same wavelength set, variation with focus of ensquared energies within 0.25×0.25 arcsec and 1×1 arcsec apertures is plotted in Figs. 2 and 3, respectively.

Fig. 4 gives the tolerable focus error for the cameras as a function of wavelength (using the 20% Strehl ratio degradation criterion), and Fig. 5 the tolerable focus error for the spectrographs. In the case of the spectrographs, the small aperture criterion is the only driver, the 1% degradation for the large aperture always being met even for very large focus errors.

For the cameras we have also studied the effect of focus change on the radius containing 80% of the energy as well as on the sharpness criterion S defined as

$$S = \sum_i P_i^2 / (\sum P_i)^2 \qquad (2)$$

where P_i is the intensity at pixel i. These are plotted in Figs. 6 and 7, respectively. Note that in Fig. 6 the best focus shifts for the 300 nm and 400 nm wavelengths. This is the result of mirror figure errors and the additional $\lambda/13$ (at 632 nm) aberration added to the system.

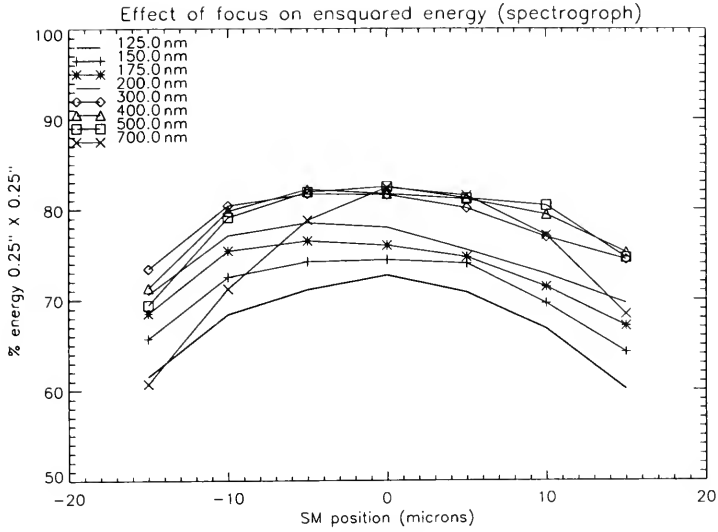


Figure 2. Variation of throughput in the $0''.25 \times 0''.25$ aperture of the spectrographs with defocus.

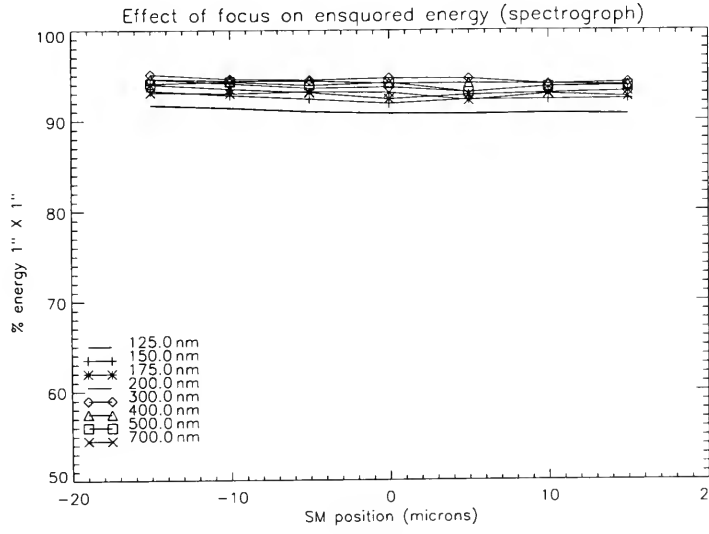


Figure 3. Variation of throughput in the $1'' \times 1''$ aperture of the spectrographs with defocus.

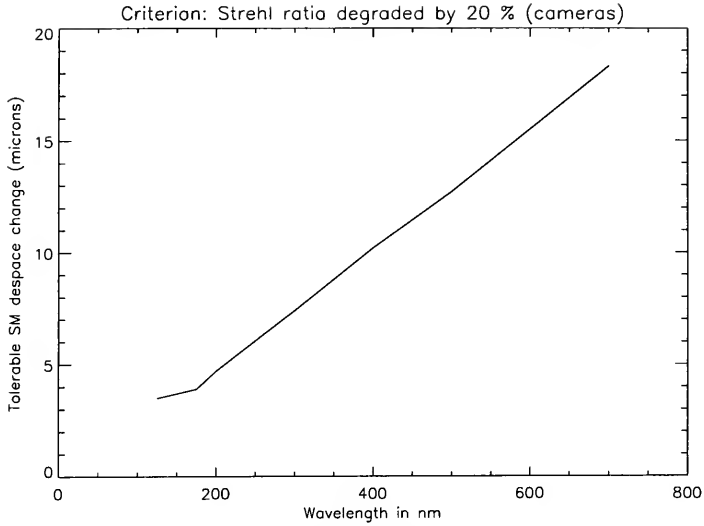


Figure 4. SM despace for which Strehl ratio degrades by 20% in the cameras as a function of wavelength.

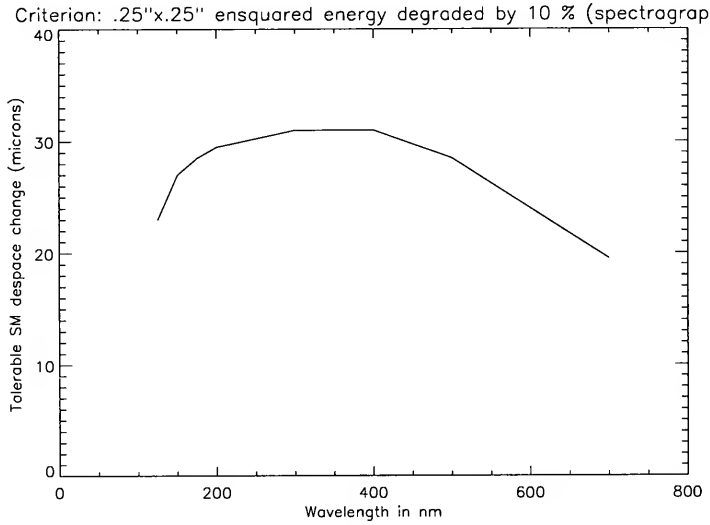


Figure 5. SM despace for which ensquared energy in the 0.″25× 0.″25 aperture in the spectrographs degrades by 10% as a function of wavelength.

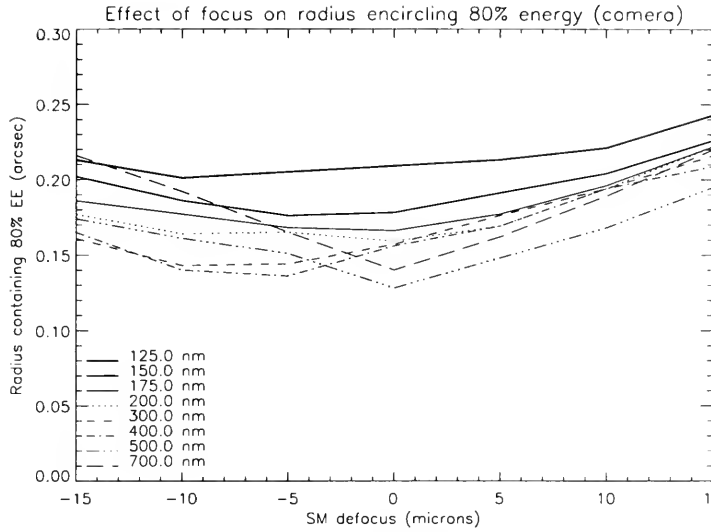


Figure 6. Variation of radius containing 80% encircled energy with defocus.

As expected, the most stringent focus requirement is set by the cameras. For them, the focus should not change by more than ± 2 microns when working in the far UV, and by not more than ± 5 microns in the blue. In terms of the radius encircling 80% of the energy, these tolerances roughly correspond to a radius increase of about 10%, while the sharpness criterion indicates an increase in exposure time from $\sim 4\%$ in the red to $\sim 35\%$ in the UV. The effect of focus changes on throughput in the large aperture in the spectrographs is barely noticeable, while the small apertures will tolerate a focus variation of about ± 10 microns without showing a degradation of throughput $> 10\%$.

It should be emphasized that these tolerances apply only to the variation of focus as seen by the instrument, and not to the pointing errors resulting from Fine Guidance Sensor (FGS) sensitivity to focus changes. The pointing error appears to be of order 20 mas, and may be more detrimental to science than the focus change itself, if left uncorrected. This pushes for establishing a deterministic model of the FGS pointing error in order to correct it via the Pointing Control System (PCS) (similar to the way corrections are done for velocity aberration or for moving targets).

References

Hasan, H. & Bely, P. Y. 1993, BAAS, 25, 1460

Hasan, H., & Burrows, C. J. 1994, *HST Calibration Workshop*, J. C. Blades, ed., Space Telescope Science Institute, Baltimore, in press

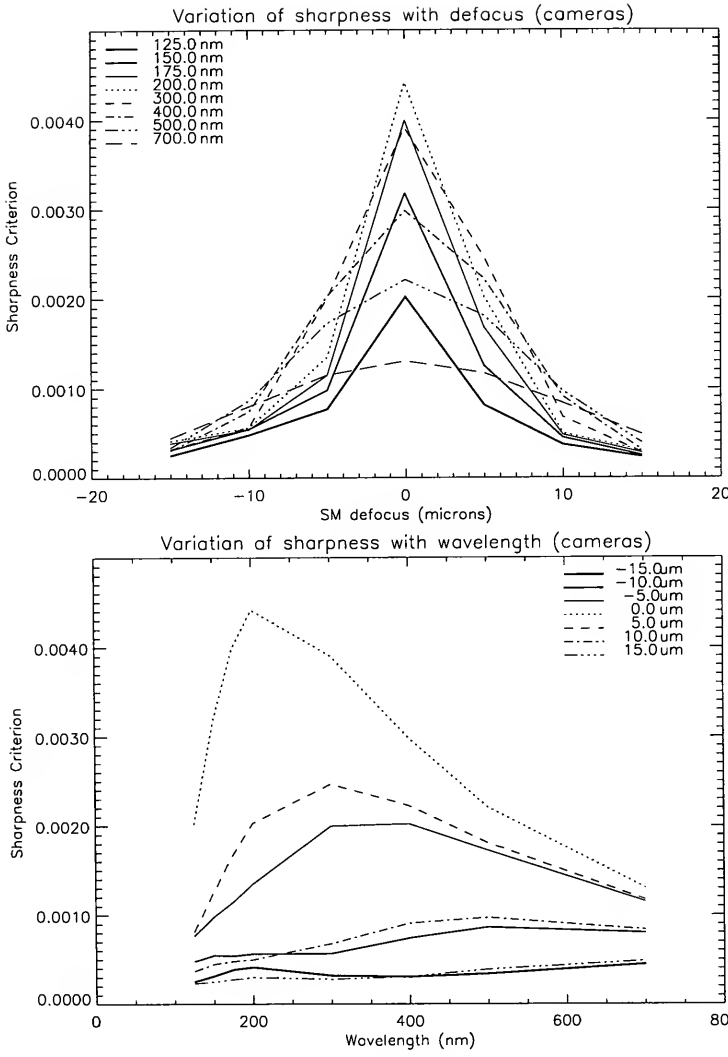


Figure 7. Variation of sharpness criterion for the cameras with (a) defocus, and (b) wavelength.

The Restoration of HST Images and Spectra II
Space Telescope Science Institute, 1994
R. J. Hanisch and R. L. White, eds.

Comparisons Between Observed and Model PSFs in WFPC Image Restoration

J. Mo and R. J. Hanisch

Space Telescope Science Insitute, Baltimore, MD 21218

Abstract. A demonstration of HST WFPC image restoration has been made using both empirical and simulated PSFs. Empirical PSFs are obtained from the WFPC PSF library and also extracted from the image to be restored. Simulated PSFs are generated by the Tiny TIM package. The best results obtained thus far rely upon being able to determine a high S/N empirical PSF from the data frame. In order to achieve accurate restorations of data with high dynamic range excellent PSF models are required. PSF models are less critical when the intrinsic source structure is smooth compared to the size of the PSF.

1. Introduction

The quality of the point spread function is the primary factor limiting the quality of image restoration for the HST. This has been demonstrated using several types of PSF for HST WFPC image restoration: (1) an empirical PSF extracted from the image to be restored, (2) an empirical PSF retrieved from the WFPC PSF library (Baggett and MacKenty 1992), and (3) a model PSF computed using the Tiny TIM software (Krist 1993). Similar results were obtained for HST WFPC images by Krist and Hasan (1993) who attribute the inaccuracies of the Tiny TIM models to incomplete knowledge of the HST primary mirror zonal error map.

Unfortunately, none of these methods yields an ideal PSF. The empirical PSFs contain noise and non-zero background levels. Frequently it is difficult to find a good PSF star in the image to be restored, and a PSF image that is well-enough exposed to show the details in the halo will often have a saturated core. Spoiler stars must be removed from the PSF image, a procedure which can itself introduce artifacts into the PSF. In using the WFPC PSF Library, care must be taken to find well-exposed images. Some images were taken with very short exposures where the effects of finite (and non-linear) shutter travel speed cannot be neglected. In addition, use of Library PSFs can introduce artifacts into the restoration owing to the known time variability of focus. Model PSFs do not reproduce all features of observed PSFs accurately, and short-term focus variability is not currently taken into account.

2. Data

2.1. Blurred Images

Tests have been carried out using three the WFPC images retrieved from the HST Archive.

R136 Observed with PC6, F555W, on 11 April 1991. This is a 256×256 subimage centered at pixel (253,638) in the original 800×800 pixel frame.

R136 Observed with PC6, F547M, on 25 August 1990. This is a 256×256 subimage centered at pixel (319,393) in the original 800×800 pixel frame.

M87 Observed with PC6, F547M, on 22 November 1991. The image is a 256×256 subimage which includes the nucleus of M87 as well as the famous jet.

2.2. PSFs

The Library PSFs, Image PSFs, and Tiny TIM PSFs are used for image restoration in this experiment.

Library PSFs The PSFs retrieved from the WFPC PSF Library cover regions of 200×200 pixels (8.″8 square). The available PSFs include one for F547M centered at position (337,328) and three for F555W centered at positions (138,525), (338,527), and (336,728). The exposure times for these PSF images may not be long enough to be totally free of the effect of the finite shutter travel speed for WFPC camera (Holtzman 1991).

Image PSFs An empirical PSF has been extracted from the R136 PC images at the position shown in Figs. 2 and 4).

Tiny TIM PSFs PSFs were computed using Version 2.4 of Tiny TIM. For the PC6 images of R136 the PSF position coincides with the position of the empirical PSF star used above. For the M87 image the PSF position coincides with the nucleus of the galaxy.

3. Image Restoration

3.1. R136 (PC6 F555W)

Two sets of restoration experiments have been done using the accelerated Richardson-Lucy method as implemented in STSDAS (200 iterations). In order to compare the quality of empirical vs. model PSFs, the R136 image was deconvolved using a) the PSF determined empirically from a star in the same image, b) a PSF from the WFPC PSF Library, and c) the Tiny TIM PSF. Then, in order to determine the quality of the Library PSFs and sensitivity to the space variance of the WFPC PSF, the R136 image was deconvolved using Library PSFs for three different positions in the field of view. Fig. 1 shows the normalized PSF images, with the three Library PSFs in the top row, the empirical PSF from the R136 image at bottom left, and the Tiny TIM PSF at bottom right.

Figs. 2 and 3 show the original R136 image and the various restorations. In Fig. 2 we show the original R136 image and restorations done using the empirical PSF from the same frame, the Library PSF, and the Tiny TIM PSF. Fig. 3 shows the three restorations done using Library PSFs for various field positions. The PSF star and the location of the Library PSFs are indicated by an arrow and a cross, respectively.

3.2. R136 (PC6 F547M)

Fig. 4 shows the original image of R136 in PC6, F547M. At top right is a restoration computed using the empirical PSF from the same frame. At bottom left is the restoration computed using the Library PSF, and at bottom right is the result obtained using the Tiny TIM PSF. The PSF star and the location of the Library PSFs are indicated by an arrow and a cross.

3.3. M87 (PC6 F547M)

The M87 image has no candidate star that can be used for extracting an empirical PSF. Thus, our experiments have been based on a Library PSF for PC6, F547M, centered on pixel (337,328), and the Tiny TIM PSF. The accelerated R-L method reaches a reduced χ-square of 0.4 after just 20 iterations. The results are shown in Fig. 5.

4. Discussion

Fig. 6 gives three sets of the azimuthally averaged profiles and the encircled energy plots. These are shown for (1) the PSFs (Image, Library, and Tiny TIM) for the image R136 F555W, and (2) the

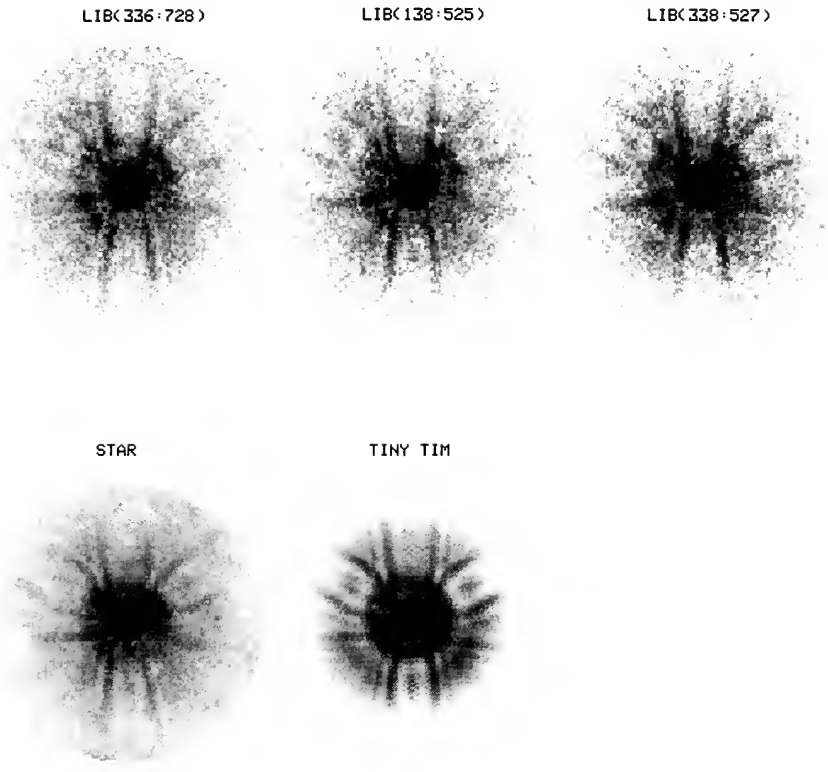


Figure 1. Normalized PSF images. Top row: Three Library PSFs with positions (336,728), (138,525), and (338,527). Bottom left: The empirical PSF from the R136 image. Bottom right: Tiny TIM PSF.

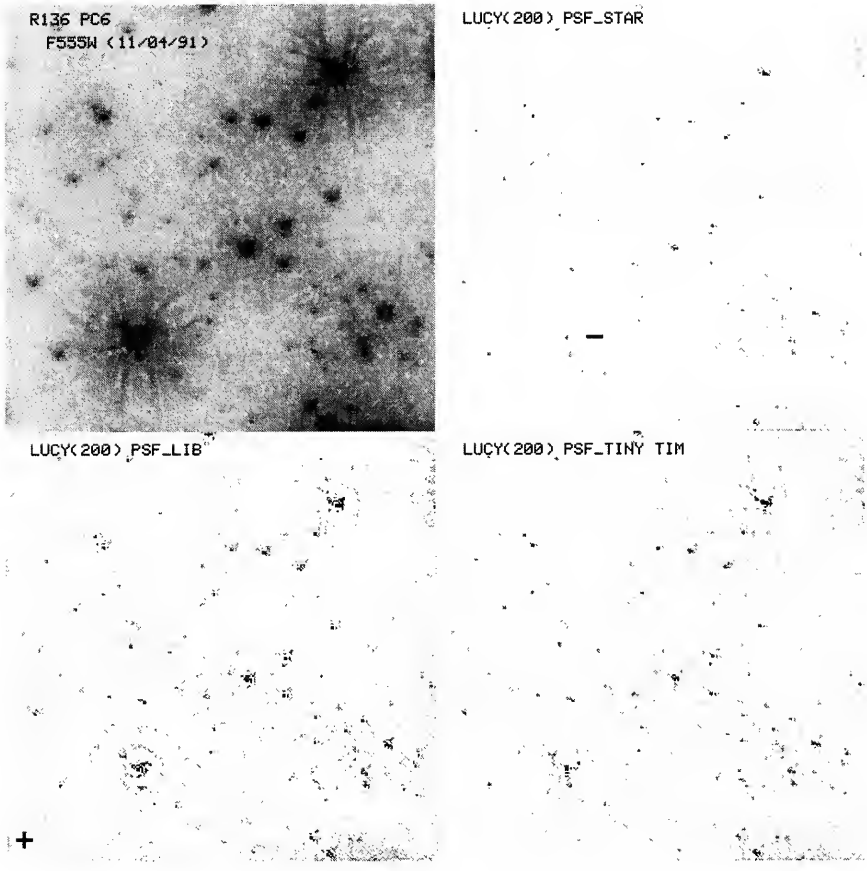


Figure 2. The original R136 PC6 F555W image taken on 11 April 1991 and the various restorations. Top left: The original R136 image. Top right: The Lucy restoration done using the empirical PSF from the same frame. Bottom left: Lucy restoration using the Library PSF. Bottom right: Lucy restoration using the Tiny TIM PSF.

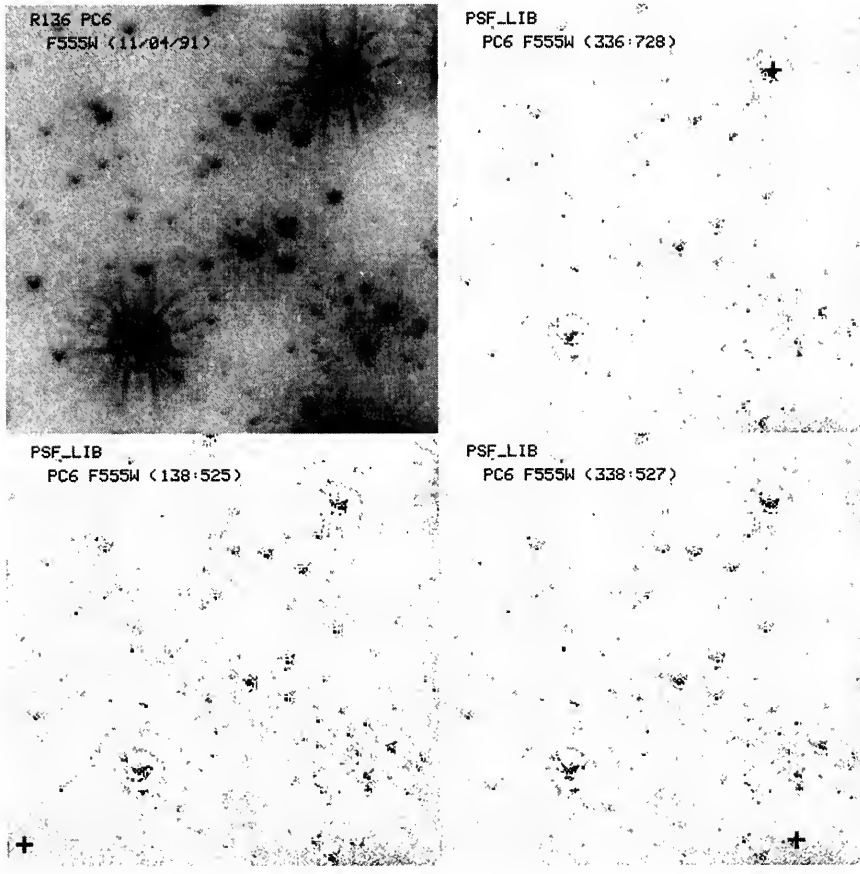


Figure 3. The original R136 PC6 F555W image taken on 11 April 1991 and the three restorations done using Library PSFs for various field positions. Top left: The original R136 image. Top right: Lucy restoration done using the library PSF at position (336,728). Bottom left: Lucy restoration using the library PSF at position (138,525). Bottom right: Lucy restoration using the library PSF at position (338,527).

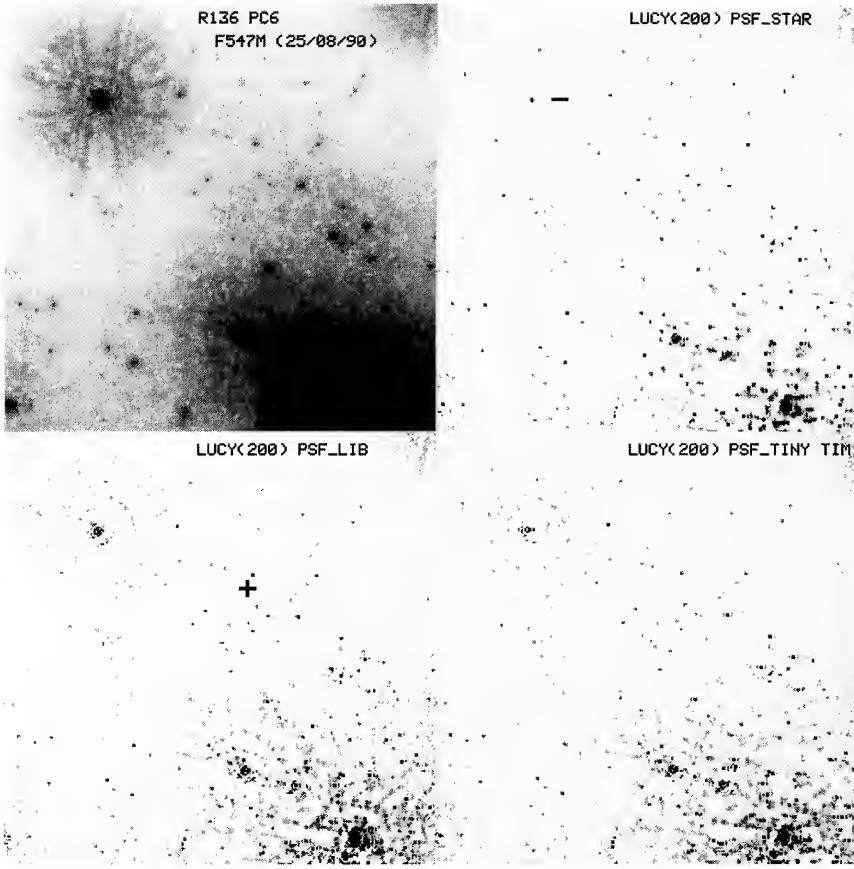


Figure 4. The original R136 PC6 F547M image taken on 25 August 1990 and the various restorations. Top left: the original R136 image. Top right: The Lucy restoration done using the empirical PSF from the same frame. Bottom left: Lucy restoration using the Library PSF. Bottom right: Lucy restoration using the Tiny TIM PSF.

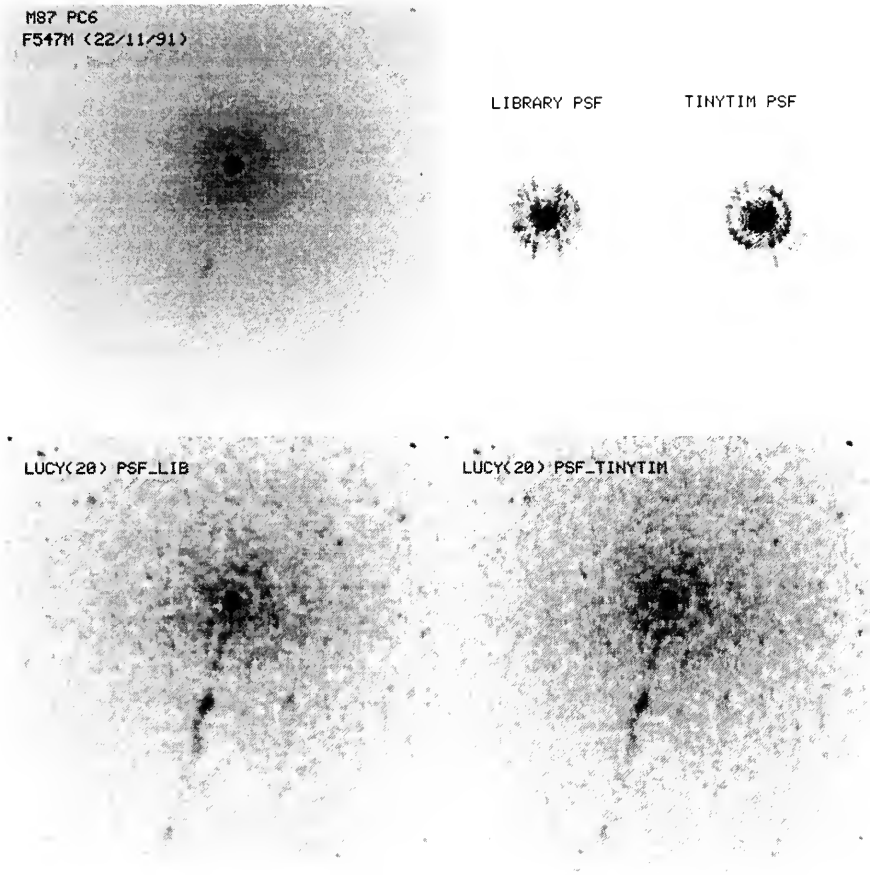


Figure 5. The original M87 PC6 F547M image, two PSFs and the Lucy restorations. Top left: The original M87 image. Top right: The Library PSF for PC6, F547M, centered on pixel (337,328) and the Tiny TIM PSF. Bottom left: The Lucy restoration done by the library PSF. Bottom right: The Lucy restoration done using the Tiny TIM PSF.

vicinity of the restored PSF star, centering on the PSF star with a radius 2.″5, for the image R136 F555W, and (3) the PSF star area in the restored image of R136 F547W.

The FWHM of the PSFs generated from different methods are all nearly the same (0.″1). The radius of the first ring of the Image PSF is very close to that in the Tiny TIM model (~0.″6) and smaller than the radius measured for the Library PSF (~0.″75). The second ring can clearly be seen in the Tiny TIM PSF but not in the others because of the noise and background (see Fig. 6, top left panel).

On the basis of the effective resolution and enhancement in dynamic range, the empirical PSF extracted from the data frame gives the best result. The peak value of the brightest star in the image restored using a PSF extracted from the same image is about four times higher than that obtained using the Library or Tiny TIM PSFs.

More residual PSF structures appear in the image restored using the Library and Tiny TIM PSFs. Bright stars sometimes split into a bright core with several faint neighbors. This may lead to the generation of a number of erroneous detections of faint stars in the restored images.

The encircled energy plot and the profile of the jet of M87 are illustrated in Fig. 7. The left panel shows the encircled energy plots in the vicinity of the galaxy nucleus in the restored images, and the profiles of the jet in the restorations are given in the right frame. The solid line in Fig. 7 shows the Library PSF for PC6, F547M, centered on pixel (337,328), and the dashed line shows the Tiny TIM PSF. The differences between the two encircled energy curves are minor. The structure of the faint jet is consistent in both M87 restorations.

There are several rings surrounding the nucleus of the galaxy in each restored image of M87. This is almost certainly an artifact of the mismatch between the true PSF and the models used, since the structure of the unrestored image is clearly dominated by the familiar features of the PSF (Fig. 8).

5. Conclusions

The best results in image restoration rely upon being able to determine a high S/N PSF from the same frame. Current PSF modeling code provides acceptable PSFs when high quality empirical PSFs are not available. PSFs from the WFPC PSF Library are about of equal value as model PSFs. Their use is limited owing to non-optimum exposures and time variability of focus. Improved modeling of the HST optical system (e.g., Redding 1994) should allow for better results in image restoration using model PSFs. In order to achieve accurate restorations of data with high dynamic range excellent PSF models are required. PSF models are less critical when the intrinsic source structure is smooth compared to the size of the PSF.

Acknowledgments. This work has been supported by ST ScI's Image Restoration Project.

References

Baggett, S. M., & MacKenty, J. W. 1992, "WFPC PSF Image Library," Space Telescope Science Institute internal report

Holtzman, J. 1991, in *Wide Field/Planetary Camera Final Orbital/Science Verification Report*, S. M. Faber, ed., Space Telescope Science Institute, Baltimore, Chapter 9, 5

Krist, J., & Hasan, H. 1993, in *Astronomical Data Analysis Software and Systems II*, ASP Conference Series 52, R. J. Hanisch, R. J. V. Brissenden, & J. Barnes, eds., 530

Krist, J. 1993, in *Astronomical Data Analysis Software and Systems II*, ASP Conference Series 52, R. J. Hanisch, R. J. V. Brissenden, & J. Barnes, eds., 536

Redding, D.C. 1994, this volume

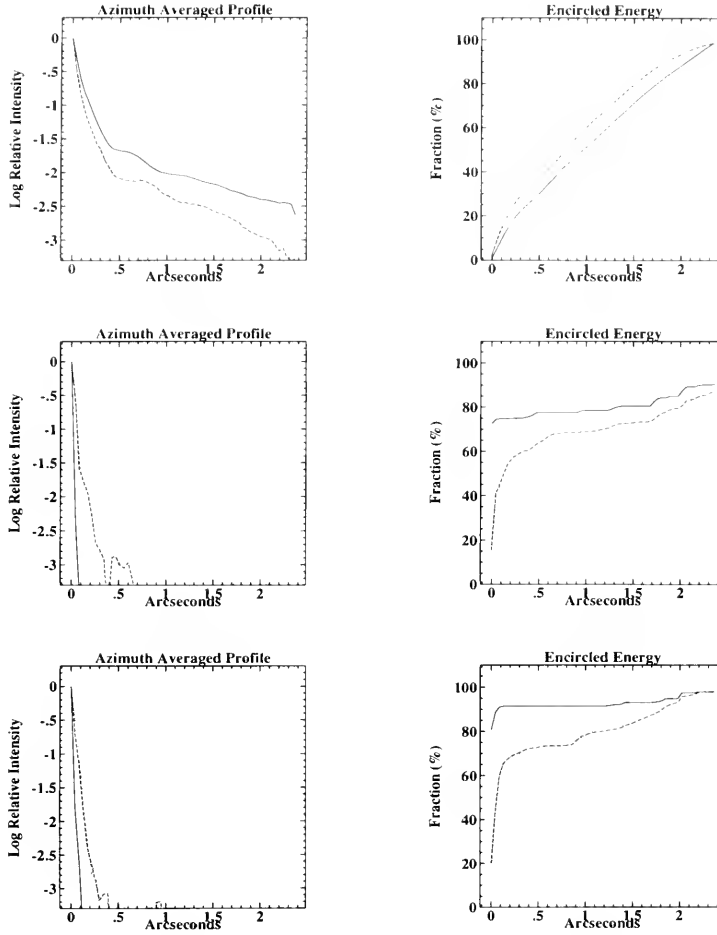


Figure 6. Three sets of the azimuthally averaged profiles and the encircled energy plots. Top: The PSFs (solid line: Image, dashed line: Library, and dotted line: Tiny TIM) for the image R136 F555W. Middle: The vicinity of the restored PSF star, centered on the PSF star with a radius 2.″5, for the image R136 F555W. Bottom: The PSF star area in the restored image of R136 F547W.

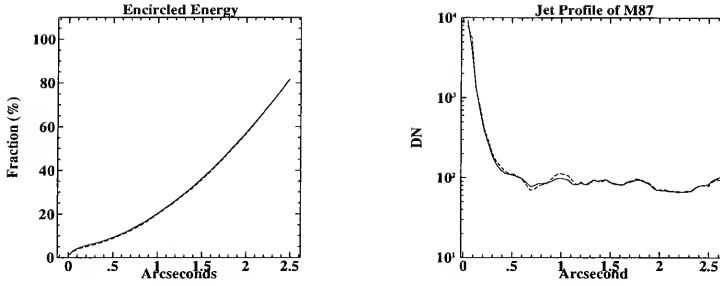


Figure 7. The encircled energy plot and the profile of the jet of M87. Left: The encircled energy plots of the vicinity of the galaxy nucleus in the restored images. Right: The profiles of the jet. The solid line indicates Library PSF for PC6 F547M, centered on pixel (337,328). The dashed line indicates Tiny TIM PSF.

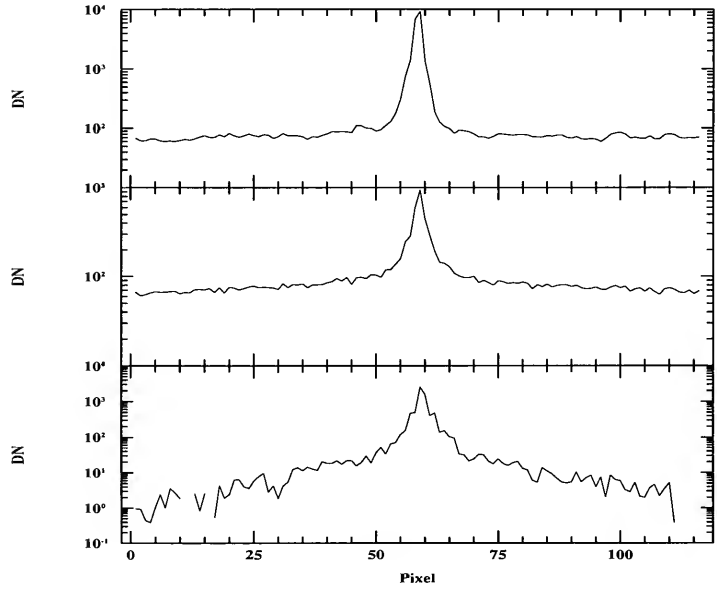


Figure 8. The profile of the Library PSF and the vicinity of the galaxy nucleus of M87. Bottom: PSF. Middle: blurred image of M87. Top: restored image of M87 using the Library PSF.

The Restoration of HST Images and Spectra II
Space Telescope Science Institute, 1994
R. J. Hanisch and R. L. White, eds.

FOC Image Restoration Using Calculated PSFs on Parallel Architectures

Raadhakrishnan Poovendran

Hughes STX/NASA–Goddard Space Flight Center, Greenbelt, MD 20771

John E. Dorband and Jan M. Hollis

NASA–Goddard Space Flight Center, Greenbelt, MD 20771

Abstract. We describe our image restoration efforts which use an analytical model of the point spread function (PSF) of the Hubble Space Telescope Faint Object Camera (FOC). Our approach is based on the Zernike polynomial modeling of the wave front phase using two phase retrieval algorithms. The difficulties involved in modeling the PSF and validating our approach are illustrated using actual FOC and simulated data which have been processed on a massively parallel computer.

1. Introduction

For a particular wavelength, given two FOC scenes of the same extended astronomical object taken within a few hours of one another and an observed FOC point spread function taken at nearly the same epoch, deconvolution of the extended images using the observed PSF should lead to similar small-scale structure. We have found that this is not the case. The probable reason for such differences is that the FOC system is time variant, coupled with the complication of noise on both the images and the observed PSF. It is well-known that observed PSFs taken in a sequence seem to "breathe" as if the focus of the camera system is changing with time. Parametric Zernike polynomial modeling of the point spread function core and wings, coupled with phase retrieval techniques constrained by the observed PSF, can produce a calculated noise-free (analytic) PSF. It may be that this analytic PSF can be used in the deconvolution of both observed extended scenes to obtain similar small scale structure by empirically changing only a few Zernike coefficients (e.g., the focus term).

2. Approach for Calculating an Analytical PSF

Since image restoration algorithms are highly sensitive to the accuracy of the calculated PSF, a significant portion of our research effort is devoted to accurately modeling the PSF which is a highly non-linear problem. Our approach is based on analytical modeling of the wavefront phase using Zernike polynomials. The wave front phase is reconstructed using a phase retrieval algorithm. First, we must make reasonable guesses of significant Zernike polynomial coefficients, and use these as input to a phase retrieval algorithm (i.e., Gerchberg and Saxton 1972, Fienup 1982) to produce an estimate of the FOC phase map, which is an array of complex numbers. In the Fourier domain, we scale the phase map by an observed PSF to more accurately account for the HST/FOC combination. After scaling, the inverse transform provides the required information in the pupil domain. We mask and scale the inverse transform data by the well-known geometry of the components contributing to the pupil (e.g., spider supports, mirror handling pads, central obscuration, as well as primary and secondary mirror surface errors). We then repeat this process several times to converge with minimal differences between the result and the observed PSF. The final result is an improved estimate of the phase map. The resulting map has to be smoothed and unwrapped to obtain a continuous phase

map. We then use this result as input to a total least squares estimate algorithm (QR factorization) to provide improved Zernike coefficients which ultimately permit a final calculated PSF whose point-to-point differences are minimized with respect to the observed PSF. Of course, at this point, the observed and calculated PSFs are compared by a variety of other means including inspection of the core and wings to insure that morphologically they are as close as possible — if not, some important Zernike term may have been left out. Moreover, one has to contend with achieving a local rather than a global minimum in this process, a problem that has yet to be solved for the phase retrieval stage. We present the mathematical methodology for modeling the PSF in the next section.

3. Mathematical Methodology

The observed PSF is proportional to the power spectrum of the plane wave incident at the entrance pupil of the telescope (Goodman 1968). This can be written as:

$$\mathrm{PSF}(\omega_1, \omega_2) = k \left| \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} e^{\phi(x,y)} e^{-i(\omega_1 x + \omega_2 y)} dx dy \right|^2 \tag{1}$$

where $\phi(x, y)$ is the phase of the plane wave incident on the entrance pupil of the telescope and k is a proportionality constant. Eq. (1) assumes that the effect of the telescope assembly can be expressed by a single equation and, hence, is an approximation in itself. However, Eq. (1) expresses the PSF as a function of the wave front phase at the entrance pupil. Hence, the observed PSF can be useful in reconstructing the wavefront at the entrance pupil of the telescope.

Zernike polynomials and their variants are known for better modeling of different aberrations of the phase function (Zernike 1935). Analytical representation of the phase in terms of the Zernike polynomials is given by:

$$\hat{\phi}(x, y) = \sum_{i=1}^{i=N} z_i P_i(x, y) \tag{2}$$

where z_i is the Zernike coefficient corresponding to the Zernike polynomial $P_i(x, y)$ and N is the number of Zernike polynomials that are necessary to represent the analytical phase. Hence, the problem of analytic PSF generation reduces to estimating the Zernike polynomial coefficients that are necessary to represent the entrance pupil wave front.

For phase retrieval we used the iterative transform algorithm (ITA) of Fienup (1982), who based much of this ITA on the Gerchberg and Saxton (1972) algorithm, and a variation of the ITA. These two algorithms are shown in Fig. 1. For Algorithm 1, which corresponds to annular Zernike polynomials, we used the telescope geometrical information such as the mirror sizes, mirror pads, spiders and the obscuration geometry in the pupil plane and the observed PSF in the Fourier transform domain. Algorithm 1 stagnated after a few thousand iterations. We used a robust QR algorithm (Golub and Van Loan 1992) to analytically represent the phase using the annular Zernike polynomials. When compared with the observed PSF, we noted that the simulated PSF did not have detailed structures such as the clover and pronounced wings that are so essential for the image restoration.

In an attempt to improve the phase retrieval results, we used Algorithm 2, which corresponds to the classical radial Zernike polynomials. Unlike Algorithm 1, Algorithm 2 does not impose any constraint in the pupil domain. This is equivalent to assuming that the wave front is continuous at the entrance pupil. We used a circular mask in the pupil domain region of interest and retrieved the phase. This algorithm also stagnated after a few thousand iterations. Since the classical radial Zernike polynomials are orthogonal over the circular pupil domain, the corresponding Zernike coefficients are obtained using a simple inner product. PSFs generated with Algorithm 2 led to better representation of the inner core structures than Algorithm 1, but still did not compare favorably to the observed PSF for better image restoration.

Fig. 2 presents the schematic diagram of the process involved in the PSF simulation. All the functions shown in the diagram are implemented on a massively parallel computer — the MasPar

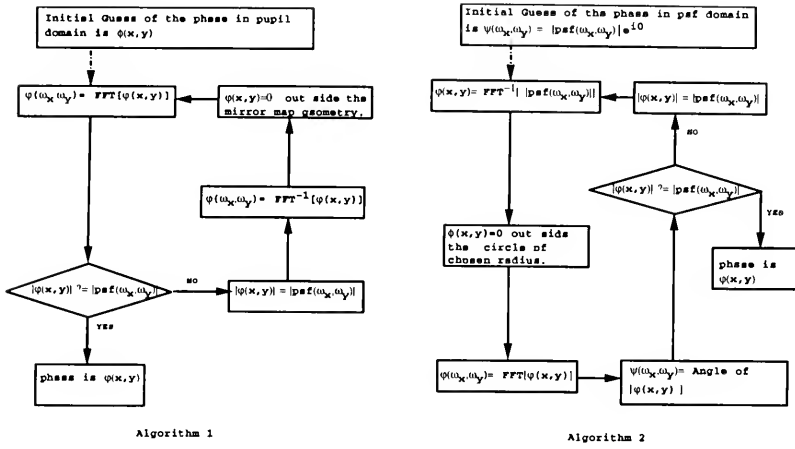


Figure 1. Schematic diagram of the phase retrieval algorithms.

with 16,384 processors. Zernike polynomials are generated "on-line" during the computation and are not stored.

4. Results and Discussions

In this section we describe different restoration results with FOC and simulated data. Before describing the restoration details we discuss the restoration method.

4.1. Maximum Entropy Image Restoration

We used the maximum entropy method (MEM) of Hollis, Dorband, Yusef-Zadeh (1992) who restored images that were plagued by the FOC detector saturation which corrupted the data in certain regions. Use of a mask was required to ensure that the saturated data were ignored in the resultant restoration which these authors obtained with an observed PSF. We have determined that the size and shape of the mask affects the detailed structures of the MEM restoration for a given number of iterations, but not the global structure. Moreover, we found the detailed structures are independent of the mask geometry when allowed to iterate long enough.

4.2. Deconvolution: Observed Versus Analytic PSFs

Hollis and colleagues obtained two public domain FOC F/96 images of the binary system R Aquarii and an observed PSF both in the light of twice ionized oxygen which emits a spectral line at 5007 Å. These images were plagued by spherical aberration and, in the case of R Aquarii, also by detector saturation. The MEM technique described above was used to restore one R Aquarii image with the observed PSF to the original design resolution of the HST/FOC combination. It was precisely because the observed FOC PSF was larger than the saturated region of the R Aquarii image that the restoration was successful (see Hollis, Dorband, and Yusef-Zadeh 1992). However, as noted in the Introduction, restoration of two extended FOC images taken at two different times does not lead to similar small-scale details when using the observed PSF. We used the two FOC images and the observed PSF used by Hollis, Dorband, and Yusef-Zadeh (1992) and repeated the MEM deconvolution. Restored images after 1000 iterations are shown in Fig. 3. These images are not

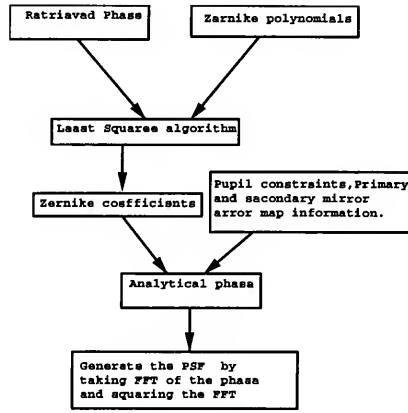


Figure 2. Schematic diagram of the PSF generation procedure.

identical in details, suggesting that that the observed PSF does not correspond well enough to the images for deconvolution purposes. Hence, it is desirable to be able to synthesize two different analytical PSFs to remedy the differences between the two extended images of interest.

In our effort to compute and use an analytical PSF in the restoration process, we noted that the structures prominent in the center core of the observed PSF are better modeled by the 31st and the 32nd Zernike "clover" coefficients which correspond to the $\cos(3\theta)$ and $\sin(3\theta)$ polynomial terms. Restoration results of the FOC R Aquarii images using an analytical PSF, computed using our estimated annular Zernike polynomial coefficients ($z_4 = -1.65, z_{11} = -0.52, z_{31} = -0.0214$, and $z_{32} = -0.0214$), are shown in Fig. 4. It can be noted that this single analytical PSF leads to fairly consistent enhanced details compared to the restoration using the observed PSF. This may be due to the noise-free analytical PSF. However, some of the faint structures shown in the restoration using the observed and the analytical PSf are computational artifacts. This suggests that neither the observed PSF nor the analytical PSF matches the scene to be restored and, moreover, two different analytical PSFs will be eventually required to restore two images separated in time. The following simulation study demonstrates that these artifacts are a measure of the accuracy with which the PSFs are modeled.

4.3. Effect of PSF Mismatch on the Deconvolution

A test image was generated by convolving five spatially distributed delta functions with the "correct" PSF which was calculated by means of Zernike coefficients. Using the MEM algorithm, the test image was subsequently deconvolved with both the correct PSF and a "mismatched" PSF which was calculated using the same Zernike coefficients as in the correct PSF except that z_4, the focus term, was changed by 0.1 waves rms. When logarithmically stretched, both image restorations gave a subtle mottled pattern with extremely weak linear artifacts that seem to bridge the gaps between delta functions, but the mismatched PSF artifacts were more exaggerated. Increasing the number of iterations for each type restoration reduces the prominence of the artifacts in the correct PSF case. Moreover, the mismatched PSF restores more slowly than the PSF which was used to generate the image in the first place. Hence, for a given number of iterations, the correct PSF will restore points to larger numerical values with fewer artifacts than the mismatched PSF. This property of "maximized signal and minimized artifacts" could be used in an empirical

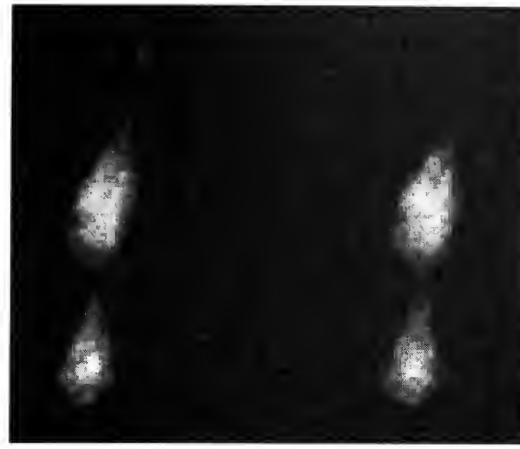


Figure 3. Restoration of FOC R Aqr data using an observed PSF. Left: R Aqr image taken 8/23/90 at 00:26 UT and right: R Aqr image taken 8/23/90 at 02:01 UT. The observed PSF image used in these restorations taken on 8/28/90.

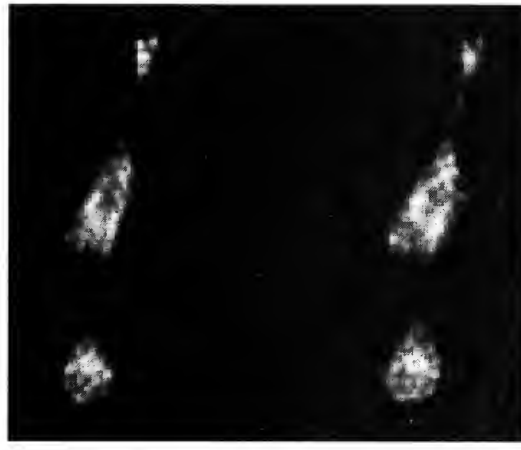


Figure 4. Restoration of FOC R Aqr data using an analytic PSF with Zernike coefficients that include clover terms (see §4.2). Left: R Aqr image taken 8/23/90 at 00:26 UT and right: R Aqr image taken 8/23/90 at 02:01 UT.

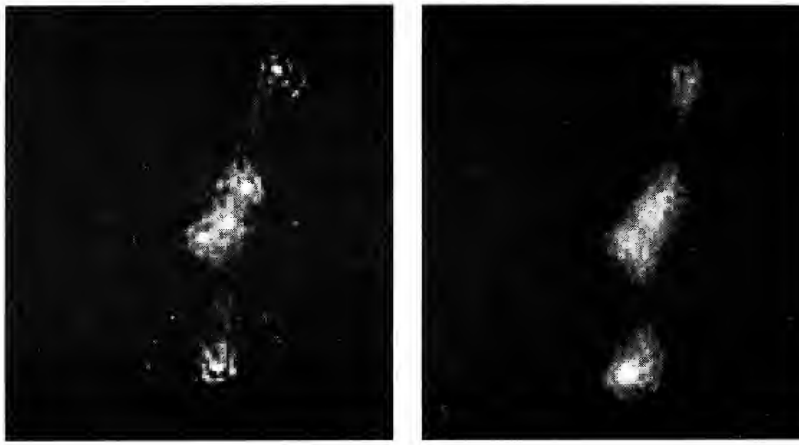


Figure 5. The right image is the result of restoring FOC R Aqr data using an analytical "ST ScI PSF" (see §4.3). The left image is the result of restoring a test image using an analytical PSF which is different from the PSF that created the test image in the first place. Comparison of the left and right images shown here suggest that the subtle mottled structure and other artifacts are caused by a mismatch of the PSF used to deconvolve each scene.

fashion to decide between the results given by two different PSF restorations for a fixed number of iterations when using this algorithm. Unfortunately, this method would be very time consuming, and thus, a convergence criterion based on some statistical property of the converging image, or some property grounded in the physics, or some smoothness property is sorely needed. Fig. 5 shows a 1000 iteration restoration of the test image with a mismatched PSF along side of a 1000 iteration restoration of FOC data of the R Aquarii Jet using a calculated PSF whose Zernike coefficients ($z_4 = -1.267, z_5 = 0.077, z_6 = -0.021, z_7 = -0.025, z_8 = 0.0881, z_{11} = -0.533$) were published in the STScI Newsletter of March 1992 (hereafter the "ST ScI PSF"). Comparison of these highly similar results suggest that both the test image and the actual HST data restorations are plagued by mismatched PSFs.

Our objective is to be able to calculate the correct PSF for use in subsequent deconvolution of an extended image. Thus, it is of interest to compare the results of a 1000 iteration restorations of FOC data of the R Aquarii Jet using the ST ScI PSF and an observed PSF, and these two results are shown in Fig. 6 as a three-dimensional plot. The results are very similar, and from a science information point of view, are indistinguishable. Considering the fact that the observed PSF data were taken 5 days later than the R Aquarii Jet data by the FOC and HST secondary mirror movement is suspected, it is not unexpected that this observed PSF would be mismatched to R Aquarii Jet data for restoration purposes. However, in a relative sense, note that the signal-to-noise on the image produced with the observed PSF is less than the image produced by the ST ScI PSF. Here the noisy areas lay outside of the areas dominated by regions of intense signal. This suggests that the observed PSF is better matched to the data than the calculated ST ScI PSF.

In conclusion, we still need to restore two scenes of the same object separated in time to validate the contention that we should be able to do better with a calculated PSF than an observed PSF. Once we are able to do this, then, in principle, we can sample a calculated PSF at any desired resolution and generate imagery that exceeds the design resolution of the HST/FOC combination (see Dorband and Hollis 1992). Our quest in this effort is still on-going.

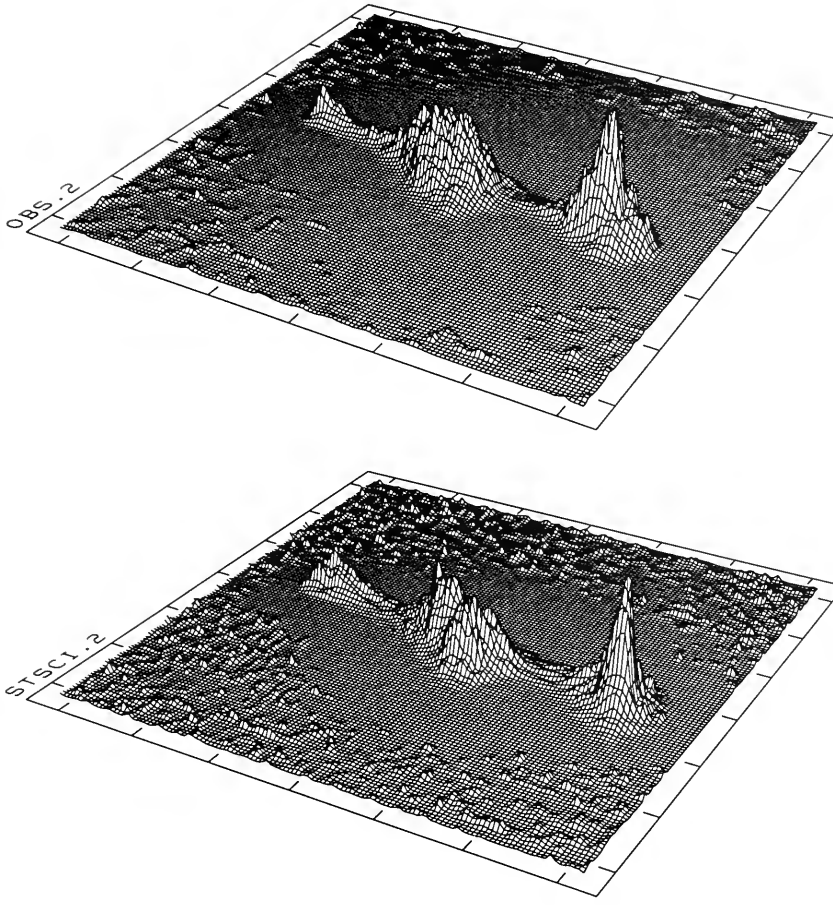


Figure 6. Upper: A three dimensional perspective of a restoration using FOC R Aqr data and an observed PSF after 1000 iterations. Lower: Same as above, but restoration performed using the analytic "ST ScI PSF". Note that the relative signal-to-noise on the upper figure is higher than on the lower figure, suggesting that the noisy observed PSF has provided a better restoration than the noise-free analytic PSF (see §4.3).

Acknowledgments. We gratefully acknowledge the help of Mr. Rick Lyon of RADEX and Mr. Anthony Gruszak of HDOS.

References

Dorband, J. E, & Hollis, J. M. 1992, in *Astronomical Data Analysis Software and Systems I*, D. M. Worrall, C. Biemsderfer, & J. Barnes, eds., ASP Conference Series, 25, 200

Hollis, J. M, Dorband, J. E, & Yusef-Zadeh, F. 1992, ApJ, 386, 293

Goodman, J. W 1968, *Introduction to Fourier Optics*, McGraw-Hill, San Francisco

Gerchberg, R. W, & Saxton, W. O. 1972, Optik, 35, 237

Fienup, J. R. 1982, Appl.Optics, 21, 2758

Zernike, F. 1943, MNRAS, 94, 377

Burrows, C. 1992, in ST ScI Newsletter, Space Telescope Science Institute, Baltimore, 9, 10

Golub, G. H, & Van Loan C. F. 1992, *Matrix Computations*, The John Hopkins University Press, Baltimore

The Restoration of HST Images and Spectra II
Space Telescope Science Institute, 1994
R. J. Hanisch and R. L. White, eds.

Approximate Shift-Invariance by Warping Shift-Variant Systems

Scott R. McNown

Optical Sciences Center, University of Arizona, Tucson AZ, 85719

Bobby R. Hunt

Electrical and Computer Engineering, University of Arizona, Tucson AZ, 85719

Abstract. A method is presented in which a signal, degraded by a linear shift-variant system, will undergo a warping such that the resulting warped signal will be approximately described by a warped original signal filtered by a linear shift-invariant system. The warping is a limited class of coordinate transformations, for which adjacent points do not cross each other after the transformation. This results in a signal that may appear stretched in some places and compressed in others (and curved if the signal is two-dimensional). The purpose of this distortion is to make the space-variant impulse response (which can be viewed as a space-invariant impulse response which has been warped in the original signal domain) vary as little as possible. In particular cases, a transformation can be found which will result in no impulse response variations. For most cases, however, the impulse response will still have some space variance, which the warping seeks to minimize. The residual variance will be ignored (this error must be small in order for this method to work well), and an "average" impulse response in the warped domain will be assumed. This allows for space-invariant restoration of the warped signal, with all of its attendant advantages in speed and reduced complexity.

1. Introduction

Image formation is typically described as a linear process. Such systems are described by the Fredholm equation of the first kind,

$$g(\boldsymbol{x}) = \int d\boldsymbol{u}\, h(\boldsymbol{x}, \boldsymbol{u})\, f(\boldsymbol{u}), \tag{1}$$

where, for 2-dimensional systems, the object coordinates are $\boldsymbol{u} = (u_1, u_2)$ and the image coordinates are $\boldsymbol{x} = (x_1, x_2)$.

Such a general description of a linear system is termed shift-variant, because the shape of h, the impulse response (or point spread function, PSF), might change as the object position, $\boldsymbol{u}$, is shifted. The Hubble Space Telescope (HST) is an example of a shift-variant imaging system. The primary contribution to image degradation in the HST is spherical aberration. This aberration varies only with the exit pupil geometry, and not with field position — it is intrinsically space-invariant. The exit pupil geometry itself, however, changes with field position. This is due to the different longitudinal positions (for the unfolded system) of the stop (primary mirror), the central obscuration, and the primary and central obscuration of the Cassegrain relays in the WF/PC, all of which undergo effective lateral shifts for different field positions (causing vignetting). Hence, the exit pupil geometry changes with field position, so the system response changes with field position (i.e., is shift-variant). The field extent (thus the vignetting in the exit pupil) for the WF/PC is large enough that the shift-variance of the imaging system is substantial.

If the *shape* of h does not change as the object position is changed, then h is said to be shift-invariant (alias space-invariant, isoplanatic, or stationary), and the problem is substantially

simplified. Of course, the *position* of h in the image plane changes as the object position changes. The formulation of image formation after this simplification is

$$g(\boldsymbol{x}) = \int d\boldsymbol{u}\, h(\boldsymbol{x} - \boldsymbol{u})\, f(\boldsymbol{u}). \tag{2}$$

Shift-invariant systems are much more easily inverted (given g, find f) than shift-variant systems. All shift-invariant systems have complex exponentials as eigenfunctions (Gaskill 1978), so Fourier transforms reduce the above convolution into a simple scalar multiplication. Inversion, within the pass-band of the system, then consists simply of a scalar division (in the complex exponential basis), possibly combined with a regularization term to attack the problem of noise amplification. Since very efficient algorithms are available to take these transforms for discrete systems (FFTs), the solution time of f is quick in comparison to space-variant inversion. Shift-invariance also has the advantage of being well and widely understood; there are many restoration and processing tools in place which depend upon this simplification (Wiener filtering (Jain 1989), statistical estimation (Jain 1989), and super-resolution (Gerchberg 1989 and Papoulis 1975), to name a few).

Space-variant inversion, excluding null functions, can be achieved from a singular value decomposition (SVD) of h (Andrews and Hunt 1977). This method involves finding *two* sets of eigenvectors, and transforming g into one of the eigenvector bases and f into the other. Although theoretically this is the same process as discussed above, in practice it is intractable for large systems due to the need to explicitly solve for two (typically large) sets of eigenvalues (whereas one set, complex exponentials, were needed for shift-invariant systems, and were known a priori).

The Landweber algorithm (Landweber 1951) provides an iterative solution to this space-variant problem without having to resort to finding the eigenvectors and eigenvalues needed for SVD. However, even if the impulse response does not change, the algorithm must be fully repeated for each image. This algorithm is an inefficient way to restore a large number of images generated by the same imaging system.

A formulation of space-variant imaging is desired for which the image and object can both be expressed in a common basis (complex exponentials are preferred, since efficient mechanisms for their manipulation are already in place). This amounts to reformulating the space-variant problem into a space-invariant problem. Sawchuck (1972, 1973, 1974), Robbins (1970), and Robbins and Huang (1972) found that for certain problems (coma-like aberrations and motion blur), the image plane could be remapped so that the space-variant problem could be exactly transformed into a space-invariant form. But only impulse responses which could be decomposed into

$$\begin{aligned} h(\boldsymbol{x}, \boldsymbol{u}) &= \alpha(\boldsymbol{x})\, \beta(\boldsymbol{u})\, h_I(\boldsymbol{c}(\boldsymbol{x}) - \boldsymbol{b}(\boldsymbol{u})) \\ &\leftrightarrow \quad h(\boldsymbol{y}, \boldsymbol{v}) = \alpha(\boldsymbol{c}^{-1}(\boldsymbol{y}))\, \beta(\boldsymbol{b}^{-1}(\boldsymbol{v}))\, h_I(\boldsymbol{y} - \boldsymbol{v}), \end{aligned} \tag{3}$$

where the coordinate transformations are

$$\begin{aligned} \boldsymbol{c}(\boldsymbol{x}) &= (c_1(x_1, x_2),\ c_2(x_1, x_2)) = (y_1, y_2) = \boldsymbol{y} \\ \boldsymbol{b}(\boldsymbol{u}) &= (b_1(u_1, u_2),\ b_2(u_1, u_2)) = (v_1, v_2) = \boldsymbol{u} \\ (\boldsymbol{c}(\boldsymbol{x}) - \boldsymbol{b}(\boldsymbol{u})) &= (y_1 - v_1,\ y_2 - v_2) = \boldsymbol{y} - \boldsymbol{v}, \end{aligned}$$

could be transformed. No general method for finding this decomposition was given; this is the goal of this paper. The advantage of obtaining this formulation of imaging is clear: the hard work is done up front — once an impulse response is given and the coordinate transforms are found, the restoration can be done using the space-invariant tools already mentioned. In general, it is impossible to decompose every possible shift-variant impulse response into equivalent shift-invariant responses. If this exact representation is unattainable, then an approximate shift-invariant response is sought.

From Eq. 3, note that the object and image planes can be warped by different transformations ($\boldsymbol{b}$ and $\boldsymbol{c}$, respectively). Different transformations are desired only if there is a warping caused

by the imaging system (e.g., the Seidel distortion aberration, non-perpendicular imaging system geometries, magnification, etc.). Under many imaging processes (e.g., blurring, for which the centroid of the impulse response is at the intended image location), the sample positions in the restoration plane must be the same as the sample positions in the image plane. For most restoration procedures, the sample positions in the image plane *define* the sample positions in the object plane. A geometric warping caused by the imaging system may be corrected separately. For non-warped imagery, which is assumed in all that follows, the object and image planes will undergo the same transformation: $(\boldsymbol{x}, \boldsymbol{u}) \rightarrow (\boldsymbol{b}(\boldsymbol{x}), \boldsymbol{b}(\boldsymbol{u}))$. The dimensionality of the problem is then reduced by one-fourth. $\alpha(\boldsymbol{x})$, $\beta(\boldsymbol{u})$, and $\boldsymbol{b}(.)$ still need to be found.

2. Sample Shifting

Moving the sample locations to achieve the desired warping will result in an irregularly sampled image plane. These irregularly spaced samples will then be brought back to regularly spaced positions, hence warping the image. Typical restoration techniques can then be performed using these regularly spaced samples. A method of optimally finding these sample positions is given below. Only the one-dimensional case will be considered here; the two-dimensional case will be attempted in the future.

2.1. Error Metric

The linear mapping of Eq. 1 is put into discrete form, with uniformly spaced samples of $g(\boldsymbol{x})$ and $f(\boldsymbol{u})$. In one dimension, Eq. 1 then becomes

$$g(x_i) = \sum_j h(x_i, u_j) f(u_j), \tag{4}$$

where uniform sampling, $x_i = x_0 + i\Delta_x$ and $u_j = u_0 + j\Delta_u$, is often assumed. The sampled impulse response, $h(x_i, x_j)$, is a function of the n sample positions, $\{x_i\}_{i=0}^{n-1}$. In particular, $h(x_{i+k}, x_i)$ describes the amount the i^{th} object sample affects the image k samples away from the i^{th} image sample.

The sample locations are to be moved so that the resulting impulse response is approximately space-invariant. To this end, the invariant term of Eq. 3 can be found, in sampled form for a one dimensional mapping, to be

$$h_I(x_i, x_j) = \frac{h(x_i, x_j)}{\alpha(x_i)\beta(x_j)}.$$

Since α and β are not predetermined, their explicit dependence on x is unnecessary. Let $\alpha'_i = 1/\alpha(x_i)$ and $\beta'_j = 1/\beta(x_j)$. A space-invariant impulse response should not depend on the coordinate of the object, i.e., a space-invariant impulse response should obey $h_I(x_{i+k}, x_i) = h_I(x_{j+k}, x_j)$, $\forall i, j, k$. If this equality does not hold, then the difference between the different impulse responses describes an error from the assumption of shift-invariance.

A metric which measures the total error over the field is

$$E = \sum_{i=N_l}^{N_u-1} \sum_{j=i+1}^{N_u} \sum_{k=N'_l-i}^{N'_u-j} \left[\alpha'_{i+k}\beta'_i h(x_{i+k}, x_i) - \alpha'_{j+k}\beta'_j h(x_{j+k}, x_j)\right]^2 \tag{5}$$

$$= \frac{1}{2} \sum_{i=N_l}^{N_u} \sum_{j=N_l}^{N_u} \sum_{k=N'_l-\min(i,j)}^{N'_u-\max(i,j)} \left[\alpha'_{i+k}\beta'_i h(x_{i+k}, x_i) - \alpha'_{j+k}\beta'_j h(x_{j+k}, x_j)\right]^2 \tag{6}$$

where the limits of the image are $N_l \le i, j \le N_u$. The limits on the summations keep the sample points within the image boundaries. Allowing the k sum to go outside the image boundaries

($N_l' < N_l$ or $N_u' > N_u$) includes some "anchor" samples outside the image, which tends to stabilize the solution trajectory when E is minimized. For example, if all of the sample positions may be moved, there is a global minimum ($E = 0$) when all of the samples are co-located. If all of the sample positions which start off inside the image are forced to remain in the image, then keeping a fixed sample position outside the image removes this zero point. Note also that k describes the distance from the "center" of the space-variant impulse response. The limits on k could therefore be constrained PSFs with support more compact than that of the entire image (which is usually the case). However, it is the actual distance $x_{i+k} - x_i$, not just the magnitude of k, which needs to be within the support of the PSF. Since the solution will necessarily change the sample locations, the limits on k will depend upon the resulting sample locations.

The goal here is to minimize E. This will provide a sampling scheme which will yield the most space-invariant impulse response.

2.2. Gradients

For minimization, the gradient elements dE/dx_n are needed.

$$\frac{dE}{dx_n} = \sum_{i=N_l}^{N_u} \sum_{j=N_l}^{N_u} \sum_{k=N_l'-\min(i,j)}^{N_u'-\max(i,j)} \left[\alpha'_{i+k}\beta'_i h(x_{i+k}, x_i) - \alpha'_{j+k}\beta'_j h(x_{j+k}, x_j)\right] \times \left[\alpha'_{i+k}\beta'_i \frac{\partial}{\partial x_n} h(x_{i+k}, x_i) - \alpha'_{j+k}\beta'_j \frac{\partial}{\partial x_n} h(x_{j+k}, x_j)\right] .$$

An application of the chain rule gives

$$\begin{aligned} \frac{\partial}{\partial x_n} h(x_{i+k}, x_i) &= \frac{\partial}{\partial x} h(x,u)\Big|_{\substack{x = x_{i+k} \\ u = x_i}} \frac{dx_{i+k}}{dx_n} + \frac{\partial}{\partial u} h(x,u)\Big|_{\substack{x = x_{i+k} \\ u = x_i}} \frac{dx_i}{dx_n} \\ &= h_x(x_{i+k}, x_i)\delta_{i+k,n} + h_u(x_{i+k}, x_i)\delta_{i,n} . \end{aligned}$$

When expressed in the form of Eq. 6, the value inside the summations of E are symmetric with respect to i and j, so there are redundant terms when forming the product in the summations of dE/dx_n. Accounting for these redundant terms with a scaling factor of two, and rearranging,

$$\frac{dE}{dx_n} = 2 \sum_{i=N_l}^{N_u-1} \sum_{j=N_l}^{N_u} \sum_{k=N_l'-\min(i,j)}^{N_u'-\max(i,j)} \left[\alpha'_{i+k}\beta'_i h(x_{i+k}, x_i) - \alpha'_{j+k}\beta'_j h(x_{j+k}, x_j)\right] \times \alpha'_{i+k}\beta'_i \left[h_x(x_{i+k}, x_i)\delta_{i+k,n} + h_u(x_{i+k}, x_i)\delta_{i,n}\right] .$$

The lower ($\delta_{i,n}$) term sifts through the i sum in a straightforward manner. The upper ($\delta_{i+k,n}$) term can sift through either i or k. Since the i sum will be sifted in the lower equation, and will typically have more terms than the k sum (making sifting through i computationally beneficial), the i sum will be sifted through the upper term as well.

The sifting for the $\delta_{i+k,n}$ term is still not straightforward, since the limits on k are a function of i. The bookkeeping is as follows:

1. The limits on k are $N_l' - \min(i,j) \le k \le N_u' - \max(i,j)$, and at $i = j$ this range is at its maximum: $N_l' - j \le k \le N_u' - j$. For any i, the limits of k fall within this range.
2. For $\delta_{i+k,n}$, the possible values of k run over $n - N_u \le k \le n - N_l$.
3. The net permissible range of k is that range for which both of the above conditions hold, i.e., the intersection of the above ranges.

$$\left\{k : \{N_l' - j \le k \le N_u' - j\} \bigcap \{n - N_u \le k \le n - N_l\}\right\} = \{k : \max(n - N_u, N_l' - j) \le k \le \min(n - N_l, N_u' - j)\} .$$

With these new limits on k, the substitution $i = n - k$ completes the sifting process of $\delta_{i+k,n}$. If compact support of h is also included ($w_1 \leq k \leq w_2$), then the permissible values of k are

$$\{k : \max(w_1, n - N_u, N_l' - j) \leq k \leq \min(w_2, n - N_l, N_u' - j)\}.$$

The gradient elements then become

$$\frac{dE}{dx_n} = \tag{7}$$

$$2 \sum_{j=N_l}^{N_u} \left\{ \sum_{k=\max(w_1, n-N_u, N_l'-j)}^{\min(w_2, n-N_l, N_u'-j)} \left[\alpha_n' \beta_{n-k}' h(x_n, x_{n-k}) - \alpha_{j+k}' \beta_j' h(x_{j+k}, x_j)\right] \alpha_n' \beta_{n-k}' h_x(x_n, x_{n-k}) \right.$$

$$\left. + \sum_{k=N_l' - \min(w_1, n, j)}^{N_u' - \max(w_2, n, j)} \left[\alpha_{n+k}' \beta_n' h(x_{n+k}, x_n) - \alpha_{j+k}' \beta_j' h(x_{j+k}, x_j)\right] \alpha_{n+k}' \beta_n' h_u(x_{n+k}, x_n) \right\}. \tag{8}$$

2.3. Minimization

The minimization used is a version of the steepest descent technique for finding zeros of functions. Each sample position is incremented by $\Delta_{x_i}^{(m+1)}$ ($x_i^{(m+1)} = x_i^{(m)} + \Delta_{x_i}^{(m+1)}$), where

$$\Delta_{x_i}^{(m+1)} = -E \frac{dE/dx_i^{(m)}}{\sum_{j=N_l}^{N_u} \left(dE/dx_i^{(m)}\right)^2},$$

and m is the iteration counter. This expression describes the shortest Euclidean line between the current point, x, and the $E = 0$ plane, along the hyperplane which is tangent to x. This technique was used because of its simplicity — errors in the software were easily tracked because each step could be followed. This method is designed to find zeros, not minima, so it was adapted by decreasing the size (but maintaining the direction) of sample position jumps which would create an E larger than that of the previous iteration. When minima were reached, the magnitude of the sample position jumps got cut drastically ($\Delta_{x_i}^{(m)}{}_{\text{used}} \sim 2^{-40} \Delta_{x_i}^{(m)}{}_{\text{calculated}}$). If this was suspected to be a *local* minima then a large increment was permitted, effectively searching a new solution space (similar to simulated annealing).

Another constraint added to this procedure was that sample points were not to cross over one another (Let $\{x_i^{(0)}\}_{i=N_l}^{N_U}$ be an increasing sequence. $\{x_i^{(m)}\}_{i=N_l}^{N_U}$ then must be an increasing sequence $\forall m > 0$). If any pair of sample locations approached each other too fast in any given iteration step (e.g., if $[x_{i+1}^{(m+1)} - x_i^{(m+1)}] \leq 0.5[x_{i+1}^{(m)} - x_i^{(m)}]$) then $\Delta_{x_{i+1}}^{(m)}$ or $\Delta_{x_i}^{(m)}$, as appropriate, was decreased so that $(x_{i+1}^{(m+1)} - x_i^{(m+1)}) = 0.5(x_{i+1}^{(m)} - x_i^{(m)})$.

3. Results

The results of a simple trial for sample shifting section are now reported. The one-dimensional image range was normalized (i.e., the image existed for $x \in [0, 1]$). A sample space-variant "impulse response" used as a test on the system was

$$h(x, u) = \frac{1}{\sqrt{2\pi(.1)^2}} e^{\frac{(x^2-u^2)^2}{2(.1)^2}}. \tag{9}$$

$h(x, u)$ is not a true impulse response, since it is not normalized. But $h(x, u)$ provides a smooth function which possesses a clearly defined transformation, $b(x) = x^2$. Then $h(b(x), b(u)) =$

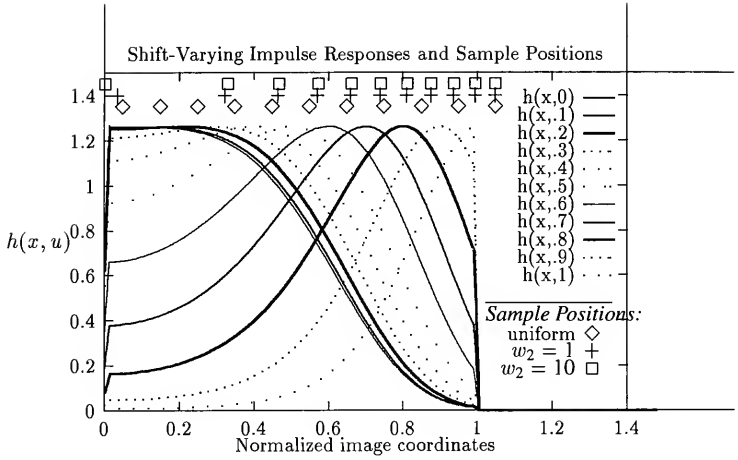


Figure 1. Impulse responses and sample positions for an example space-variant system.

$h_I(b(x), b(u))$, allowing the minimization to concern itself with finding only the sample positions, and not the multipliers α and β. The image plane was uniformly sampled at 11 points, 10 points inside the image and one point outside the right boundary of the image ($N_l = 0$, $N_u = 9$, $N'_u = 0$, and $N'_u = 10$). E was then minimized for two cases: when $w_1 = -1$ and $w_2 = 1$ (greatly restricting the range of the PSF used for calculating E), and again for $w_1 = -10$ and $w_2 = 10$ (which maximizes the range of the PSF). For the greatly restricted trial ($w_2 = 1$), the algorithm came upon a number of local minima but was eventually was able to find a global minimum. For the unrestricted trial ($w_2 = 10$), the global minimum was found directly. Both trials were run for 100 iterations. Fig. 1 shows examples of the impulse responses of Eq. 9 for different object positions, the uniformly spaced sample positions used as an initial estimate of the sample positions, and the sample positions which resulted from both the greatly and lightly restricted applications of the algorithm. Table 1 lists the sample positions and $x^2_{i+1} - x^2_i$, which should be constant for ideal sampling in this case. In the unrestricted trial, the ideal sample locations were found after 100 iterations with negligible error, as expected. But, even for the greatly restricted trial, the sample locations were were found to a good approximation. The advantage of the restriction was a substantially decreased computing time per iteration. The difficulty with the restricted trial was the abundance of local minima which needed to be negotiated.

4. Conclusion

The concepts of shift-variance and approximate shift-invariance have been presented in this paper. A metric which measures the degree of shift-variance was found and used to develop an algorithm which moved the position of image samples so that the resultant image was equivalently filtered by an approximately space-invariant system. A simple example was shown, where a transformation that made the system exactly shift-invariant was found.

The work done here is just a first step — a great deal more needs to be done. α and β need to be included in the minimization procedure. The minimization of E needs to be improved with a quicker algorithm (e.g., conjugate gradient). This would also likely include a better way to assure the continued order of the sample positions (likely by an additional constraint term to the error metric). There is a question about the types of systems on which this technique will work well. One example of trouble is an impulse response which changes topologically with object position. For

	uniform sampling		$w_1 = -1$, $w_2 = 1$		$w_1 = -10$, $w_2 = 10$	
sample number	sample location	E_{loc}[a]	sample location	E_{loc}[a]	sample location	E_{loc}[a]
0	0.05	0.02	0.0361	0.1038	0.0046	0.1101
1	0.15	0.04	0.3242	0.1131	0.3318	0.1103
2	0.25	0.06	0.4671	0.1104	0.4695	0.1102
3	0.35	0.08	0.5732	0.1100	0.5750	0.1103
4	0.45	0.10	0.6622	0.1103	0.6640	0.1103
5	0.55	0.12	0.7408	0.1102	0.7424	0.1103
6	0.65	0.14	0.8118	0.1106	0.8133	0.1101
7	0.75	0.16	0.8773	0.1093	0.8784	0.1103
8	0.85	0.18	0.9375	0.1117	0.9391	0.1103
9	0.95	0.20	0.9953	0.1119	0.9961	0.1103
10	1.05	na	1.05	n/a	1.05	n/a

Table 1. Sample positions and quadratic differences in the normalized image.

[a] $E_{loc} = x_{i+1}^2 - x_i^2$ (consistency check). If the sample locations for the quadratically-varying Gaussian impulse response are correctly identified, E_{loc} should be constant for all i.

example, if an impulse response has no bumps in it (is monotonically increasing until its maximum, and monotonically decreases thereafter) in one position, but has bumps in it in another position, no amount of stretching or contracting will make them identical. Also, until this technique is tried on real images, it remains unknown how this kind of restoration affects image quality although, for the examples of exact transformations in Robbins and Huang (1972) and Sawchuck (1975), the results appear promising). The propagation of noise through the system, concentrating on the statistics of the noise in the warped image plane, needs to considered (Robbins and Huang just assumed that the noise in the warped image plane was additive and stationary). It is unknown, for example, if an MMSE solution in the warped image plane relates to an MMSE solution in the real image plane. The effects of performing the resampling from interpolated uniformly sampled images, rather than using perfect sample values at new positions, needs to be considered. Extending this procedure to two-dimensional signals appears to be straightforward, but may cause further constraints upon the types of systems for which this method will work well.

References

Andrews, H. C., & Hunt, B. R. 1977, *Digital Image Restoration*, Prentice-Hall, Englewood Cliffs

Gaskill, J. D. 1978, *Linear Systems, Fourier Transforms, and Optics*, Wiley, New York

Gerchberg, R. W. 1989, IEEE Trans. on Acoust., Speech and Sig. Proc., 37, 1603

Robbins, G. M., & Huang, T. S. 1972, Proc. IEEE, 60, 862

Jain, A. K. 1989, *Fundamentals of Digital Image Processing*, Prentice-Hall, Englewood Cliffs

Landweber, L. 1951, Am. J. Math., 73, 615

Papoulis, A. 1975, IEEE Trans on Circ. and Sys., 22, 735

Robbins, G. M. 1970, Pattern Recognition, 2, 91

Sawchuck, A. A. 1972, Proc. IEEE, 60, 854

Sawchuck, A. A. 1973, Journal Opt. Sci. Am. 63, 1052

Sawchuck, A. A. 1974, Journal Opt. Sci. Am. 64, 138

The Restoration of HST Images and Spectra II
Space Telescope Science Institute, 1994
R. J. Hanisch and R. L. White, eds.

Improved Prescription Retrieval and PSF Modeling Code

David Redding, Meemong Lee, and Sam Sirlin

Jet Propulsion Laboratory, California Institute of Technology, 4800 Oak Grove Drive, Pasadena, CA 91109–8099

Abstract. The success of HST image restoration depends critically on the ability to obtain, either through direct observation or through detailed optical modeling, a good estimate of the point spread function. PSF models are, in principle, superior to observed PSFs owing to the lack of noise and the ability to compute the PSF on a higher density pixel grid. We describe an approach to PSF modeling for the HST which utilizes two complementary programs: (1) a sophisticated prescription retrieval code, and (2) a hybrid PSF modeling code that combines both ray-tracing and diffraction propagators. The PSF modeling code will be made available to the HST user community through the ST ScI.

1. Introduction

Image restoration is the process of removing artifacts caused by telescope imperfections from images. A good understanding of these imperfections is essential to the process. As part of the ST ScI Image Restoration Project, and continuing under new NASA funding, we are exploring improved ways for determining telescope imperfections, and for using that information to compute accurate point spread functions (PSFs) for restoration. We will be making the products of our work available to the community.

The core of our approach is a sophisticated optical modeling code that generates accurate PSFs directly from telescope optical prescription data. The prescription specifies the physical state of the telescope, namely the location, orientation and figure of the lenses, mirrors, obscurations, and detectors that comprise its optics. Our code utilizes a hybrid geometric and physical optics approach to generate PSFs from this information. It captures the essential physics of the image formation process, including the effects of induced aberrations and diffraction from multiple obscuring surfaces. It provides a rigorous basis for accurately predicting PSFs over widely varying field angles and focus settings without the need for data matching at each point of interest.

The accuracy of this approach is limited by the accuracy of the prescription information. A good starting-point prescription is usually available from the telescope design specifications and from post-fabrication optical test. Changes to the telescope due to environmental effects (or manufacturing errors) can occur, however. In this event, an improved estimate of the prescription data can be obtained using a technique termed "prescription retrieval."

Prescription retrieval "inverts" image data to determine the prescription of the telescope that made the images. It is a nonlinear parameter estimation technique, which works by varying prescription parameters so as to match diagnostic images with simulated images. It is capable of identifying optical element parameters that affect image location and quality, subject to limitations that we discuss below. Prescription retrieval works best with deliberately defocused, off-axis diagnostic images. We successfully developed prescription retrieval as a tool for determining the conic constant of the HST primary mirror (Redding, Dumont, and Yu 1993).

For image restoration purposes, prescription retrieval to determine a large number of telescope parameters will be required once following a major configuration change of a telescope. For HST, this means post-launch, post refit, and possibly following resetting of the secondary mirror decenter.

Parameters to be identified include camera/telescope alignments and camera and OTA mirror figure parameters.

During normal operations of the telescope, minor prescription retrieval, probably to identify a single time-varying parameter such as telescope focus, may be desirable. For HST, this would provide a means of handling the effects of OTA "breathing." This could be performed using the science images — diagnostic images are not required. Once prescriptions are determined for the telescope at a particular time, they can be archived and should be useful for restoring images taken at about that time.

Tracking the telescope prescription over a period of time should also help identify the root causes of changes, such as thermal cycling. It may then be possible to implement corrections to the prescriptions to be used in restoration based on the telemetry stream. For instance, temperature measurements could be used to adjust assumed secondary mirror position, yielding improved PSFs without retrieval.

The immediate product of our work will be a computer program for generating PSFs for each of the HST cameras, both pre and post-refit. This program will use the best current prescription data, and will incorporate retrieved prescriptions as they become available. It will be available from ST ScI early in 1994.

Our ultimate objective, which we will pursue over the next three years jointly with the ST ScI, is to provide a high-performance integrated prescription-based image restoration application. This application will provide functions and features such as PSF generation, prescription retrieval, prescription archiving, and restoration using spatially-variant PSFs. The application will be portable to different platforms and different architectures, including desk-top workstations and massively parallel supercomputers. When completed the application will support distributed processing across heterogeneous computing environments, allowing efficient utilization of networked workstations and supercomputers. We hope it will provide a flexible tool for rapid, high-accuracy restoration of images from HST and other instruments.

2. Image Formation

Ideal image formation can be modeled as the convolution of the true brightness distribution of the object with the telescope point spread function. The PSF is the spatial frequency impulse response of the telescope, which is to say, the image of an ideal point source such as a star. An approximate model of the PSF can be computed as the Fourier transform of the field in the exit pupil of the telescope. For unaberrated, unobscured telescopes, this pupil function consists of a spherical wavefront centered on the detector, with uniform intensity within the geometrically-defined circular aperture. The resulting PSF has the well-known form of the Airy pattern.

Real telescopes have PSFs that differ from the Airy pattern ideal. Optical aberrations due to misalignments, thermal gradients, outgassing, manufacturing errors and other effects change the configuration of the optics, causing the phase of the wavefront propagated through the telescope to deviate from a perfect sphere. This results in decreased amplitude and increased width of the PSF, causing blurring of the image and decreasing the sensitivity of the instrument. Vignetting or shadowing incurred at stops, spatial filters, and obscurations such as mirror support pads, spiders and secondary mirrors, further decreases the PSF amplitude and can alter the halo of the PSF, adding diffraction rings and tendrils. These effects are very apparent in images taken with the first set of HST instruments. They are also present, albeit at much lower levels, in images taken with the refurbished HST cameras.

For most telescopes, wavefront phase and vignetting effects vary as a function of field angle (or focus setting). These induced aberrations cause the PSF to vary both in structure and amplitude from point to point over the detector. Again, HST images show the effects of strong spatial variance of the PSF. Detector characteristics are also significant. These include pixel size as well as non-ideal behaviors such as spatially-variable quantum efficiency or pixel cross-talk. Other effects, such as high-frequency telescope jitter or atmospheric turbulence, can cause smearing of the PSF.

Image restoration can compensate for telescope aberrations, induced aberrations and detector effects by deconvolving a "known" PSF from the image data. Known jitter and turbulence effects can be compensated similarly. The variation of the PSF over the field can be handled by using multiple PSFs, each valid within a sub-region of the detector. The result of restoration using accurate PSFs is a recovered image whose accuracy is limited only by the noise content of the original image. These noise effects can be reduced by taking several images of the same source and averaging.

The success of image restoration depends critically on the accuracy of the known PSF. Our objective is to provide good predicted PSFs for restoration of HST images. To do this the PSF generation code should account for the effects of:

- Aberrated optics. Figure errors, such as the OTA primary conic constant error, and zonal aberrations from optical test data.
- Misaligned optics.
- Induced aberrations, both phase and vignetting effects.
- Detector errors, such as pixel cross-talk and variable quantum efficiency.

Detector errors will be more significant for the refurbished cameras than for the original HST instruments (e.g., Burrows 1994).

3. Prescription-Based Optical Modeling

An ideal model of the image formation process would integrate Maxwell's equations directly to solve for the fields in the telescope. This is quite a formidable task. Finite-element solutions of Maxwell's equations work for simple instrument geometries, but require very large amounts of computation, and do not provide a practical aproach for Hubble. Fortunately, simpler theories are available that are quite good.

Optical designers have long relied on geometric optics, or ray-tracing, to predict performance of optical instruments. Geometric optics makes the assumption that the wavelength of the light is infinitesimally small compared to the diameter of the beam. This turns out to be a very good assumption while the beam is away from a focus. Geometric optics provides an accurate means of determining both the wavefront phase and vignetting patterns in the expanded beam of a telescope, even in highly aberrated systems. It does not require simplification of the geometry of the instrument, and works directly from the optical prescription. It enables accurate computation of the effects of figure errors, induced aberrations and misalignments. Ray tracing alone is not capable of reproducing images in the far field with any accuracy, however.

For modeling images and highly diffracted beams (such as laser cavity modes) physical optics theories offer an alternative. Fresnel physical optics makes a different set of assumptions, such as time-invariance and slowly-varying phase, to provide a simplified set of propagation integrals that can be solved rapidly using Fourier transform techniques. As commonly implemented, physical optics codes unfold an optical system into a simplified linear sequence of "thin lenses," which collapse the phase effects of optical elements into a plane. Single-plane PO models then lump all of the phase effects onto a single plane at the exit pupil of the telescope; images are created by propagating the exit pupil irradiance to the far field, at the detector. Multiple-plane physical optics models propagate from planar element to planar element in the near field, and then to the far field at the detector. Neither approach allows for accurate prediction of induced aberrations without phase retrieval. Multiple-plane models do allow for multiple planes of obscuration, which can be important in capturing details of the point spread function.

We have implemented a hybrid approach, which exploits the advantages of both geometric and physical optics. We use ray-tracing to track the phase and determine vignetting in the expanded beam (Fig. 1). This information is used to drive Fresnel near- and far-field diffraction propagators, which compute the beam complex amplitude and, ultimately, the image or PSF. This approach

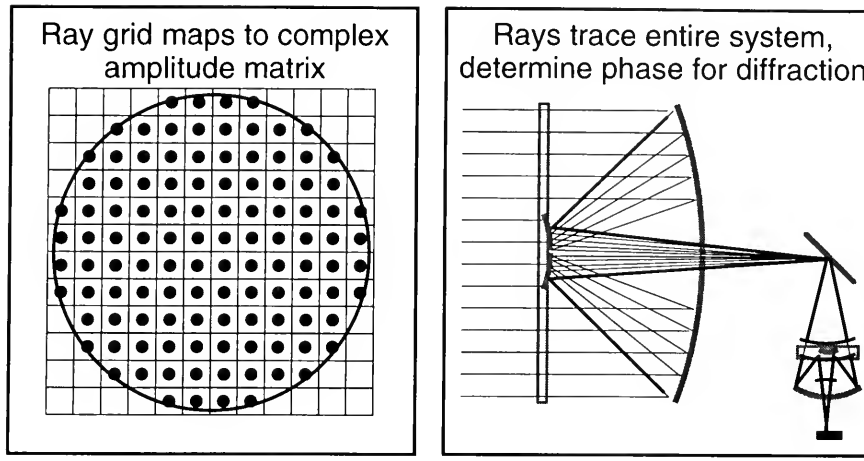


Figure 1. Sketch of a typical telescope.

accurately determines the effects of vignetting, aberrations and induced aberrations, based on the actual optical instrument geometry and prescription. It also enables us to model multiple planes of obscuration and other important diffraction effects.

This hybrid approach is realized in the Controlled Optics Modeling Package (COMP) (Lee et al. 1990), which is available from the NASA COSMIC software library (COSMIC can be reached at (706)542-3265, or by e-mail at service@cossak.cosmic.uga.edu). COMP is a full-featured optical modeling code that can be used either as a stand-alone application, or, through a subroutine interface, to provide optics functions in-line to other programs (Table 1). The subroutine version is called SCOMP. The point-spread function generator we are writing incorporates SCOMP, providing a simplified user interface and reduced function set specialized to the HST cameras.

COMP associates a physical optics complex-amplitude matrix with a geometric optics ray grid (Fig. 1), so that each ray acts as a phase and obscuration probe for a corresponding diffraction cell in the beam. Ray-tracing determines the wavefront aberrations and the vignetting effects of obscurations based on the instrument physical configuration and field angle. Phase errors due to figure errors, misalignments or other optical aberrations are introduced into the modeled system in the same way they occur in the actual system. Fresnel near- and far-field diffraction algorithms use the optical path length information to propagate the complex amplitude matrix through the system to the focal plane to generate the PSFs. Multiple obscuring surfaces can be specified. COMP incorporates various features to ensure correct sampling for the diffraction calculations and to avoid aliasing. Polarization effects can be included using polarization ray-trace and vector diffraction functions.

Using COMP to generate PSFs for image restoration allows us to avoid some of the limitations of a pure physical optics approach. Prescription-based PSF generation uses a fundamental parameterization that is capable of predicting accurate PSFs across the entire operational range of the instrument. Equivalent performance using a pure physical optics approach would probably require iterative data-matching or prescription retrieval at each field point. The ability to accurately predict PSFs is especially crucial for restoration of extended-object images, where objects for matching are not available.

Table 1. COMP capabilities.

Analysis Capabilities	Element Types
Geometric optics: ray-trace	Surface types: Conic, conic plus Zernike-specified deformations, conic plus data-specified deformations, general aspheres
Polarized light: polarization ray-trace	Optic types: Reflective, refractive, diffractive (HOEs), obscuring, polarizing
Linear analysis: differential ray-trace	Lenticular arrays
Physical optics: Scalar and vector Fresnel near- and far-field propagation	Segmented mirrors
Composite-object image generation	Spatial filters
Multi-spectral image generation	Pixel-array detectors
Linear optical model generation and export	Reference and return surfaces
COMP can be used as a stand-alone application or as a subroutine	

4. Prescription Retrieval

Performance of prescription-based PSF generation is limited by the accuracy of the prescription data. Prescription retrieval provides a means of improving estimates of prescription parameters by computer processing of image data (Fig. 2). Prescription retrieval takes as a starting point the optical design prescription, including the location and orientation of each optical element specified in global coordinates, plus figure (conic constant, radius, and Zernike or other deformation parameters) and index. Higher spatial-frequency figure errors determined by optical testing of the elements are added to the respective element surfaces. Obscurations are placed at the appropriate physical coordinates.

The prescription retrieval process takes image data as input. It uses SCOMP functions to generate images that match the data on a pixel-by-pixel basis. Residual pixel differences are minimized in an iterative optimization loop that varies selected prescription parameters. We have used Levenberg-Marquardt nonlinear least-squares and the NPSOL sequential quadratic programming algorithms with good success. The result, when the process is converged, is improved estimates of those prescription parameters. The best results are obtained using several images simultaneously, preferably images that use different field and focus points. Image diversity improves the observability and separability of individual parameters.

We developed prescription retrieval as part of the Hubble Aberration Recovery Project. The main objective here was to obtain an accurate estimate of the true conic constant of the OTA primary mirror, which is the source of most of the spherical aberration seen in pre-refit HST images. It turned out that other effects were significant also, especially conic constant errors in the WFPC repeater cameras. Alone among several groups analyzing the HST images we were able to determine those effects and separate them from the primary mirror errors (Redding, Dumont, and Yu 1993). The resulting estimate of the primary conic constant was in agreement with the results of workers who examined the HST primary mirror fabrication and test hardware, and is validated by the excellent quality of the images from the refurbished cameras.

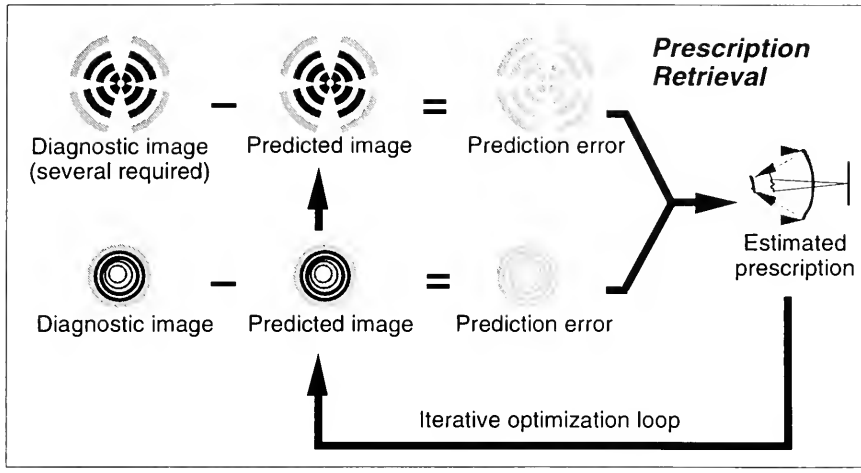


Figure 2. Prescription retrieval is an iterative optimization process.

Example simulated and actual Hubble/PC6 images are shown in Fig. 3. The simulated image was computed tracing 65,000 rays to update a 512×512 diffraction grid. Two images, taken at different focus and field settings, were used in the retrieval process (the second is not shown). A total of about 40 parameters were estimated, including various element alignments and figure parameters. Convergence required many days running on a SPARC 1+ workstation.

The images show good agreement with respect to major features such as the spherical aberration rings and the location and structure of the diffracted obscuration patterns. The match is not perfect and can be improved by the addition of the primary mirror zonal aberrations and other effects.

We also applied prescription retrieval to diagnostic star images that were returned by the Mars Observer Narrow-Angle Camera during its approach to Mars. This camera was equipped with heaters on its primary mirror that, when activated, cause a nearly pure focus change to be imparted to the system. A typical example image showing simulated and actual data is given in Fig. 4. Retrieval here was complicated by limited dynamic range and noise contributed by a lossy data compression algorithm used in the approach phase of the flight.

Results of prescription retrieval using 7 images taken on both sides of focus and across the camera field indicated that the primary mirror had deformed slightly, with the main error being 45-degree astigmatism. Other possible errors, such as decenter of the secondary mirror, did not accurately reproduce the observed images. Performance at the best focus setting was good, however, with a 1–3 pixel PSF.

There are several potential problems to be considered when applying prescription retrieval. These are:

Noise Detection noise, background noise, self-noise, stray light and similar effects obscure structure that can be important in the retrieval process.

Local minima Nonlinear least squares and similar optimization algorithms can be trapped by local minima. The likelihood of this is reduced with a sufficiently diverse data set and a good initial guess.

Inseparable parameters If two parameters have a proportionally similar effect on the images, the contribution of one cannot be distinguished from that of the other. Separability can be

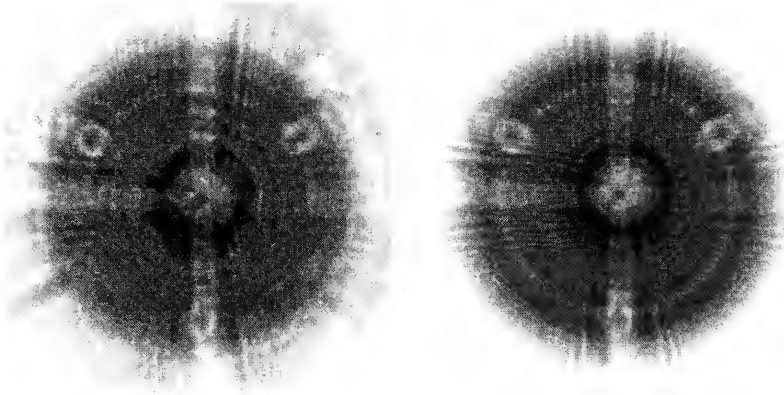


Figure 3. Actual and simulated HST PC6 diagnostic image.

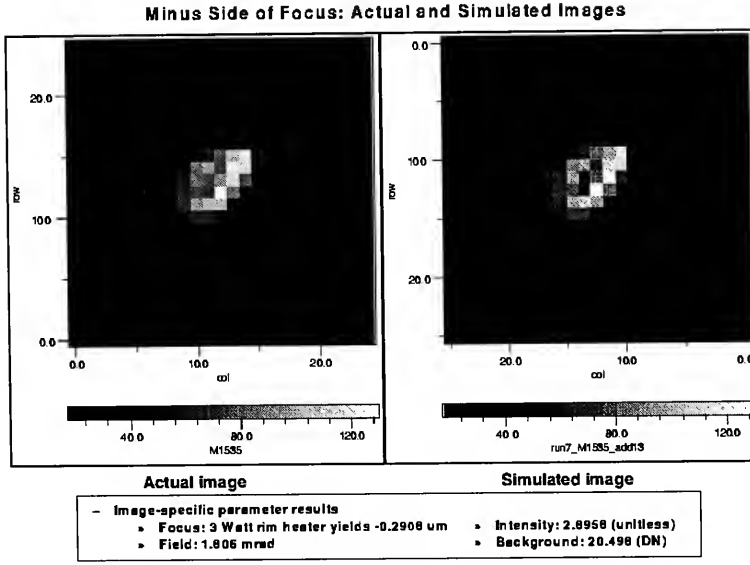


Figure 4. Actual and simulated Mars Observer Camera image.

improved by looking at different field angles and focus positions. Some parameters that cannot be separated, need not be separated – their effect on the image is accounted for without knowing precisely the contribution of each parameter individually.

Inadequate model It is important that major factors in the imaging process be captured by the model. This means getting the physics right, both in the optics and in the detectors. It also includes correctly identifying which factors are common to particular images, and which change between images, so they can be correctly solved for. Unmodeled changes between images, such as incomplete pyramid mirror rotations or thermal deformations of the telescope structure, add confusion and reduce the quality of the solutions.

Meaningless parameters Solving directly for noise or jitter was found to improve the optimization cost function without improving the accuracy of individual parameter estimates.

Besides its role in image restoration, prescription retrieval also has significant advantages for optical test and calibration, and for active optical control. These advantages include:

- It provides the ability to optically test aspheric elements or systems without null correctors or interferometers, removing a major potential source of testing error.
- It provides end-to-end instrument testing, including the science detector.
- It is a non-invasive testing procedure that does not require instrument disassembly.
- It is applicable over the full operational range of the instrument, rather than at 1 or 2 field points.
- As we have demonstrated, it can be performed remotely on instruments in the field, as well as in a laboratory environment.
- For instruments equipped with low-bandwidth controllable optics, such as deformable or movable mirrors, prescription retrieval can determine corrections based on image data.

Prescription retrieval does require a large amount of computational power. The best results require simultaneous processing of numerous images while exercising an appropriate number of prescription parameters. This requires a large amount of memory as well as high-speed computation.

For most instruments, prescription retrieval will be performed intensively once, to provide the best possible understanding of the basic instrument configuration. This initial retrieval will utilize diagnostic star images taken at multiple focus and field points. This computationally intensive process need not be repeated unless the instrument radically changes.

Subsequent limited prescription retrieval might be optionally performed on a reduced parameter set, without the need for special diagnostic images. Here the objective would be to account for small, localized shifts that are known to occur relatively frequently, such as the HST thermally induced "breathing" effects. This process will be much less computationally intensive. Retrieved prescription data will be archived by the application and referenced by date, so that appropriate prescriptions will be available for recovery.

5. What We Will Provide the Community

Our work is continuing under two separate programs. We have a small amount of funding from ST ScI to provide a HST-specific point spread function generation code. This code will not perform prescription retrieval. It will be released (early in 1994) with the current best prescription data describing each of the WF/PC–1, FOC, WF/PC–2 and COSTAR/FOC cameras. It will be available as a stand-alone application, generating PSFs in an interactive session. These can be exported as FITS or binary format files for use in restoration. The code will be available in executable form for

UNIX workstations. The code will also be available as a FORTRAN subroutine that can be called directly by a user code. For further information, contact Bob Hanisch at ST ScI (hanisch@stsci.edu).

We are also beginning a new task in collaboration with ST ScI, under the NASA Advanced Information Systems Research Program, to develop an integrated prescription-based image restoration application. The program will be written to make efficient use of a variety of new computer architectures, ranging from scalar workstations to one or more of the new massively parallel supercomputers. It will provide functions and features such as:

- PSF generation. PSFs will be generated and exported in several formats.
- Prescription retrieval.
- Prescription archiving.
- Other optical design and analysis functions.
- Image restoration using spatially-variant PSFs. Two or three of the best-performing algorithms from the ST ScI Image Restoration Project will be modified to work with space-variant PSFs and included in the application.
- Easily programmed scripting language.
- Easily extensible by users to implement alternative algorithms, optical models, or prescription or phase retrieval techniques.
- Usable without modification for most optical instruments.

The application will exploit recent developments at JPL in parallel processing and distributed computing to provide a major increase in processing power, while retaining source-code portability, a user-friendly interface and the convenience of access through UNIX desk-top workstations. The user interface will provide a near-plain English command language and macro capability. Extensive graphical user interface features will be included.

The application will exploit an architecture-independent parallel programming environment that allows source code level portability for applications on various distributed memory architectures (Lee et al. 1990). This is achieved through an object-oriented data distribution function library, offering multiple distribution types. An application programmer can choose a proper distribution type according to the data access pattern of an application without considering the architectural constraints of a parallel system. We will exploit a message passing interface protocol (Dongarra et al. 1993) to achieve source-code level portability to various user environments so that a flexible mapping between the task and all available resources at the user site can be performed with one application program. Only a limited amount of machine-specific coding will be required to implement the application on a new architecture.

The code will be developed and released in phases, with a first (scalar) version available in about a year. When completed (in three years) the application will support distributed processing across heterogeneous computing environments, allowing efficient utilization of networked workstations and supercomputers.

In parallel with the code development, we will be performing extensive validation and verification of the various functions of the code. The performance limits of prescription retrieval with respect to noise, data diversity and other factors will be explored. Performance of prescription-based restoration will be compared to other methods. Working with images from the refurbished HST images, we will be pursuing ambitious objectives, such as recovery of the "fuzz" (underlying galaxies) of quasars, detecting faint companions of nearby stars, and extending the magnitude range over which stellar photometry can be done. Our ultimate objective is to make the very best quality restoration a matter of routine for all images from instruments such as HST.

References

Burrows, C. 1994, this volume

Dongarra, J. J., Hempel, R., Hey, A. J. F., & Walker, D. W. 1993, "A Proposal for a User-Level Message Passing Interface in a Distributed Memory Environment"

Lee, M., Groom, S., Mazer, A., & Williams, W. 1990, JPL Pub. No. 88–32, Rev. 1, Vol. 1–3

Redding, D., Dumont, P., & Yu, J. 1993, Appl. Opt., 32 (10), 1728

Redding, D., Needels, L., Wallace, K., Yu, J., & Levine, M. 1992, "Controlled Optics Modeling Package User Manual," Ver. 1.0, JPL Doc. D–9816

The Restoration of HST Images and Spectra II
Space Telescope Science Institute, 1994
R. J. Hanisch and R. L. White, eds.

Better HST Point-Spread Functions: Phase Retrieval and Blind Deconvolution

Richard L. White

Space Telescope Science Institute, 3700 San Martin Drive, Baltimore, MD 21218

Abstract. The accuracy of HST image restorations is often limited by the quality of point-spread functions that are available. Phase retrieval methods are being applied to HST PSFs in an effort to improve the agreement between observed PSFs and PSFs computed using optical modeling programs. The use of blind deconvolution to make better use of observed PSFs is also discussed.

1. Introduction

The most difficult practical problem to be solved when restoring HST images is usually not the choice of a restoration algorithm, but rather finding a good point-spread function (PSF). Even the most sophisticated algorithm cannot produce a decent restored image using a bad PSF, and even the least-sophisticated algorithm will do a pretty good job given a good PSF.

The problem is that the PSF changes as a function of wavelength, position in the camera field of view (for the WF/PC, WFPC2, and FOC with COSTAR), and time. There are enough variables that it is not practical to maintain a library of high-quality, observed PSFs covering all possibilities. Computed PSFs using programs such as Tiny Tim (Krist 1993) present a more attractive alternative because they can be computed at any wavelength and camera position, they are noise-free, and they can be computed on sub-sampled pixel grids. Unfortunately the PSFs computed with existing optical modeling programs are usually in relatively poor agreement with the observations, so that they are useful only for deconvolving observations with fairly low signal-to-noise ratios. We are pursuing a number of possible improvements for calculated PSFs, including more sophisticated optical models (Redding et al. 1994) and restoration methods that can adjust the PSF to get better restored images (see the discussion of blind deconvolution in §6.)

Another promising approach is to use *phase retrieval* to develop better optical models for HST. In this paper I briefly introduce the topic of phase retrieval and show some results that have been obtained for HST images.

2. What Is Phase Retrieval?

The PSF for an optical system is determined by the amplitude and phase of the (approximately) spherical wavefront as it converges on the point of focus. The amplitude $A(u,v)$ measures the intensity of the wavefront at each point (u,v) on the sphere and is usually approximately uniform across the entire pupil, except where it is obscured by objects in the light path such as the secondary mirror and its supports. The phase $\Phi(u,v)$ measures the deviation of the wavefront from the sphere (a perfectly focussed wavefront has zero phase error). Usually the phase error is measured in units of the wavelength of light being observed.

The PSF P for the given amplitude and phase is

$$\begin{aligned} P(x,y) &= \left| \int\int du\, dv\, A(u,v) e^{i2\pi[ux + vy + \Phi(u,v)]} \right|^2 \\ &= \left| \int\int du\, dv\, W(u,v) e^{i2\pi(ux+vy)} \right|^2 . \end{aligned} \tag{1}$$

Stated simply, the PSF is the square of the amplitude of the Fourier transform of the complex pupil function, $W(u,v) = A(u,v)\exp[i2\pi\Phi(u,v)]$. Note that this equation assumes that the wavefront is not too strongly curved over the pupil; if the curvature of the wavefront is large one must use a Fresnel transform rather than a Fourier transform.

Phase retrieval is the process of trying to recover the wavefront error $\Phi(u,v)$ (and possibly the amplitude A as well) given a measurement of the PSF. Phase-retrieval methods have been used since the discovery of the aberration in the HST primary mirror (Burrows et al. 1991, Fienup et al. 1993) to characterize the HST optical system. Previous phase-retrieval efforts have been aimed mainly at an accurate measurement of the spherical aberration. We are now using similar techniques (and some newly developed methods) to try to improve our understanding of the HST and WFPC/FOC optical systems. Our goal is to be able to compute better PSFs for use in image restorations.

Phase retrieval has much in common with deconvolution, and many of the techniques used for image restoration have counterparts for phase retrieval. However, the equation relating the phase and the observed PSF is non-linear in the phase-retrieval problem, which makes phase retrieval considerably more difficult than image restoration. A particular problem in phase retrieval is that maximum likelihood approaches to finding the phase tend to get stuck at local maxima of the likelihood rather than finding the globally best solution. Fienup and Wackerman (1986) discuss this and other phase-retrieval problems.

3. A New Phase Retrieval Iteration

I have been using a new phase retrieval iteration that is derived from the maximum likelihood equation for Poisson statistics in the data. The joint likelihood $\mathcal{L}$ of getting the observed counts $D(i)$ in each pixel given the expected counts $I(i)$ is

$$\ln \mathcal{L} = \frac{1}{N} \sum_i D(i) \ln I(i) - I(i) - \ln D(i)! \,. \tag{2}$$

Here the continuous variables (u,v) and (x,y) have been replaced by discrete variables, and the equations are written as one-dimensional for simplicity of notation. The image I is equal to the PSF times a constant. If we allow the amplitude of the complex pupil function W to vary, then the maximum likelihood solution occurs where all partial derivatives of $\mathcal{L}$ with respect to $W(j)$ are zero:

$$\begin{aligned} \frac{\partial \ln \mathcal{L}}{\partial W(j)} &= 0 \\ &= \frac{1}{N} \sum_k \left[\frac{D(k)}{P(k)} - 1 \right] e^{-i2\pi jk/N} \sum_l W_l e^{i2\pi lk/N} \,. \end{aligned} \tag{3}$$

The iteration is simply

$$W_{new}(j) = W(j) + \frac{\partial \ln \mathcal{L}}{\partial W(j)} \,. \tag{4}$$

This is a "gradient step" for updating the complex pupil amplitude W. If one wants only the phase Φ to to vary during the iteration, holding the pupil illumination function A fixed, it is possible to renormalize W to a constant amplitude after each iteration.

4. Examples

An example of how phase retrieval can improve computed PSFs is shown in Fig. 1. At this wavelength in the middle of the optical band, the Tiny Tim PSF is in reasonably good qualitative agreement with the observations. However, there are some discrepancies between the observed PSF and the Tiny Tim PSF; note especially that the shape of the PSF core is not well modeled and the position and brightness of the bright ring are not exactly correct. The most likely explanation is that

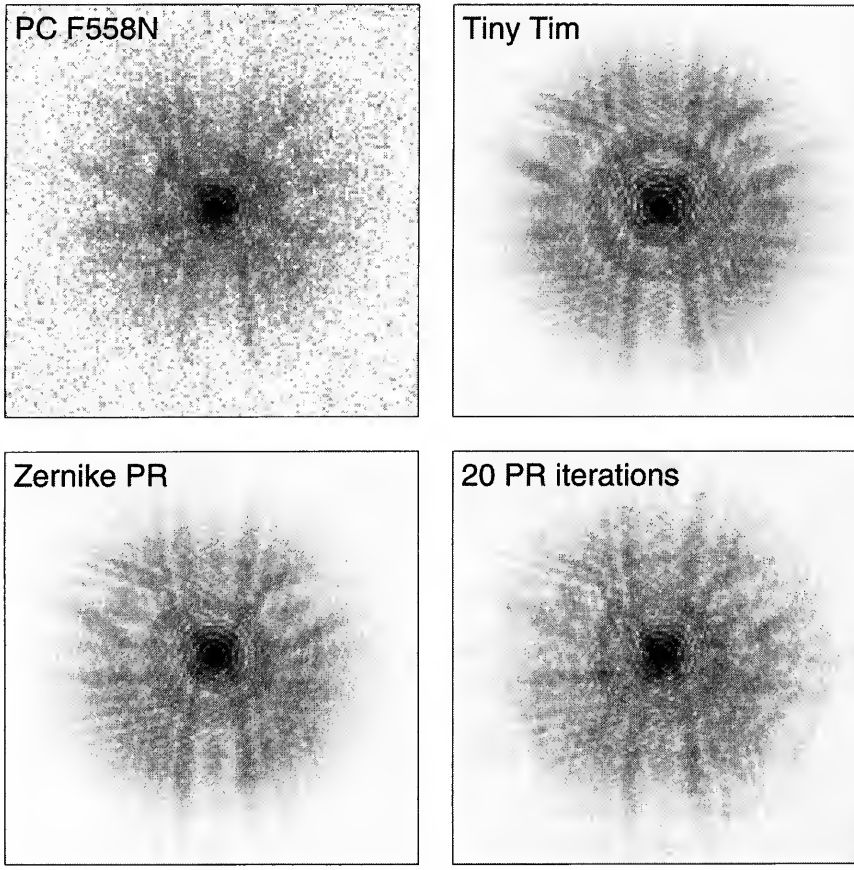


Figure 1. Results of phase retrieval for PC F588N PSF. Top left: observed PSF. Top right: Tiny Tim PSF computed using 3 wavelengths across filter bandpass. Lower left: phase retrieval result using low order Zernike polynomials to model wavefront errors. Lower right: phase retrieval result using 20 iterations of the Poisson maximum likelihood method. Initial guess was taken from the Zernike phase retrieval model.

the mirror map and aberration model used by Tiny Tim is not accurate enough. Any discrepancies in the optical model become more obvious at shorter wavelengths.

The simplest phase retrieval method is to model the wavefront error $\Phi(u, v)$ using a low order polynomial in (u, v). For HST the appropriate polynomials are the Zernike polynomials given in the *HST Optical Telescope Assembly Instrument Handbook* (Burrows 1990). The various Zernike polynomial coefficients correspond to wavefront tilt (which shifts the position of the object), focus, spherical aberration, astigmatism, and so on.

The lower left panel of Fig. 1 shows the result of adjusting these coefficients to fit the observed PSF. A conjugate gradient optimization method was used to search for the set of Zernike coefficients that maximizes the Poisson log likelihood. The agreement with the observations is improved, though there are still differences. For example, note the bright knot in the observed PSF where the tendril trailing off towards 8 o'clock crosses the bright ring. Also note that the outer dark ring, seen easily in the Zernike PSF, is hardly visible at all in the observed PSF. Other differences are readily visible when the PSFs are blinked on an image display.

The agreement with the data can be improved further by solving for a map of wavefront errors across the pupil using the method described above. The results of 20 iterations of such a method are shown in the lower right panel of Fig. 1. Note the improvement in features such as the strength and position of the "tendrils" that extend downward from the center of the PSF.

Fig. 2 shows the wavefront error maps derived by the two phase retrieval methods.

5. Using PSFs for Image Restoration

Fig. 3 shows the result of deconvolving the observed PSF using the three computed PSFs. Differences in the structure left in the PSF halo are apparent. There are also substantial differences in flux conservation and in the sharpness of the stellar images among the different restorations, as shown in Table 1 below.

Table 1. Properties of Restored Images

PSF	Total Counts	Peak Brightness
Tiny Tim	84,000	45,400
Zernike	80,900	52,400
PR Iteration	78,900	62,900

The observed PSF has a total of 74,400 counts, so flux is conserved better with the phase-retrieval PSFs. (The lack of exact flux conservation is due to the CCD readout-noise in the WF/PC data. In the absence of readout noise, the Richardson-Lucy iteration conserves flux exactly.)

6. Blind Deconvolution

Holmes (1992) has derived an iterative algorithm for finding the maximum likelihood solution for both the unblurred image and the PSF in the presence of Poisson statistics. This problem is usually referred to as *blind deconvolution*. A maximum likelihood approach to blind deconvolution is certain to fail unless one can place very strong constraints on the properties of the PSF. For example, if one applies the Holmes iteration to HST images with no constraints on the PSF, the solution found is that the unblurred image looks exactly like the data and the PSF is a delta-function. (A perfect fit to the data, but ludicrously far from the truth!) If one applies a band-limit constraint on the PSF based on the size of the HST aperture, the result is little better: the PSF solution looks like a diffraction-limited PSF rather than the true spherically aberrated PSF.

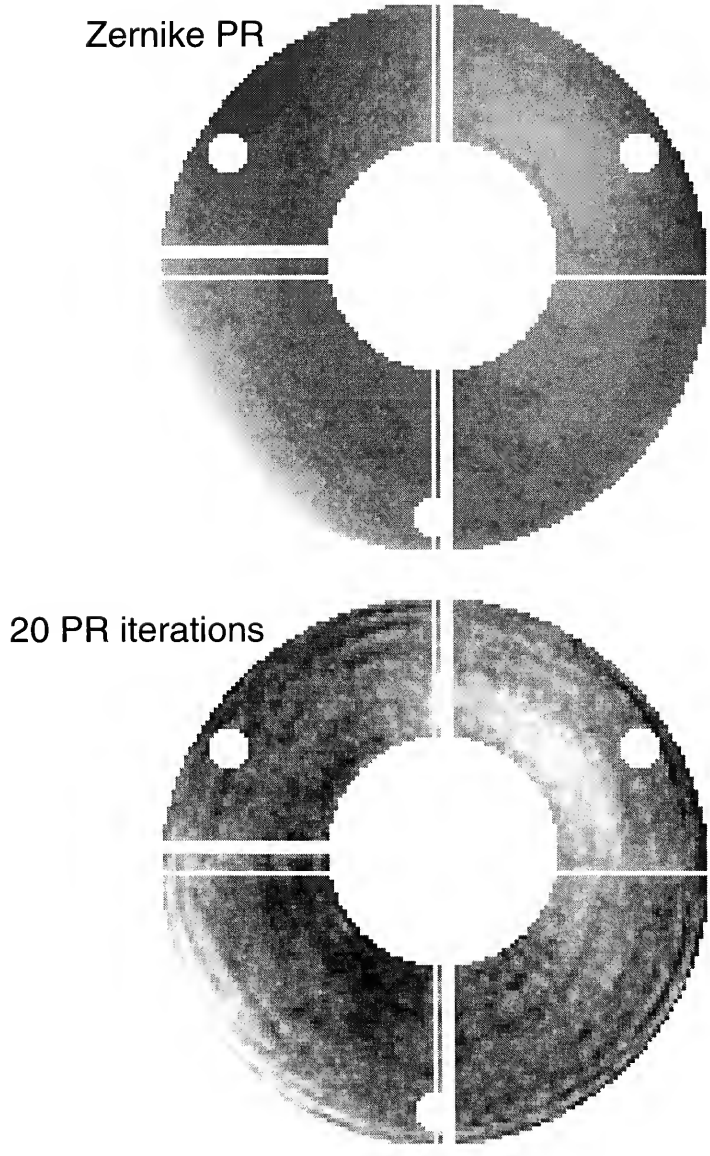


Figure 2. Maps of phase error found by Zernike phase retrieval and maximum likelihood method. The white areas are obscured by the HST secondary and supports, the WF/PC secondary and supports, and the pads on the HST primary mirror. The spherical aberration and focus terms have been removed to make these smaller aberrations easily visible. Typical wavefront variations are about 0.1 waves.

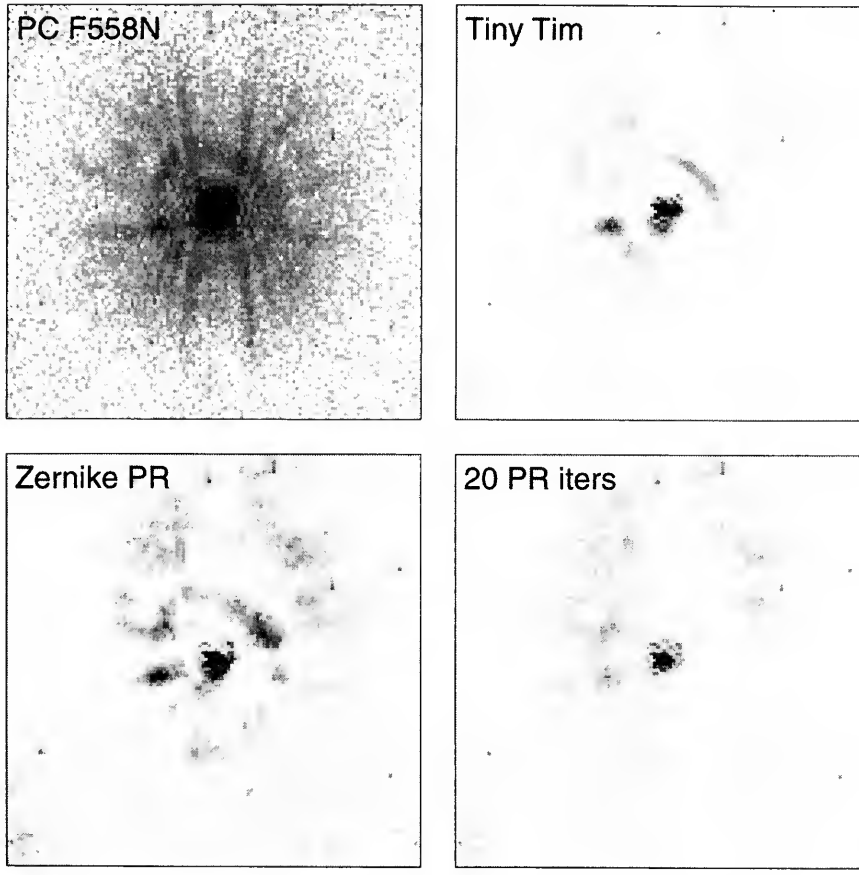


Figure 3. Deconvolution of observed PSF using computed PSFs with Richardson-Lucy iteration.

If one has either a star within the image or a separate PSF observation, one can do much better by forcing the selected PSF objects to be point sources in the restored image. One simple approach is to start with an initial guess for the image that has a delta-function at the position of the star and is zero in the other nearby pixels. Then it is really possible to get estimates of both the unblurred image and the PSF from the Holmes iteration.

This may look to be of only academic interest, but in fact it may be the solution to an important problem for HST images. Most HST deconvolutions are carried out using observed PSFs, because theoretically computed PSFs (using the Tiny TIM software, for example) are usually not a very good match to the observations. The noise in observed PSFs presents a problem, however: when using a noisy PSF, the restoration algorithm ought to account for the fact that both the image and the PSF are noisy. The Richardson-Lucy method (and other commonly used image restoration methods) incorrectly assume that the PSF is perfectly known. The blind deconvolution approach allows one to construct an algorithm that takes as input a noisy image and a noisy PSF observation and to construct as output estimates of both the PSF and the unblurred image.

Early experiments indicate that this method may be especially useful in crowded stellar fields where it is difficult to find isolated PSF stars. In that case one can start with a very noisy PSF star (either on the edge of the crowded region or from a separate observation) and then can improve the PSF estimate using the overlapping PSFs of stars in the crowded region (Fig. 4). The final PSF is considerably more accurate than the original noisy PSF, which in turn leads to better restorations of the unblurred image.

7. Summary

Our hope is that from these phase retrieval calculations it will be possible to derive better mirror maps for the HST/WFPC and HST/FOC optical systems. These improved maps then could be used in Tiny Tim or other optical modeling programs to compute PSFs that would be useful for a wider range of HST image restoration problems.

Blind deconvolution may allow observed PSFs to be used for deconvolution in a way that allows the uncertainty of the noisy PSF to be properly incorporated into the image restoration iteration. The ideal blind deconvolution algorithm would doubtless incorporate a phase-retrieval iteration as part of the PSF estimation (see paper by Schulz in these proceedings for such an approach.)

References

Burrows, C. J. 1990, *HST Optical Telescope Assembly Handbook*, Space Telescope Science Institute, Baltimore

Burrows, C. J. et al. 1991, ApJ, 369, L21

Fienup, J. R., Marron, J. C., Schulz, T. J., & Seldin, J. H. 1993, *Applied Optics*

Fienup, J. R., & Wackerman, C. C. 1986, J. Opt. Soc. Am. A, 3, 1897

Holmes, T. J. 1992, J. Opt. Soc. Am. A, 9, 1052

Krist, J. 1993, *The Tiny Tim User's Manual*, Space Telescope Science Institute, Baltimore

Redding, D., Lee, M., & Sirlin, S. 1994, this volume

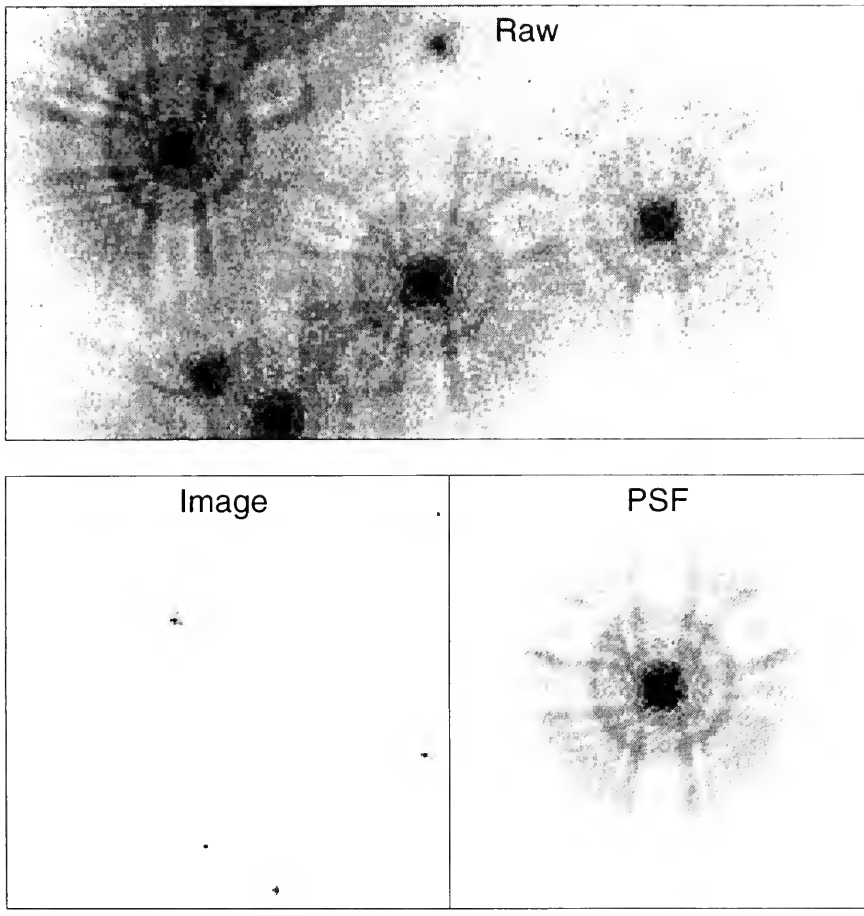


Figure 4. Demonstration of blind deconvolution on simulated PC image. Bottom panel shows raw image (wrap-around boundary conditions were used for convenience.) Star on right was used as PSF star by setting right half of initial guess to zero except at the pixel where the star is located. Three of the five stars on the left are brighter than the PSF star. Top panels show image and PSF computed using blind iteration. Note that the computed PSF has considerably higher signal-to-noise than the PSF star, indicating that the algorithm has managed to use the stellar images of the brighter stars to improve the PSF estimate.

The Restoration of HST Images and Spectra II
Space Telescope Science Institute, 1994
R. J. Hanisch and R. L. White, eds.

Simultaneous Phase Retrieval and Deblurring for the Hubble Space Telescope

Timothy J. Schulz and Stephen C. Cain

Michigan Technological University, Houghton, MI 49931–1295

Abstract. Most methods proposed for restoring images acquired by the Hubble Space Telescope rely on prior knowledge of the telescope's point-spread function; however, for many images, this function is not known precisely and must be inferred from the noisy measured data. In this paper, we address this problem and discuss a maximum-likelihood estimation technique for simultaneously determining the nature of the aberrations and for recovering the underlying object from a noisy, degraded image.

1. Introduction

Various methods have been proposed to restore images acquired by the Hubble Space Telescope (HST). Many of these techniques rely on prior knowledge of the telescope's point-spread function (PSF). However, as noted by R. White (White 1993),

> Most HST deconvolutions are carried out using observed PSFs, because theoretically computed PSFs (using Tiny TIM software, for example) are usually not a very good match to the observations. The noise in observed PSFs presents a problem, however: when using a noisy PSF, the restoration algorithm ought to account for the fact that both the image and the PSF are noisy.

This problem — the simultaneous estimation of both the object and the PSF — has been referred to as *blind deconvolution*. Unfortunately, the naive application of a blind deconvolution procedure to HST data will surely lead to difficulties since the trivial solution in which the PSF is estimated as a point or impulse and the object is estimated as the data will produce a perfect match to the data. Therefore, some type of constraints must be used to avoid both this solution and the equally trivial one in which the object is estimated as a point source and the PSF is estimated as the data.

For a related problem in ground-based astronomy, Schulz (1993) has recently proposed a technique for processing a sequence of images degraded by turbulence-induced phase errors. With this technique, the trivial solutions mentioned in the previous paragraph are avoided by using the prior knowledge that the point-spread functions are determined by phase errors distributed across the telescope's aperture. In this paper, this approach is used to derive an image restoration technique applicable to data acquired by the HST.

2. Data Model

We describe the CCD image data collected by the HST by using the model discussed in detail by Snyder, Hammoud, and White (1993) in which the data collected at pixel y of the CCD detector, $d(y)$, is described as

$$d(y) = n_{\text{obj}}(y) + n_0(y) + g(y), \quad y \in \mathcal{Y}\,, \tag{1}$$

where $n_{\text{obj}}(y)$ is the number of object-dependent photo-electrons, $n_0(y)$ is the number of background-dependent photo-electrons, $g(y)$ is CCD read-out noise, y is a two-dimensional index, and $\mathcal{Y}$ is the support region for the detector array. We assume that the random variables $n_{\text{obj}}(y)$, $n_0(y)$, and $g(y)$

are statistically independent of each other and of $n_{\text{obj}}(y')$, $n_0(y')$, and $g(y')$, for $y \neq y'$, and that the object-dependent photoelectrons $\{n_{\text{obj}}(y),\ y \in \mathcal{Y}\}$ are Poisson distributed with the mean function

$$i(y) = \beta(y) \sum_{x \in \mathcal{X}} h(y, x; \alpha) o(x), \quad y \in \mathcal{Y}\ , \tag{2}$$

where $\{o(x),\ x \in \mathcal{X}\}$ is the object's intensity function, $\{\beta(y),\ y \in \mathcal{Y}\}$ is the flat-field response function, $\{h(y, x; \alpha),\ y, x \in \mathcal{Y} \times \mathcal{X}\}$ is the PSF, and α is a collection of parameters that are determined by the optical path-length error (Goodman 1968) due to the spherical aberration, focusing error, or other aberrations. Additionally, we assume that the background-dependent photoelectrons are Poisson distributed with a known mean function $\{i_0(y),\ y \in \mathcal{Y}\}$, and that the CCD readout noise $\{g(y),\ y \in \mathcal{Y}\}$ are identically-distributed Gaussian random-variables with known mean m and variance σ^2.

Because of the parameterization by α, the PSF cannot take an arbitrary functional form, and the trivial solutions previously discussed are avoided. In the monochromatic, space invariant situation, the parameterized PSF is obtained by spatially sampling the continuous function

$$h_c(v, z; \alpha) = \left| \int A(u) e^{\jmath \frac{2\pi}{\lambda} W(u;\alpha)} e^{\jmath \frac{2\pi}{\lambda} u \cdot (v - z)} du \right|^2\ , \tag{3}$$

where $A(u)$ is the telescope's pupil function, λ is the wavelength of the observed light, $W(u; \alpha)$ is the optical path-length error due to the optical aberrations, and u is a two-dimensional spatial variable that indexes over the effective telescope pupil. The path-length error is commonly specified through either a point-by-point description

$$W(u; \alpha) = \alpha(u)\ , \tag{4}$$

or a polynomial expansion

$$W(u; \alpha) = \sum_{n=1}^{N} \alpha_n z_n(u)\ , \tag{5}$$

where the polynomials $\{z_n(u)\}$ are typically chosen as the Zernike polynomials, orthogonal over an annular region (Burrows 1990). For a system such as the HST, the discrete PSF can be created by averaging the continuous PSF over detector regions. Mathematically, this procedure is described as

$$h(x, y; \alpha) = \int \int w(x, y, v, z) h_c(v, z; \alpha) dv dz\ , \tag{6}$$

where $w(x, y, v, z)$ is a jointly discrete-continuous function that describes the averaging and sampling performed by the CCD detector array.

3. Problem Statement

Following Snyder, Hammoud, and White (1993), we simplify the noise model by defining the modified data

$$\tilde{d}(y) = d(y) + \sigma^2 - m\ , \tag{7}$$

and by considering the situation in which the CCD read-out noise variance is large. In this case, the modified data are approximately equal in distribution to a Poisson process whose mean function is

$$\tilde{i}(y) = i(y) + i_0(y) + \sigma^2\ . \tag{8}$$

Therefore, the pseudo-log-likelihood takes the form

$$\mathcal{L}(\alpha, o) = -\sum_{y \in \mathcal{Y}} \tilde{i}(y) + \sum_{y \in \mathcal{Y}} \tilde{d}(y) \ln \tilde{i}(y)\ , \tag{9}$$

where terms that do not depend on the unknown object o or parameters α have been omitted. The maximum likelihood estimator of these parameters will maximize $\mathcal{L}$ for a particular data set $\{\tilde{d}(y),\ y \in \mathcal{Y}\}$, subject to the physical constraint that the object estimate be a nonnegative function. An explicit, closed form solution to this problem is not known. Therefore, in the next section we propose a technique for producing estimates numerically by using the expectation-maximization (EM) algorithm (Dempster et al. 1977).

4. Solution

When α is known, the PSF is known and the numerical procedure is the modification of the Richardson-Lucy iterations (Richardson 1972, Lucy 1974) described by Snyder (1990) and Snyder, Hammoud, and White (1993):

$$o^{\text{new}}(x) = \frac{1}{\bar{\beta}(x)} o^{\text{old}}(x) \sum_{y \in \mathcal{Y}} h(y, x; \alpha) \frac{\beta(y)\tilde{d}(y)}{\tilde{i}^{\text{old}}(y)} , \tag{10}$$

where

$$\bar{\beta}(x) = \sum_{y \in \mathcal{Y}} \beta(y) h(y, x; \alpha) ,$$

and

$$\tilde{i}^{\text{old}}(y) = \beta(y) \sum_{x \in \mathcal{X}} h(y, x; \alpha) o^{\text{old}}(x) + i_0(y) + \sigma^2 .$$

When α is unknown, the problem is more difficult. For our initial investigation, we consider the simplified situation for which the flat-field response function is a constant $\beta(y) = \beta$, the PSF is spatially invariant, and the light is monochromatic. For this case, the EM algorithm for updating both the object and the aberration parameters becomes,

$$o^{\text{new}}(x) = \frac{1}{H_0} o^{\text{old}}(x) \sum_{y \in \mathcal{Y}} h(y - x; \alpha^{\text{old}}) \frac{\tilde{d}(y)}{\tilde{i}^{\text{old}}(y)} , \tag{11}$$

and

$$\alpha^{\text{new}} = \arg\max_{\alpha} \left[\sum_{x \in \mathcal{X}} s(x) \ln h(x; \alpha) \right] , \tag{12}$$

where

$$H_0 = \sum_{y} h(y; \alpha) , \tag{13}$$

and where

$$s(x) = h(x; \alpha^{\text{old}}) \sum_{y \in \mathcal{Y}} o^{\text{old}}(y - x) \frac{\tilde{d}(y)}{\tilde{i}^{\text{old}}(y)} . \tag{14}$$

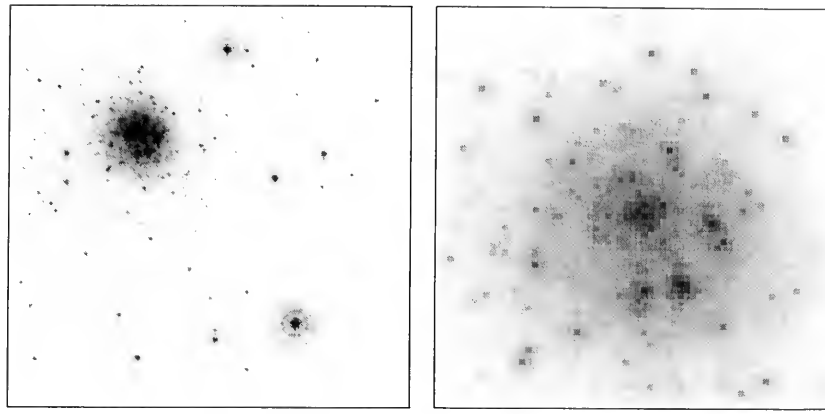
Notice that H_0 is independent of the aberration parameters α. The optimization problem defined in Eq. 12 can be performed by using standard optimization techniques; however, the maximum need not be found. Instead, if α^{new} is chosen such that

$$\sum_{x \in \mathcal{X}} s(x) \ln h(x; \alpha^{\text{new}}) \geq \sum_{x \in \mathcal{X}} s(x) \ln h(x; \alpha^{\text{old}}) , \tag{15}$$

then the iterations will be a *generalized* EM algorithm (Dempster et al. 1977) and still have the desirable property

$$\mathcal{L}(\alpha^{\text{new}}, o^{\text{new}}) \geq \mathcal{L}(\alpha^{\text{old}}, o^{\text{old}}) . \tag{16}$$

Eq. 15 can be satisfied, for instance, by performing a few iterations of any numerical optimization procedure.



(a) (b)

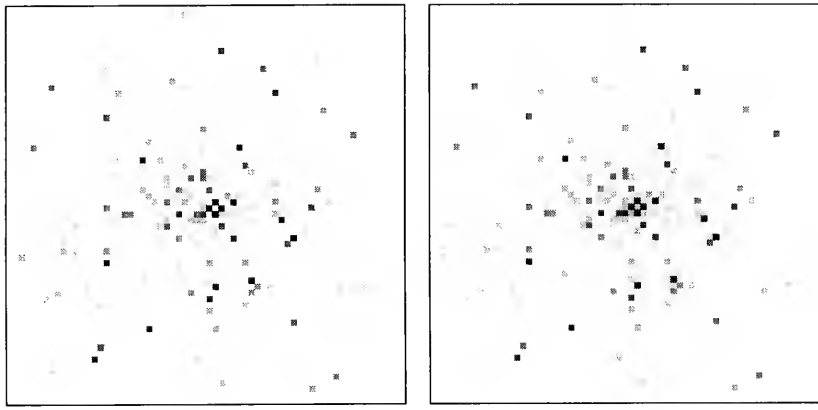
Figure 1. Simulated data used in the example. (a) 256×256 array of data. (b) 64×64 subarray of data taken from crowded region in the upper left of (a). The data in (b) are spatially magnified by a factor of 4.

5. Example

To demonstrate the use of this technique, we performed a restoration for the simulated star cluster provided by the Space Telescope Science Institute. The simulated image is shown in Fig. 1. The crowded region of the true object and of the restored object are shown in Fig. 2. The original point spread function (PSF), a noisy estimate of the PSF taken from the lower right region of the data, and the PSF estimate are shown in Fig. 3. The PSF estimate was obtained by estimating the point-by-point phase errors due to the telescope's aberrations and using these errors to create a predicted PSF.

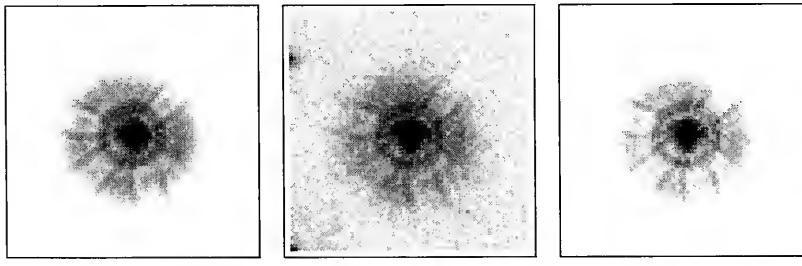
6. Conclusions

In this paper we have discussed a technique for the simultaneous estimation of the unknown object and system aberration-parameters from data acquired by the Hubble Space Telescope. We have presented an image-restoration algorithm very similar to the modified Richardson-Lucy technique, but generalized to allow for the estimation of the unknown system parameters, and, as such, applicable to the blind-deconvolution problem. In contrast with other approaches in which the spatial values of the point-spread function are estimated, with our approach the aberration parameters are estimated. This makes the technique directly applicable to situations in which the aberration-induced point-spread function is spatially varying. Although the computational burden will surely increase, the extension of the model to handle this situation is straightforward. The use of this technique on simulated data has produced promising results; however, the technique should be tested on real data, and its use with spatially varying, non-monochromatic point-spread functions should be investigated.



(a) (b)

Figure 2. (a) Original object. (b) Object estimate after 2000 iterations.



(a) (b) (c)

Figure 3. (a) True PSF. (b) Noisy PSF taken from the measured data. (c) PSF obtained from the estimated optical path-length error.

References

Burrows, C. 1990, *Hubble Space Telescope Optical Telescope Assembly Handbook*, Space Telescope Science Institute, Baltimore

Dempster, A. D., Laird, N. M., & Rubin, D. B. 1977, J. R. Stat. Soc. B, 39, 1

Lucy, L. B. 1974, AJ, 79, 745

Richardson, W. H. 1972, J. Opt. Soc. Am., 62, 55

Schulz, T. J. 1993, J. Opt. Soc. Am. A, 10, 1064

Snyder, D. L. 1990, in *The Restoration of HST Images and Spectra*, R. L. White & R. J. Allen, eds., Space Telescope Science Institute, Baltimore, 56

Snyder, D. L., Hammoud, A. M., & White, R. L. 1993, J. Opt. Soc. Am. A, 10, 1014

White, R. L. 1993, in *Image Restoration Newsletter*, R. Hanisch, ed., Space Telescope Science Institute, Baltimore, 1, 11

The Restoration of HST Images and Spectra II
Space Telescope Science Institute, 1994
R. J. Hanisch and R. L. White, eds.

Blind Deconvolution of HST Simulated Data

Julian C. Christou[1]
Starfire Optical Range, Phillips Laboratory, PL/LIG, Kirtland AFB, NM 87117

Stuart M. Jefferies[2]
Bartol Research Institute, University of Delaware, Newark, DE 19716

Mark W. Robison[3]
School of Physics & Astronomy, University of Minnesota, Minneapolis, MN 55455

Abstract. We apply an iterative deconvolution algorithm, which has the capability to recover both the object and point spread function from a single image or multiple images, to simulated HST star cluster data. The algorithm uses error metric minimization to enforce known physical constraints on both the reconstructed object and point spread function. The reconstructed object is shown to preserve the photometry inherent in the observed image. The use of multiple observations improves the signal-to-noise ratio of the reconstructed object.

1. Introduction

An iterative deconvolution algorithm based on the method of error metric minimization has been developed (Jefferies & Christou, 1993). This is a blind deconvolution algorithm in that it restores both the object and point spread function (PSF) from an observed image. A detailed discussion of the algorithm, which is based on the work of Lane (1992), has been previously presented by Jefferies & Christou (1993). Here we only highlight the main points.

In order determine both the object and PSF from a noisy measurement of their convolution, it is necessary to enforce known physical constraints. The basic constraints are that the restored object and PSF at each iteration should: (1) convolve to the observed image, (2) have real, non-negative values, and (3) exist over finite regions of space. In addition, other constraints can also be applied. These include: (1) a band-limit constraint which prevents the trivial solution of the observed image and a delta function, (2) the use of multiple observations, (3) the use of only the high signal-to-noise region of the Fourier domain, and (4) the use of object Fourier spectrum information, e.g., as extracted from speckle imaging. We note that the multiple observations constraint can be handled in two ways, either sequentially or in parallel. The former uses the reconstructed object from one observation as the input for the next observation. The latter, which we utilize, constrains the object to all observations simultaneously.

The algorithm is deterministic unlike maximum likelihood techniques which explicitly model the noise in the observations. However non-systematic noise is handled by the use of multiple

[1] Steward Observatory, University of Arizona, Tucson, AZ 85721

[2] Visiting Astronomer, National Solar Observatory, Tucson, AZ 85726. NSO is operated by AURA, Inc. under contract to the National Science Foundation.

[3] Summer Research Student, National Solar Observatory, Tucson, AZ 85726.

observations. Here the noise and the PSF are different for each image and it is only the object that is common to all the images.

Constraining the reconstruction to initial estimates of the object's Fourier components, i.e., the amplitude and phase, also prove to be very powerful especially for application to speckle interferometric data. The Fourier components, obtained from an ensemble average power spectrum for the amplitude and cross-spectra or bispectra for the phases, have significantly improved SNR over individual specklegrams, by a factor of the square root of the number of specklegrams and is typically of the order of 30. When using the Fourier amplitude as a constraint the algorithm now approximates one of "phase retrieval." When using both amplitude and phase estimates it approximates an iterative deconvolution algorithm. We note that use of either of these constraints is not limited to the speckle imaging application but can also be applied to single images if an estimate of the PSF is known. Of course, in this case, the Fourier component SNR will be as noisy as the data and appropriate weighting has to be applied. Typically, we apply the Fourier domain constraints for the first few iterations and then relax them so as to avoid systematic errors in the Fourier quotients due to differences in the object and reference PSFs.

2. Application to Simulated Data

The algorithm has been successfully applied to both real and simulated astronomical data (Jefferies & Christou, 1993; Jefferies 1993) and its ability to accurately preserve photometry in the observed image has been demonstrated. In this section we quantify this photometric linearity by application to the simulated wide field camera star cluster data made available through ST ScI (Hanisch, 1993).

2.1. Noiseless Case

In order to investigate the algorithm's performance without the influence of detector and photon noise we took the "truth" data set and convolved it with the given PSF to generate a noiseless image. We then reduced this using the Fourier modulus constraint and extracted the true object distribution and PSF. The restored object showed no residuals. This demonstrated the ability of the algorithm to extract both object and PSF and also the linearity of the restored object.

2.2. Single Object Input

The algorithm was applied to the "sim1" data. In the blind-deconvolution mode the algorithm does not recover the true PSF with any fidelity. However, extra information is contained within the Fourier modulus which is obtained from the simple quotient of the Fourier components of the object and the given PSF. Thus we used the algorithm in its phase retrieval mode. The initial object and PSF estimates were the observed objects and PSF respectively. Fig. 1 shows the convolution image, "sim1", and Fig. 2 shows the reconstructed object and PSF using the Fourier modulus constraint. Note the reduction of the extended "halos" in the object sources which are now "point-sources."

2.3. Two Object Input

For the next experiment we applied the algorithm in its blind deconvolution mode to the images "sim1" and "sim2" simultaneously. These two images have different noise realizations and a slightly different PSF. We used one of the convolution images as the initial object estimate and the brightest well isolated star, to the image top right, as the PSF estimates. The restored object and the two restored PSF's are shown in Fig. 3. Note that the PSFs show strong Airy rings as well as the filaments but have not converged to the measured PSF result. However, the object shows single pixel sources.

2.4. Four Object Input

For this experiment we increased the number of multiple images from two to four using the previous two images along with two images of the simulated star cluster obtained with WFPC 2, i.e. "sim1c"

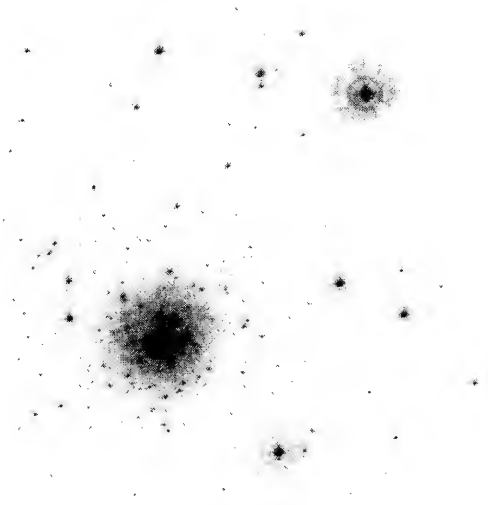


Figure 1. The simulated star cluster image "sim1".

Figure 2. Restored object (left) and restored PSF (right). Note that the PSF is displayed with a factor of 2 magnification.

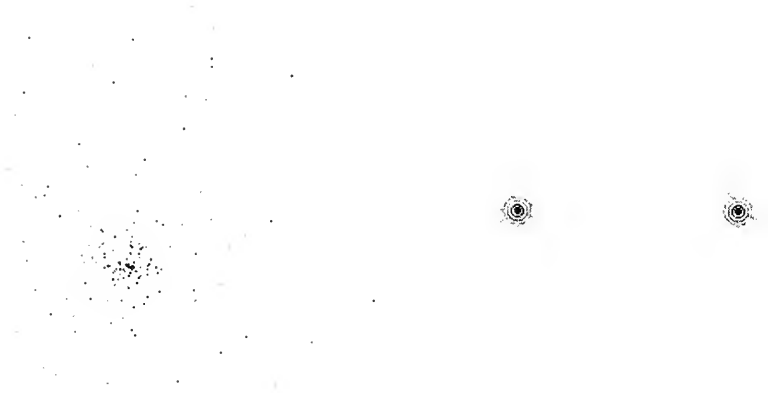


Figure 3. Restored object (left) and restored PSF's (right) for "sim1" and "sim2" respectively.

and "sim2c". We thus had four independent observations of the same field, each with a different PSF and noise realizations. As for the previous case, the algorithm was run as a blind deconvolution using one of the WFPC 2 images as the initial object estimate and the brightest star in each image as the corresponding PSF estimates. Due to the sizes of images, each of them 256×256, there were in excess of 300,000 variables to minimize. The nature of the problem forced us to reduce these data on the San Diego Supercomputer Center Cray YMP/2. The restored object and PSFs are shown in Fig. 4. Note the noise in the restored PSFs. The algorithm has found a solution which places the noise into the four different PSFs as it is not systematic and is therefore not common to the object. These PSFs are noisier than those restored for the two previous cases.

3. Linearity

An important question about any deconvolution algorithm is its linearity. In order to quantify this we followed the approach used by Busko (1993). This involves investigation of the photometric residuals as a function of the source brightness. Because of the crowded field of these simulations we only used the value of the restored object at the pixel locations given by the truth list. We defined the brightest source in the field, i.e., the bright star in the upper right quadrant, as having zero magnitude and normalized the photometry of the reconstructions to this source.

Fig. 5 shows the peak photometry for the "raw" data sets. Note the difference before and after the correction for the spherical aberration. There is still some power distributed into the PSF "halos" as is evidenced by the mean offset from zero. The scatter of the residuals is due to the crowded field where there are overlapping PSF's which contribute to increasing the intensities. Fig. 6 shows similar plots but for the three restored images shown in the previous section. Note that the three have means close to zero and that the dispersion of the residuals is substantially less than that for the "raw" images. The dispersion of the residuals increases with fainter magnitude. This behavior is to be expected as the SNR of the individual sources decreases with increasing magnitude. It can also be seen that as more input information is used, i.e., an increased number of observations, then the residual dispersion decreases showing a trend, based on the limited data presented here, which is proportional to the square root of the number of input images.

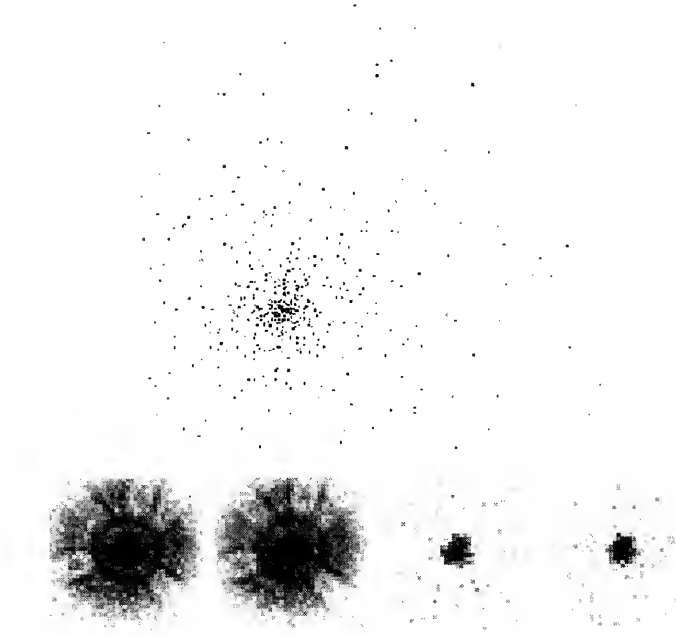


Figure 4. Restored object (top) and restored PSF's (bottom) for (left to right) "sim1", "sim2", "sim1c", and "sim2c".

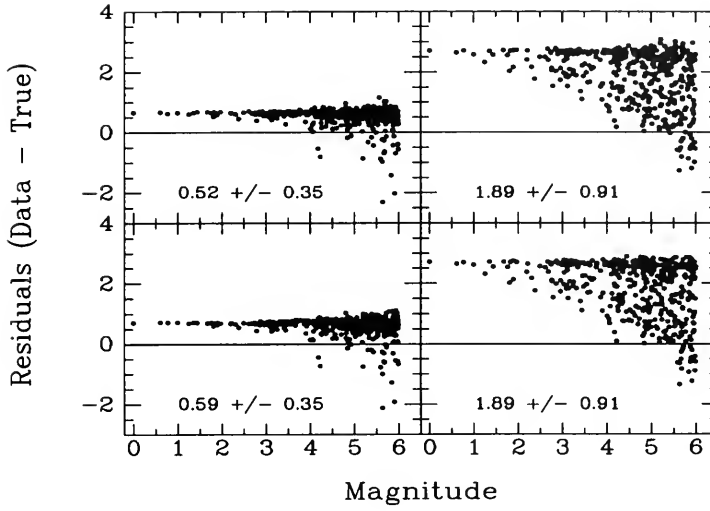


Figure 5. Linearity measurements for the "raw" images. Clockwise from top left to bottom left: "sim1c", "sim1", "sim2", & "sim2c". The means and dispersions of the residuals are shown.

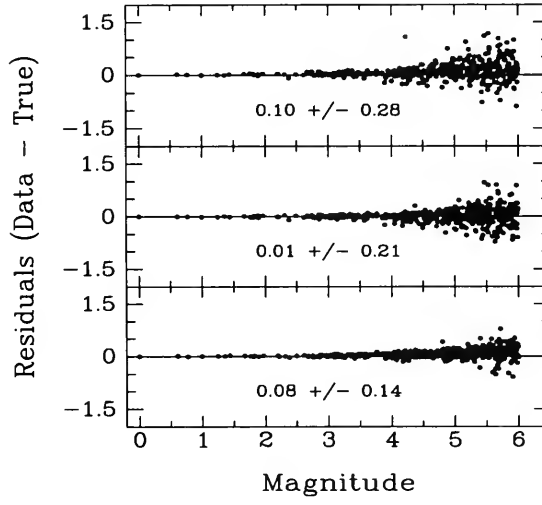


Figure 6. Linearity measurements for the three object reconstructions. From top-to-bottom, a single image input, two image input, and four image input. The means and dispersions of the residuals are shown.

Acknowledgments. SMJ was supported by NSF Grant OPP-9219515. MWR was supported by an REU scholarship from the National Science Foundation.

References

Busko, I. 1993, in *Restoration Newsletter*, R. J. Hanisch, ed., Space Telescope Science Institute, 1, 60

Hanisch, R. J. 1993, in *Restoration Newsletter*, R. J. Hanisch, ed., Space Telescope Science Institute, 1, 76

Jefferies, S.M. 1993, in *Eighth Cambridge Workshop on Cool Stars, Stellar Systems, and the Sun*, J.-P. Cauillaut, ed., ASP Conference Series, in press

Jefferies, S. M., & Christou, J.C. 1993, ApJ, 415, 862

Lane, R. G. 1992, J. Opt. Soc. Am. A, 9, 1508

Practical Problems

The Restoration of HST Images and Spectra II
Space Telescope Science Institute, 1994
R. J. Hanisch and R. L. White, eds.

Some Problems of Practical Image Restoration

Ivan R. King

Astronomy Department, University of California, Berkeley, CA 94720

Abstract. This discussion begins by noting that stellar photometry is often done quite effectively on unrestored images. Even when restoration is necessary, one must be wary of sophisticated methods, many of which distort photometric values. Fourier methods are photometrically reliable, and Wiener filtering leads to reasonably good restorations. The statistics of pixels in a restoration presents new problems, which have been only partly solved. A completely unsolved problem is presented: estimating the arithmetic difference between two images that have different PSFs and different S/N. The discussion concludes with a plea for methods that are available and transparent to the ordinary user of HST.

This is likely to be the least technical and most naïve of all the papers presented at this meeting. I come before you not as an expert but as a user, and not as a contributor but as a consumer. I could suggest that I am on the invited program only because I can't convince Bob Hanisch that I don't know anything about image restoration, but that would not be totally fair. I think that Bob really did want you to hear from some people who represent the user community of image restoration. I do have a certain amount of experience with it, and I have some scars to show for it.

1. To Restore or Not to Restore?

The first question I want to ask is, "Why restore images at all?" The reason for asking such an apparently ridiculous question is to emphasize that there are times when image restoration is not the magic potion that is needed. Most of the work that has gone on in my own group involves stellar photometry, and we have found that this is better done on the unrestored image than on any restoration. Specifically, we use PSF fitting by DAOPHOT. The trick is to have a good enough (and big enough) PSF — but discussion of that belongs to a different paper, at a different meeting.

2. Problems of Photometry on Restored Images

Photometry of restored images has two problems. First, many restorations are photometrically unfaithful, and you get the wrong answer. Second, in a restoration the pixels are no longer statistically independent, and this makes a mess of processes like least-squares fitting. It is important to note this characteristic. You don't want to use DAOPHOT on a restored image, because DAOPHOT chooses the weights of the pixels on the assumption that they are statistically independent. (Note also that the variances are no longer Poisson, which is what DAOPHOT assumes.)

What is needed instead is a least-squares method that works on correlated data points. I have actually worked out how to do least squares correctly on a restoration, but I haven't carried it all the way through to a programmable algorithm, because I felt I didn't have any real use for the result. Briefly, the normal equations look, in a formal way, exactly the same as the normal equations for least squares on independent data points, except for one horrible complication: instead of weights of individual pixels you have the covariance matrix of the pixels. One certainly doesn't want, for a 512×512 image, to generate a covariance matrix whose size is 512^4, or 68 billion elements. Undoubtedly there is a way of dealing with small parts of the image — say, the surroundings of a single star; but as I say, I haven't wanted to pursue this further.

I have taken a dim view of photometry on restored images, but I do hope that by the end of this meeting I will have heard about restoration methods that don't distort the photometry, and even methods that do better photometry on the restored image than you can do on the original. In that case I think that it will be valuable to resuscitate the study that I have just mentioned and to bring it to a practical fruition.

And of course there are situations where it is absolutely necessary to restore the image. There is no way of doing correct photometry on an extended object without restoration. But again, we have to be very sure that the restoration process doesn't distort the photometric values. And that often happens. My former postdoc, Adam Stanford, made up a test image that is closely similar to the center of M31, with its double peak. He then convolved it with a PSF, and added noise. We ran the *lucy* task that is in STSDAS, for the usual 80 iterations, and it gave the wrong answer; the ratio of the heights of the two peaks was wrong by about 10 or 20%. We sent the test to one of the practitioners of maximum entropy, and he failed it. On the other hand, Tod Lauer has told me that he has tested "lucy" and found it to reproduce a known answer correctly. Maybe Tod's test wasn't severe enough; maybe we did something wrong in our test. I would certainly like to know what the answer is.

Also, we have run "lucy" on a star field and then done photometry. The faint stars were systematically wrong. I have been told by at least one expert that what you need to do is to run "lucy" to complete convergence, but we haven't been able to do that — at least with the *lucy* task in STSDAS; the field begins to fill up with false stars that come from noise peaks. But note the distinction: I am not talking about the Richardson-Lucy algorithm in general; I am only telling what happened when I used the version in STSDAS. Maybe others can do it better. As I said, I'm not an expert.

Nor do I wish to detract from all the excellent restoration methods that are available, but just to say, "buyer beware." If you just turn a crank, there is a real danger that all you'll get is sausage.

3. Problems of Fourier Restorations

As a result of not knowing how to do any better, I have stuck to Fourier restorations, which can be counted on to be linear. Even so, there are precautions to be taken. My associates and I have learned, the hard way, that you have to avoid discontinuities at the edges, match opposite sides and preferably taper with a cosine bell, that you have to subtract out stars that overlap the edge, and so on.

Even a method as simple-minded as Fourier has one difficulty: filtering. Here the best that I have found is the STSDAS task *wiener*, which creates a so-called "optimum filter." The task has a number of options, though; my experience has been that on HST images it gives the best results when run with the default options. (But since the time of the Workshop I have heard that others have achieved better results by using a signal model that is rather sharp.) Even so, I've gotten some results that had obvious noise artifacts and needed a further low-pass filtering. And it's annoying that the restoration of a piece cut out of an image gives a result different from the way that piece looks when you restore the entire image. (This is because of the way "wiener" goes out to the edge of the transform of whatever image it has, to estimate the noise.)

4. An Example of a Perplexing Problem

I would like to spend the rest of my time showing an example of the kind of dilemma one can get into when doing restorations. In this particular case we were dealing with an FOC ultraviolet image of the center of M31, and a PC *V*-band image of the same region, scaled, rotated, and aligned. We wanted to take the quotient of the two images, to look for color effects. But first, of course, we had to get the light back where it belongs, by doing restorations.

Notice that for this problem one cannot use the sophisticated methods that have been shown to produce such handsome restorations. The degree to which they restore depends on S/N, and the two images would get sharpened by different amounts, producing an erroneous quotient.

In doing the restorations we tried to treat the two images identically (except of course for using its own PSF for each image). What we did was to run "wiener" on the noisy FOC image, and extract the filter that was generated in the process. Then we did Fourier deconvolutions of each image by hand, using that *same* filter on each. (This left the PC image less than optimally filtered, but it should assure that each image has the identical amount of blurring.) The result showed a small, well-marked central peak that certainly represents a real ultraviolet excess at the center of M31. Surrounding this peak is a large, round, shallow depression. M31 is elongated, but this depressed "bowl" is round. The only component in the problem that is round is the PSF. Does the bowl represent a real UV feature of M31, or is it possible that the bowl is just an artifact that comes from the restoration process? We have tried using somewhat changed PSFs, but we haven't been able to make the bowl go away. We are going to keep trying such experiments, but I certainly hope that some one who is more expert than I am can make the suggestion that I have been lacking so far.

I have described a specific problem here, but instead I can pose an abstract problem that, as far as I know, has never been addressed. It goes as follows: "Images A and B show the same extended object with identical placement, but were taken under different circumstances. Each is accompanied by its own PSF. Image A has low S/N, while B has high S/N. Wanted: a good estimator of the true arithmetic difference, pixel by pixel, between A and B, properly corrected for the different blurring in the two images." This is not at all a rare situation, and I think the problem is worthy of consideration.

5. A Plea to the Learned

The foregoing is an example of the difficulties that come up when an innocent ventures into image restoration. (And incidentally, the astronomical answer in this case is awfully important.) But let me conclude by emphasizing this position of innocence and ignorance. The average user of HST images knows even less than I do. One can of course argue that a good scientist learns whatever he or she needs to know in order to get meaningful answers. But there is a limit to how many things a person can become expert in, and the more time that you spend learning abstruse methods of image reconstruction — and especially the practical details that you need to know to make them work — the less time there is to reduce and measure those voluminous images. So my real message is that you experts need not only to devise elegant and powerful methods of image restoration; you need also to make them accessible, and even transparent, to the ordinary HST observer — without the danger that he or she will unwittingly produce wrong answers without knowing it. May I remind you of the phrase "user-friendly"?

6. Conclusion

I know I haven't said much to enlighten you, but I look forward to learning a lot from the other papers that will be presented, and I particularly hope that they will be written up in a way that makes their content clear to the ordinary astronomer — and if we may hope for such a thing, that they will soon be embodied in tasks that can be made available in STSDAS.

The Restoration of HST Images and Spectra II
Space Telescope Science Institute, 1994
R. J. Hanisch and R. L. White, eds.

WFPC and WFPC 2 Instrumental Characteristics

J. Biretta

Space Telescope Science Institute, 3700 San Martin Drive, Baltimore, MD 21218

Abstract. We summarize instrumental properties and problems in the WFPC and WFPC 2 cameras which are likely to complicate image restoration.

1. Introduction

The spherical aberration in the HST primary mirror has generated considerable interest in restoration and deconvolution of WFPC images. The new WFPC 2 camera, which will be installed in HST in December 1993, should completely correct the aberration. Nonetheless, deconvolution may remain useful in removing optical artifacts such as Airy rings, diffraction spikes, and scattered light from WFPC 2 images. Herein we discuss the characteristics of both instruments, with emphasis on effects and problems which seem likely to impact restoration and deconvolution of these images. We begin with a brief summary of the layout and parameters of both instruments. We then examine problems which corrupt the PSF, effects which contribute noise to the images, and finally artifacts in the images which will complicate deconvolution.

2. Instrumental Parameters

The Wide-Field / Planetary Camera (WFPC), and its replacement camera, WFPC 2, have nearly identical optical layouts (Fig. 1). In both cases light from the Optical Telescope Assembly (OTA) is reflected off a "pickoff mirror", passes through the shutter and filter assemblies, and is then divided into four separate channels by a "pyramid" mirror which also serves as a field flattener. Light from each channel is then reflected off a flat "fold" mirror, through a Cassegrain relay system, through a MgF_2 field flattener, and finally onto an 800×800 pixel CCD detector.

An important difference between the old and new cameras is that WFPC has four Wide-Field Camera (WFC) and four Planetary Camera (PC) imaging channels. Either the WFC or PC mode can be chosen by simply rotating the pyramid mirror. Instead, WFPC 2 has only three WFC channels and a single PC channel with a fixed pyramid mirror. But perhaps the most important difference is that WFPC 2 has a corrective figure applied to the Cassegrain relay secondary mirrors, which nullifies the spherical aberration of the OTA primary mirror.

Properties of the CCD detectors for each camera are summarized in Table 1. WFPC uses thinned back-side illuminated CCDs, whereas the WFPC 2 detectors are front-illuminated. The front-illuminated devices have much better flat-field characteristics, but a price is paid in the form of lower quantum efficiency (QE) at red wavelengths. An important advantage of WFPC 2 is its lower read-out noise. The read-out noise is a factor ~ 2.6 lower than the old camera, and more than compensates for the reduced QE.

3. PSF Effects

The well-known spherical aberration in the OTA primary mirror causes most of the light for stellar targets to fall in a large halo (Figs. 2 and 3). The PSF core has nearly the correct diameter, but contains only about 15% of the light. The remaining 85% forms a complex halo about 4″ in

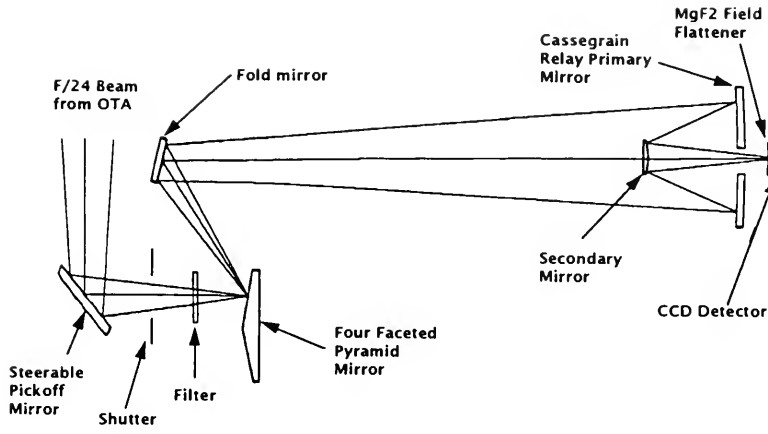


Figure 1. WFPC 2 optical configuration; the original WFPC is nearly identical. From Burrows, Clampin, and Krist (1993).

Table 1. Comparison of WFPC and WFPC 2 CCD detector characteristics.

Parameter	WFPC	WFPC 2
Device	Texas Inst. 800×800	Loral 800×800
Type	Thinned back-side illuminated	Thick front-side illuminated
Pixel size	15 μm	15 μm
Image Scale: WFC	0.″10/pixel	0.″10/pixel
PC	0.″043/pixel	0.″046/pixel
QE (6000Å)	50%	25%
Read-out noise	13 e^- RMS	5 e^- RMS

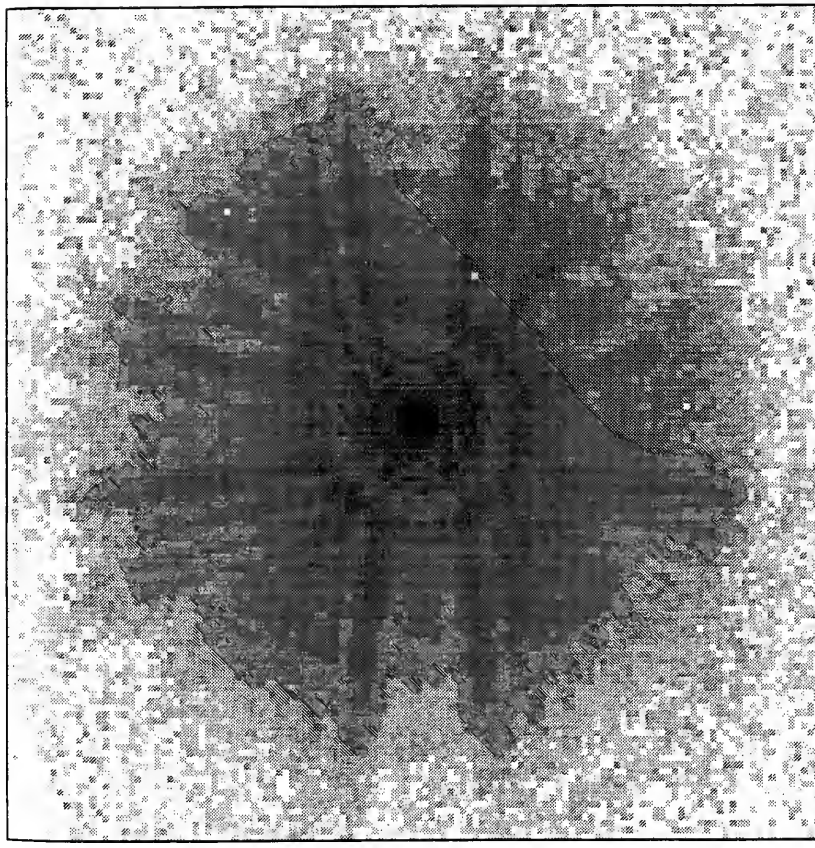


Figure 2. Typical PSF for WFPC Planetary Camera taken through a 487nm narrow band filter at the "standard" HST focus position. Gray scale is logarithmic. From Holtzman (1992).

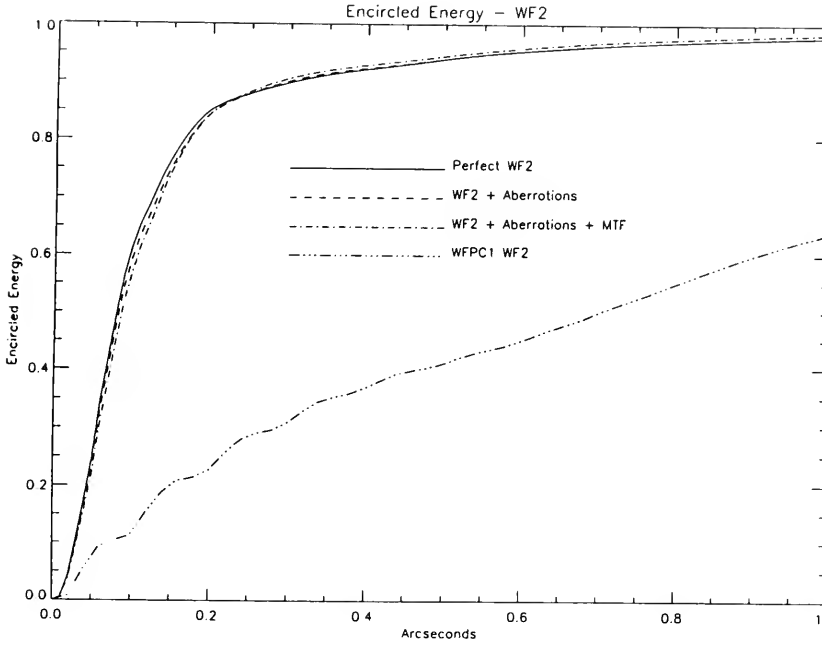


Figure 3. Encircled energy and a function of aperture radius for the Wide-Field Camera mode. The top three curves are for WFPC 2, while the bottom curve ("WFPC1 WF2") is for WFPC. Curves for Planetary Camera modes are nearly identical except at smallest radii. From Krist (1993).

diameter. The newer WFPC 2 has a corrective figure applied to the relay secondary mirrors which should completely eliminate this aberration. The correction will also reduce the PSF core diameter slightly, since light from the mirror edges falls into the PSF halo for the old camera.

Both cameras seriously undersample the PSF at most wavelengths. The PSF core radius is approximately 0.″08 at 8000 Å, but only 0.″02 at 2000 Å. The WFC pixel size is about 0.″10, and 0.″04 for the PC, so that the PSF is severely undersampled, except perhaps for PC observations in the far red.

In both the old and newer cameras, the PSF varies with both field position and time. The variations with field position are a large effect for WFPC images (Fig. 4). Due to vignetting within the relay optics, the CCD edges don't see all of the OTA primary mirror. This causes details of both the PSF core and halo to vary with field position. A notable effect is that the PSF core shape becomes elliptical near the CCD edges away from the pyramid apex. "Tendrils" and other features in the PSF halo also change shape.

The PSF also varies with time due to both OTA "breathing" and pointing "jitter." Solar heating of the OTA causes small focus variations during the orbit with typical amplitudes of about 5 μm. Occasionally variations as large as 10 to 15 μm are seen. This "breathing" of the OTA has a noticeable effect on the PSF wings for WFPC data; the wings expand and contract on a timescale comparable to the orbit period. For WFPC 2 data, the breathing only has a small effect on the PSF size.

Typical pointing jitter errors have a 3 to 5 milliarcsecond (mas) RMS amplitude in Fine Lock tracking mode. Much larger transients, exceeding 20 mas, sometimes occur at terminator crossings due to solar array oscillations. Large transients are occasionally seen at other times, as well, but their cause is not well understood. The new solar arrays installed during the HST Service Mission should reduce transients for WFPC 2 observations.

Besides these short time-scale effects, the PSF also undergoes slow variations due to long-term desorption (shrinkage) of the OTA, and then sudden changes during related focus readjustments.

Both cameras tend to distort the PSF for very short exposures ($<$ 1 second). This distortion occurs because the shutters act as a narrow traveling slit for short exposures. Any given detector pixel sees the full OTA primary mirror pupil for only a small part of the total exposure, so that the PSF is corrupted. In a related effect, light is diffracted off the edges of the traveling shutter blades, further complicating the PSF. These effects are important to remember when using observed PSF images; exposures much shorter than 1/2 second should generally be avoided.

Because WFPC 2 uses front-side illuminated CCDs, it is susceptible to various sub-pixel effects. (These effects are thought to be smaller in WFPC, and have generally gone unnoticed due to the spherical aberration.) The WFPC 2 CCDs appear to scatter light, so that only about 60% of the photons striking any given pixel are actually detected by that pixel. The remaining 40% are scattered and detected by the surrounding pixels. This has the effect of smearing out the PSF slightly, especially for the WFC detectors where the PSF is quite small compared to the pixel size. A second effect in WFPC 2 are sub-pixel quantum efficiency (QE) variations. The QE appears to vary by about 10% across the pixel faces. The exact pattern is poorly understood, but the QE is currently thought to be highest in the pixel centers. This will impact efforts at sub-pixel image restorations, as well as photometry when the PSF is severely undersampled.

4. Image Noise

The most important noise source for both the WFPC and WFPC 2 is the readout noise. This is a random noise with an approximately Gaussian distribution which occurs while reading out the CCD and measuring the charge deposited in each pixel. For WFPC the noise is about 13 electrons RMS per pixel, which is well-sampled by the 7.5 electron per DN digital conversion. For WFPC 2 the readout noise is only about 5 electrons per pixel, and is undersampled by both the 7.5 and 15 electrons per DN digital conversions. This undersampling of the noise may impact deconvolution and statistical analyses of WFPC 2 data.

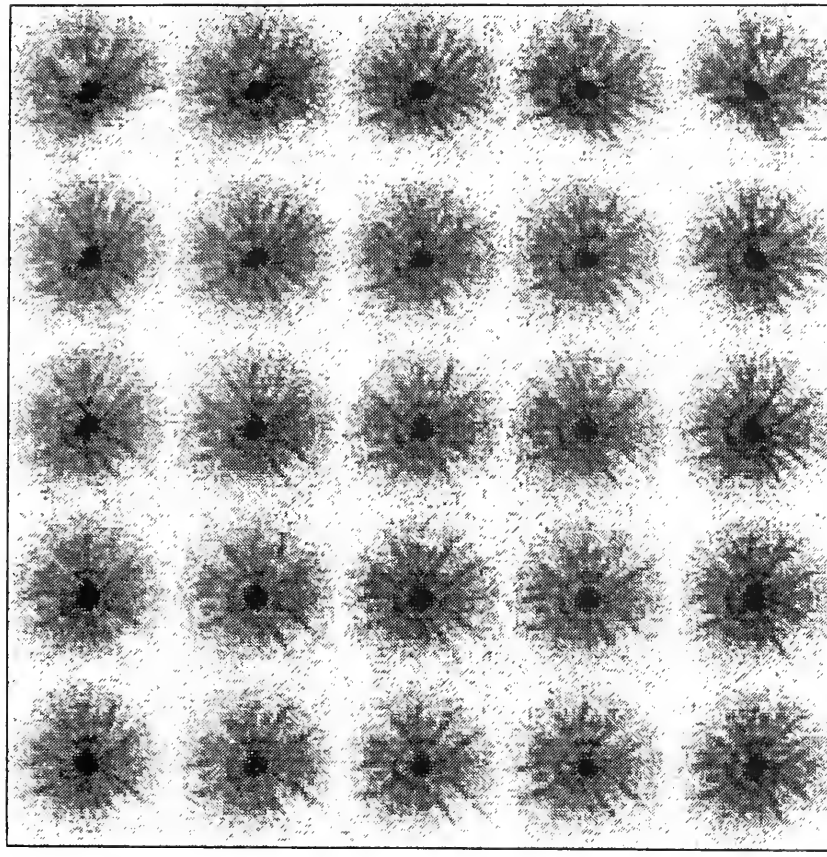


Figure 4. Sample PSFs for WFPC detector WF2 at various field locations. Location on the page is indicative of the position on the detector, with the pyramid apex (OTA optical axis) at the lower left; each PSF is magnified by a factor of three. From Holtzman (1992).

WFPC suffers from an electronic problem in its "sample and hold" circuitry which corrupts the analog-to-digital conversion of the data. Basically, as various bits in the A-to-D conversion are tested, the bits sometimes fail to set properly so that the resulting digital values are too low. This adds a small noise which is typically about 1/2 DN, but can be as large as 2 DN, depending on the pixel value. This again may impact statistical analyses on the data. WFPC 2 has a similar problem, but at about twenty times lower level.

For most targets, WFPC images must be "preflashed" with faint lamps to eliminate charge transfer errors during the readout. Images are typically preflashed to about 30 electrons, which introduces about 1 DN of noise per pixel. At present it is believed that WFPC 2 images will not require any preflash.

Table 2 summarizes the effective total noise and other factors related to faint target detection for both WFPC and WFPC 2 cameras. Fig. 5 compares observed WFPC, Richardson-Lucy deconvolved WFPC, and simulated WFPC 2 images of a field containing faint targets. While the PSF wings are reduced in the deconvolved image, noise prevents detection of faint objects easily seen by WFPC 2.

Table 2. Comparison of WFPC and WFPC 2 factors affecting faint target performance.

Parameter	WFPC	WFPC 2
Effective total noise	20 e^-	5 e^-
QE (6000Å)	50%	25%
Fraction of light in central pixel of PSF for WFC	6%	~65%

5. Image Artifacts

Pixel-to-pixel sensitivity variations of the detectors are routinely removed by flat field calibration, that is, by dividing the raw data by exposures of a uniformly illuminated target. This procedure is often imperfect, however, due to a number of effects, including time-variable contamination inside the cameras and errors in the calibration flats themselves.

For the WFPC, an important source of flat fielding error is the "persistent measles." Low volatility contaminants collect on the cold camera windows 6 mm in front of the focal plane. These particle-like contaminants cause diffractive scattering of light, resulting in patterns of concentric bright and dark circles with total scale sizes of 3 to 10 pixels. These can have adverse effects on deconvolution, since their size is close to that of the PSF core. The amplitude and number of the measles are different in different camera channels. In the WFC and most of PC6, the errors are about 1% in intensity. Larger errors of 2 to 5% are seen some of PC6, and much of the other PC CCDs. An example of measles is shown in Fig. 6. While corrections (called DELTAFLATS) are available, they are only partially successful since the measle pattern depends on the illumination pattern and spectral properties of the target. These errors tend to be larger at short wavelengths, where diffractive effects are stronger. Great care was taken to reduce contaminant levels during the manufacture of WFPC 2, so it is unlikely to suffer these problems.

Both WFPC and WFPC 2 are liable to have errors or defects in the calibration flats. This is certainly true for the current WFPC calibration flats. At small spatial scales (<10 pixels) there are often errors of a few percent, usually in the form of long, narrow streaks. In some cases these errors are as large as 10 to 15%. These streaks originate from observations of the bright earth which are used to generate the calibration flats; earth features together with spacecraft motion cause streaks

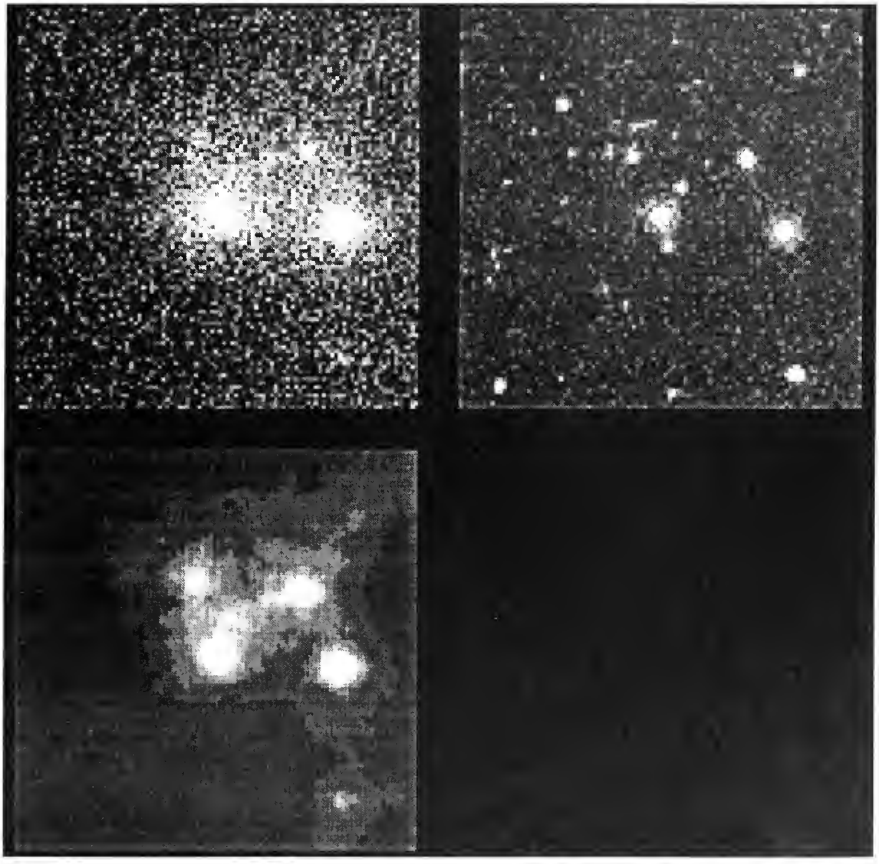


Figure 5. Comparison of observed WFPC (upper left), Richardson-Lucy deconvolved WFPC (lower left), and simulated WFPC 2 image (upper right) of a WFC field containing faint stars.

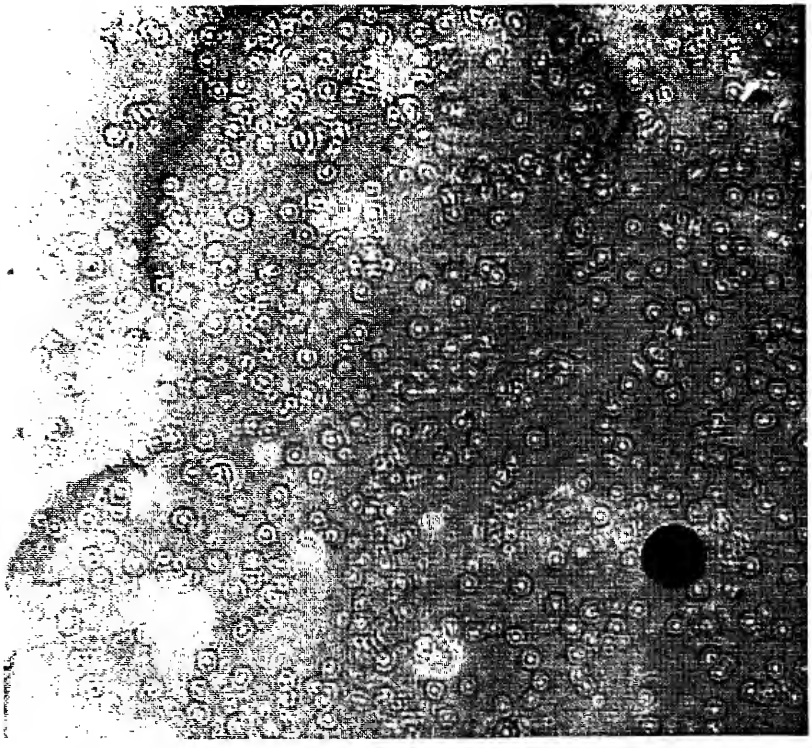


Figure 6. Example of "measles" in detector PC8 through filter F517N. The display range is $\sim \pm 7\%$ about the mean. From MacKenty et al. (1992).

in the raw images, which are not completely removed during calculation of the calibration flats. On larger scales (>100 pixels) there are sometimes errors up to about 15%; these are caused by the use of neutral density filters when observing the broad-band flats. We plan to deliver new calibration flats early in 1994, in which these errors should be greatly reduced. These problems should be unimportant for WFPC 2, due to its special internal calibration channel.

The CCDs in both the new and old cameras have manufacturing defects in the form of pixel rows or columns which are too narrow or wide. In WFPC 2 every 34th row is about 3% too narrow. The standard flat field calibration is based on the assumption that all pixels have the same size, so these anomalous rows will not be properly calibrated. In particular, stellar targets falling on these rows with standard calibration will have 3% too many counts. These rows also cause a geometric distortion for PSFs falling on them. The old WFPC camera has a similar effect, with every 33^{rd} row and column being slightly too wide.

Figs. 7 and 8 show 400×400 pixel areas from the centers of the WF2 calibration flats for WFPC and WFPC 2, respectively, and illustrate some of the features discussed above.

Cosmic rays cause artifacts in images for both cameras. These pass through the CCD detectors depositing large amounts of charge (typically >100 electrons per pixel). In WFPC most cosmic ray hits involve only 1 or 2 pixels, and it is easy to distinguish weak cosmic rays from aberrated stars. In WFPC 2, however, the cosmic rays are expected to involve more pixels per hit since the CCD is thicker. Because the cosmic ray hits will be larger, and the PSFs smaller, it is expected to be more difficult to distinguish cosmic rays from stars in WFPC 2. For both cameras, cosmic rays can be eliminated by combining several exposures of the same field with rejection of anomalously high pixels. The elimination of cosmic rays from images which are poorly aligned (*i.e.* having displacements less than 1 pixel) is an important topic requiring further study.

Hot pixels affect images in both WFPC and WFPC 2. These are single pixels whose dark current is anomalously high (>0.01 DN sec^{-1}). They cause a pattern of single pixel bright spots in long-exposure images; the pattern is more-or-less constant on time-scales of hours, but varies on time-scales of months. Each CCD has several hundred of these hot pixels, and careful calibration can significantly reduce their numbers and intensity, but cannot eliminate them completely. One solution is to take several exposures with the target on slightly different regions of the detector, so that "real" objects can be easily distinguished from these detector artifacts.

Acknowledgments. I am grateful to J. Krist and C. Burrows for helpful discussions regarding WFPC 2 characteristics and anomalies.

References

Burrows, C. J., Clampin, M., & Krist, J. 1993, *HST Wide-Field and Planetary Camera 2 Instrument Handbook*, Space Telescope Science Institute, Baltimore

Holtzman, J. 1992, in *Wide Field/Planetary Camera Final Orbital/Science Verification Report*, Ch. 9

Krist, J. 1993, private communication

MacKenty, J. W. et al. 1992, *HST Wide-Field and Planetary Camera Instrument Handbook*, Space Telescope Science Institute, Baltimore

Trauger, J. T. 1993, *WFPC 2 Science Calibration Report*

Figure 7. Example of WFPC calibration flat. This is a 400×400 pixel section from the center of detector WF2 for filter F555W. The gray scale ranges from 0.95 (black) to 1.05 (white). Some notable features are: (a) dust particles on windows, (b) large circular arc defects due to use of neutral density filter, (c) 33 column / row manufacturing errors, and (d) dust particles on the CCD itself. There is also a broad intensity gradient wherein the brightness decreases from the bottom to the top of the image; this is a defect related to the neutral density filter.

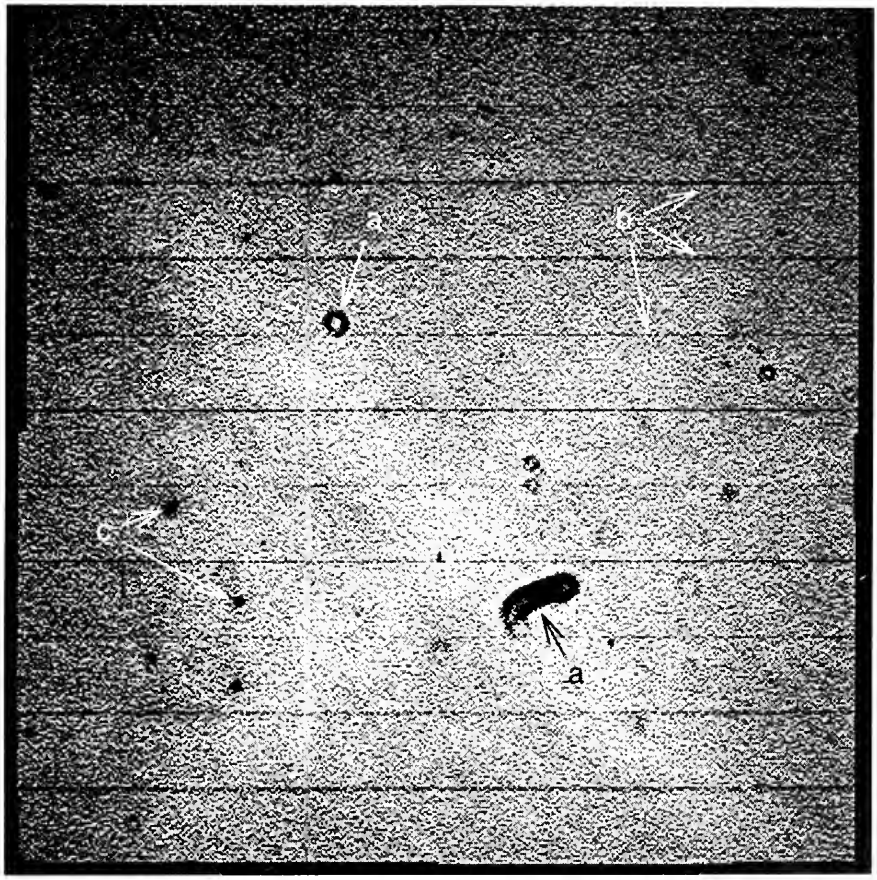


Figure 8. Example of WFPC 2 flat. This is a 400×400 pixel section from the center of detector WF2 for filter F555W. The gray scale ranges from 0.95 (black) to 1.05 (white). Data are from ground-based camera tests. Some notable features are: (a) dust particles on windows, (b) 34 row manufacturing error, and (c) dust particles on the CCD itself.

The Restoration of HST Images and Spectra II
Space Telescope Science Institute, 1994
R. J. Hanisch and R. L. White, eds.

Restoration of FOC Imaging Data: Some Considerations

Dave Baxter

Space Telescope Science Institute, 3700 San Martin Drive, Baltimore, MD 21218

Abstract. We review and discuss some of the considerations, relating to FOC imaging, which will affect or at least influence the quality of image restorations. Some of these effects derive from the nature of the instrument and the calibration procedures, however, others are the responsibility of the individuals designing the observations. We also discuss the expected status of these effects after the deployment of COSTAR.

1. Introduction

Since the discovery of spherical aberration associated with the primary mirror, the application of image restoration techniques to HST imaging data has become an almost routine procedure. It is not my intention here to argue for or against the use of these procedures generally, but simply to highlight certain features relating to the Faint Object Camera and the data it produces and how these features should influence an observer when designing observing proposals. If nothing else, this discussion may help those less experienced image restorers to understand some of the limitations of the techniques. The main sources of information for FOC Users should be the *Faint Object Camera Instrument Handbook, Version 3.0* (Pre-COSTAR) (Paresce 1992), and *Version 4.0* (Post-COSTAR) (Nota et al. 1993). Additionally, our accumulated experiences over the last three years are published in "In-Flight Performance of the Faint Object Camera of the HST: II" by Greenfield et al. (1993).

Clearly, for each individual user, the relative importance of these different effects will depend very much on the quality of their data, the type of target being observed, and the science they are trying to do, and therefore the following topics are addressed in no particular order. Readers should not attempt to infer any significance regarding their relative importance from the order in which they are presented. It should be noted that most of these effects are not independent in that their existence often influences, or is influenced by, other considerations. Note also that unless stated otherwise, all of our discussion is slanted towards the F/96 relay since this is the most used. We present a typical example in §2. In the course of each discussion we will comment on how each of the problems will be affected by the deployment of COSTAR.

2. Frame Size and Pixel Type

For many applications, a serious restriction on FOC imaging is the amount of sky coverage available. The platescale associated with the F/96 relay ($\sim$0.″0225 per pixel), in principle gives us a maximum imaging area (for a 1024×1024 image) of $\sim$23", but in reality this area is somewhat smaller because of geometric distortion. A further consideration relates to the aberrated PSF which has about 85% of its power in a halo which extends out to a radius of 100 pixels from its center. This means that a square perimeter with at least this width will be influenced, to a greater or lesser degree, by edge effects, further reducing the scientifically useful area of the image, at least as far as image restoration is concerned. The bottom line is that many attractive science targets, such as most Solar System planets, just won't fit. Even if you have determined that your target *will* fit, you should also remember that, because the scientifically useful area of any given format can be much smaller, it is almost an imperative that an interactive acquisition be included in your proposal (see later in this section).

As previously implied, successful use of the FOC usually requiresd careful planning of the exposures and filter combinations; this is especially true for those desiring to perform image restoration. These are not however the only considerations and so, as an example, let us review the factors which an observer will have to consider when deciding on an imaging format. In the first instance let us assume that the observer feels constrained, because of the target size or maybe just for maximum sky coverage, to use the largest format available, the F/96 512×1024 (zoomed) format, in which case

- Because of the large area to be scanned by the read beam, the large format has the poorest linearity characteristics, i.e., significant non-linearity is to be expected in almost all data obtained with this format since the maximum reasonable count rate allowable (for non-linearity $\leq$10%) is about 0.05 counts/pixel/sec.

- FOC pattern noise (see below), appears to be proportional in some way to the level of non-linearity present. Choosing a format with poor linearity characteristics magnifies the problem of pattern noise.

- This format only has 8-bit pixels, i.e., the maximum number of photons which can be registered by a single pixel, regardless of exposure time, is 255, at which point the pixel "wraps over" and continues counting from zero. An individual pixel can be wrapped over several times, and therefore this problem cannot be safely corrected without a priori knowledge of the intrinsic structure within the affected area.

- For point sources, the maximum reasonable count rate (i.e., peak) is somewhat higher, (about 0.3 counts/pixel/sec), however, if the proposed target is a crowded star field, high levels of non-linearity will almost certainly be present in the cores of the brightest stars, even if wrap-over is avoided.

- The 8-bit count limit, combined with the count rate considerations, implies that for all but the most diffuse, extended objects, high signal-to-noise is difficult to achieve without prohibitively long exposure times.

What are the solutions to these problems? Well, if the large format is the only one which will suffice for your purposes then there are no practical solutions except the obvious one, i.e., that you take extreme care to calculate your filter combinations and exposure times to avoid wrap-over, maybe by splitting the total exposure and obtaining multiple images, and to limit the non-linearity to acceptable levels. If the initial filter calculation included one or more Neutral Density (ND) filters, then the observer could opt to take two exposures: one with sufficient NDs to make the bright sources reasonably linear, and one with less NDs to show up the faint structure.

A second option which users should consider is the use of a smaller format. The 8-bit count limit only applies to two formats, the 512×1024 (zoomed) and the 512×512 (normal) formats, so one solution would be to use a smaller format and take multiple images with positional offsets, i.e., 2×512×512 (zoomed) images (with 16-bit pixels) instead of one 512×1024 (zoomed) image. There are other advantages to this solution in that, although your image has only half the area and all of the other problems still remain, half the area implies approximately twice the scan frequency. If the image is being scanned twice as often, the linearity characteristics are improved by about a factor of two, i.e., images are only half as non-linear. You can also expect an accompanying reduction in the pattern noise since this seems to scale with non-linearity. If you opt for the 512×512 (zoomed) format you can either

a. Obtain the same S/N with no wrap-over and twice the linearity, in the same exposure time,

b. Obtain the same S/N with no wrap-over but with the same non-linearity in half the exposure time, or

c. Obtain better S/N ($\sim \sqrt{2}$) with no wrap-over but with the same non-linearity in the same exposure time.

There are two main disadvantages involved in moving from a large format to a smaller one, particularly if image restoration is intended. First there is the reduction of the effective image area, i.e. the area free of edge effects. With the present (pre-COSTAR) PSF this is a very serious consideration since edge effects are present within a 100 pixel perimeter of every image format. The second disadvantage is that the smaller format may necessitate an acquisition image to ensure that the science target is positioned at the center of the image. Remember that the accuracy with which the telescope can be pointed depends almost totally on the accuracy of the Target and Guide Star coordinates (which can accumulate to several arcseconds of error), so when the imaging aperture is only 11" square, an observer take a great risk by not including an interactive acquisition.

Following the installation and calibration of COSTAR we have a good news/bad news situation. The bad news is that the focal ratio of the FOC cameras will be changed from F/96 to F/151 (F/48 $\rightarrow$ F/75), resulting in a reduction of the format sizes (in arcseconds) by a factor of 2/3 (F/96 plate scale is reduced to 0.″014/pixel). With these smaller format sizes, the need for an acquisition image should be even more apparent. The good news is that, if you still feel a need to restore images after COSTAR, the PSF halo should be fairly negligible (only about 20% of the total light), and therefore edge effects should only be present in a relatively narrow perimeter, leaving a larger effective science imaging area (relative to pre-COSTAR). In fact, the absolute size of the effective imaging area (in arcseconds) is not going to be very much smaller than the pre-COSTAR situation. It should also be noted that the smaller size of the pixels give a slightly better sampling of the PSF. The 8-bit counting limit will, of course, be unaffected, and so wrap-over will still occur, which in turn will force shorter exposure times because of the 4–5 fold increase in the peak intensity of the PSF.

The effects of COSTAR on non-linearity and distortion, etc., are discussed in the relevent sections.

3. Detector Blemishes

In Fig. 1 we show an example of large format (512×1024 zoomed) internal flatfield obtained with the F/96 relay of the FOC. The image has had an unsharp masking technique applied in order to make the various defects and blemishes more obvious. As a first step in this discussion, I will try to identify some of the more important features beginning with those which are there for a purpose.

- The first deliberate feature is the grid of *reseau* marks. This grid comprises 17 rows and 17 columns with a nominal spacing (prior to detector distortion) of 60 pixels. The reseau marks are etched onto the first of the bialkali photocathodes in the intensifier tube from where they cast shadows onto the TV target. This grid is used to determine the geometric distortion of the detector. Two things should be noted, firstly the grid is rotated by about 7° with respect to the frame axes, and secondly the grid is off center by about 70 pixels in the $-y$ direction, resulting in the loss of one complete row of the grid, and an absence of reseaux across the top of the image. The significance of this is discussed in the section on geometric distortion.

- The second deliberate feature visible in F/96 external images (but not internal flat fields) are the occulting fingers. The occulting fingers, which cast very clear shadows on the full format science data, project into the images from the top right and bottom left edges of the frame.

Several obvious artifacts and blemishes are also visible, most notably:

- Vignetting by a square baffle located just upstream of the detectors causes the large wedge-shaped shadow at the top left and the smaller vertical wedge at the bottom left.

- The central 512×512 format causes the square "burn in" feature seen at the center of the image.

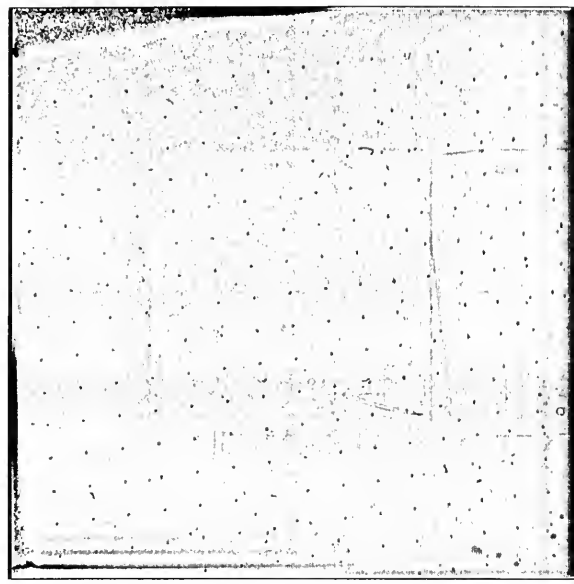


Figure 1. An F/96, 512×1024 (zoomed) internal flatfield. Reseau marks and blemishes have been highlighted using an unsharp masking technique.

- Vertical striping can be seen at the right side of the image. This is referred to as "scan distortion" (see §4).
- The horizontal "ripple" seen at the bottom left, is thought to be due to a ripple instability of the coil drivers when the frame scan begins.

Other, less obvious features include a narrow line (generally referred to as the "flyback"), which runs from the bottom left to the upper right corners due to the read beam not being completely blanked when it returns to the beginning at the end of each frame scan. This is more noticeable in the smaller formats. The narrow horizontal features at the right edge, (y = 256, 512, and 768) are due to noise glitches in the scan coil driver, caused by changes in the most significant bits of the line counter. Finally there are several patterns present; these are discussed in §7.

For any given wavelength, a single full format (1024×1024) flatfield is available and smaller formats use the appropriate subset of this image. Since the full format flatfields are themselves heavily smoothed (using a Gaussian filter with $\sigma \simeq 15$ pixels), the majority of these blemishes are not removed by the flatfielding process. Also, some of the features such as the scan distortion and the rippling at the bottom left, involve changes in the geometric distortion, not just sensitivity.

It has proven impossible to incorporate the blemish information into a data quality file primarily because they originate from several different locations, i.e., the intensifier photocathodes, the TV tube, the data store, etc., but also because many of the marks are not stationary. As an example, the reseau marks (or any other blemish associated with the photocathode), can move by 2–3 pixels (around the edge of the frame) due to small changes in the detector distortion.

None of the items discussed in this section will in any way be affected by the arrival of COSTAR.

4. Scan Distortion

Scan distortion manifests itself as a series of regular, vertical stripes down the right side of every image (see Fig. 1) in every format. It is believed that this effect is caused by the non-uniform motion of the scanning beam as it begins reading each line. If this is indeed the case, then the scan distortion will introduce variations in the geometric distortion and both the detector sensitivity and non-linear response. The scan distortion is present at the right hand edge of all formats and has a peak-to-peak scalelength of 60–70 pixels, about the same as the reseau separation. For this reason it is not removed, or even significantly modified by the geometric correction.

The intensity variation which results from the scan distortion behaves something like a damped sine wave and extends for at least 3 full oscillations (about 200 pixels), with its amplitude decreasing from about 15%, to 6%, to 2%, as it progresses inward from the right hand edge. Because of the smoothing which has been applied to the pipeline calibration flat fields (UNI files), and because the flatfield correction is applied after geometric correction, the flat fielding process does little to correct the sensitivity variation induced by the scan. It should also be noted that smaller FOC formats, which use only a subset of the 512×1024 (zoomed) format, do not include the scan distortion at all.

Since it is an inherent characteristic of the FOC cameras, there is no reason to expect that scan distortion will be changed in any way by the installation of COSTAR. Having said that, a new method of geometrically correcting the smaller formats (see the next section) may result in a reduction of the effect for these formats.

5. Geometric Distortion

Geometric distortion can be separated into three components: optical distortion, detector distortion, and format-dependent distortion. The optical component of the distortion can be described as any distortion due to the HST+FOC optical configuration, i.e., everything which originates upstream from the intensifier tube. It is determined by ray-tracing models of the optical system. The format-dependent distortions include those aspects which differ from format to format, such as the scan distortion discussed above. Most of these distortions result from events which take place downstream from the intensifier tube. Finally, the detector distortions are those which occur within the tube itself.

The accuracy of the geometric distortion correction of FOC data is currently limited by the coarseness and size of the reseau grid (60 pixel spacing). Nominally, the accuracy of the geometric correction is very good — about 0.2%. That is, the true relative separation between points after geometric correction is good to about 0.5 pixels over a separation of about 250 pixels. However, there are several conditions on this which must be fulfilled.

The geometric distortion is only explicitly defined at the positions of the reseau marks *present in the format*. The actual geometric correction is the result of a 2-D polynomial fit between the observed reseau marks and a reference grid, hence the pixel-by-pixel distortion is the result of interpolation between these known points on the assumption that the variation of the distortion has a relatively long scalelength and is both smooth and well-behaved. (Note: As discussed in the last section, the scan component of the distortion is not modelled by the fitting procedure because it varies on scale lengths approximately equal to the reseau spacing and so, after geometric correction, this component remains.) The accuracy of the fit to the distortion at any given pixel depends only on the order of the fitted function and how well it is constrained. The order of the 2-D fitting function is determined by the dimensions of the reseau grid that is present in the format being addressed, so that if there are n columns and m rows then it is normal to use a polynomial which is $[\mathrm{Int}(n/2)+1]^{\mathrm{th}}$ order in x, and $[\mathrm{Int}(m/2)+1]^{\mathrm{th}}$ order in y (for both m and $n > 2$).

If we think about the situation (see Fig. 2), it is clear that the observed reseau grid marks the limit of the constrained fit and that any areas of the image which are outside the observed grid can have no reliable geometric correction based on the reseau grid (light shading in Fig. 2.). Immediately within the reseau grid boundaries we have a region which is only constrained by a single point on the

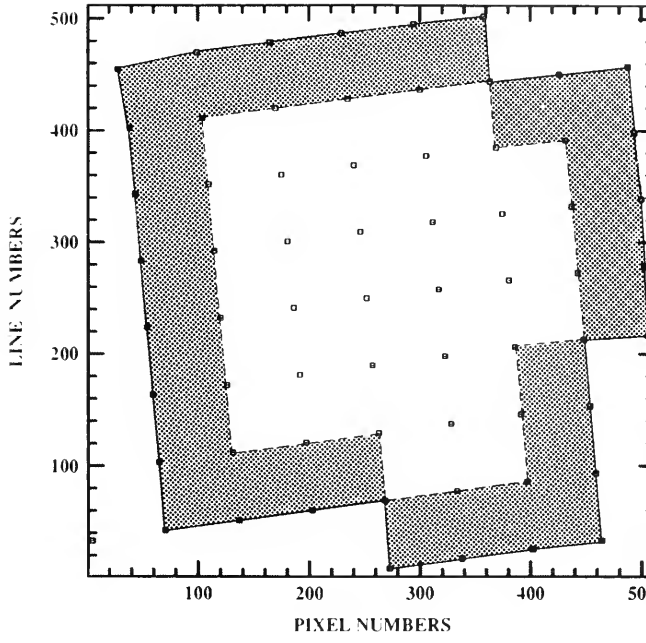


Figure 2. Observed reseau grid seen in the F96, 512×512 (normal) imaging format. Because of fitting constraints, there can be no reliable geometric correction in the light shaded regions, and only fair correction within the dark shaded region. The unshaded region at the center represents the geometrically reliable imaging area.

outside (dark shading in Fig. 2.) and in this region the geometric correction can only be considered fair, i.e., position residuals may be as much as 1–1.5 pixels. After taking these considerations into account we see that the geometrically reliable region, where the position residuals will be 0.5 pixels or less, is only the unshaded area in the center, representing about 40% of the 512×512 (normal) format. (Figures similar to Fig. 2 can be derived for other formats simply by applying the same rules.) These considerations also highlight the significance of the missing top row of reseau marks in the full 512×1024 (zoomed) format since there can be no reliable geometric correction in the region above the reseau grid which, on its own, accounts for a full 15% of the image.

The significance of this discussion is that

1. The measured distance between two points will have a larger than nominal error if one or both of the points lies in the shaded areas in Fig. 2, and

2. Structures (such as PSF haloes) which overlap the shaded areas may not be reliably corrected and may retain a residual distortion.

The advent of COSTAR itself, will make no difference to these limitations on the quality of the geometric correction. However, we are investigating a method which should produce a general improvement for most formats, with the exception of the largest (the 512×1024 zoomed). Basically, observations will be taken of a crowded star field (in 47 Tucanæ), using all of the necessary formats and, after geometrically correcting the large format, all smaller formats will be corrected using a transformation which aligns the stars with the larger format. In other words, the star positions in the geometrically corrected 512×1024 (zoomed) image will be used as the reference for all smaller formats. The standard geometric correction procedure using the reseau grid will only be used for for the large format. Since most of the smaller formats lie within the geometrically reliable area of the large format, and since the stars provide a much finer grid (a random distribution with an average separation of 10–15 pixels), it should be possible to correct these formats almost right to their edges with no significant increase in the geometric residuals. This will mean that the geometrically reliable imaging area of the smaller formats will, in effect, be virtually the whole image.

One bonus of this method is that the camera platescales will, by default, be identical in each format. A somewhat less important advantage will be that images in different formats will be automatically co-aligned.

6. Non-Linearity

Nonlinearity is probably one of the most insidious problems associated with the FOC, primarily because, until the detector becomes saturated, it is not immediately obvious to the researcher. I do not intend to discuss this in detail here, but instead refer interested parties to available documentation (Nota et al. 1993, Jedrzejewski 1993, Greenfield 1993, Baxter 1993). The FOC detectors, apart from the 8-bit wrap-over mentioned earlier, are essentially *count rate limited* — nonlinearity and saturation are not governed by the number of photons that can be accumulated, but by the rate at which they arrive and their small scale distribution, within the time taken for a single complete scan of the image. The response of the detectors differs considerably depending on whether the small scale distribution is extended or point-like. [There is also evidence that there may be a small dependence of the point source response on pixel type (zoomed or normal), however this is not certain and, for the moment will be ignored.]

As an example, if we consider the largest format (512×1024 zoomed), and specify that nonlinearity less than 10% is acceptable in our data (90% of the photons arriving at the target are registered by the detection logic), then we find that for extended sources (which include the PSF halo), the maximum acceptable count rate is 0.04 counts/pixel/sec, or 40 counts in 1000 seconds (and this with ~10% nonlinearity). For point-like distributions the response is somewhat better: for <10% nonlinearity within a 5 pixel diameter aperture centered on the star, a *peak* count rate (brightest pixel) of about 0.5 counts/pixel/sec is acceptable (we must also be aware of the halo, however, which is approaching significant levels of nonlinearity for this peak count rate).

Figure 3. Diagonal pattern noise. This effect appears to be enhanced by nonlinearity.

As a general rule-of-thumb these numbers scale inversely with the format area. If we change from a large format (with normal pixels) to a smaller one (with only 25% of the area) we get an improvement in the acceptable count rates of about a factor of 4. If the larger format is a zoomed format, and the smaller is normal, we do even better since the change from zoomed pixels to normal pixels gives a further factor of 1.75. For a change from 512× 1024 (zoom) to 512×512 (normal) we get a total improvement in the linearity characteristics of a factor of 7.

As mentioned above, even high levels of nonlinearity (>50%), are not necessarily immediately apparent when examining data visually. The only true indication is to actually check this by examining the countrates (allowing for the structure present). Saturation is sometimes even more ambiguous since, although it begins to occur when nonlinearity reaches 60%, it may not be detectable visibly (particularly for the inexperienced, examining extended sources) until the nonlinearity reaches 75–80%. When point sources saturate they begin to develop a hole slightly downstream from the peak in the scan direction (to the right of the peak), increasing in size and depth (maximum depth of zero counts) as saturation gets worse. For moderate levels of nonlinearity, (40%), a correction is possible (see references earlier in this section), however for higher levels, and for data which is saturated, it is not.

I will not elaborate on the effects of nonlinearity with respect to image restoration except to note that nonlinearity changes the shape of the PSF, particularly in the core, and therefore ignoring its presence can only have a detrimental effect on any restoration. Also, since nonlinearity is a detector characteristic, it will not be significantly affected by the installation of COSTAR except that, with the core of the PSF containing 4–5 times the power as is current, even more care will have to be taken when calculating filter combinations.

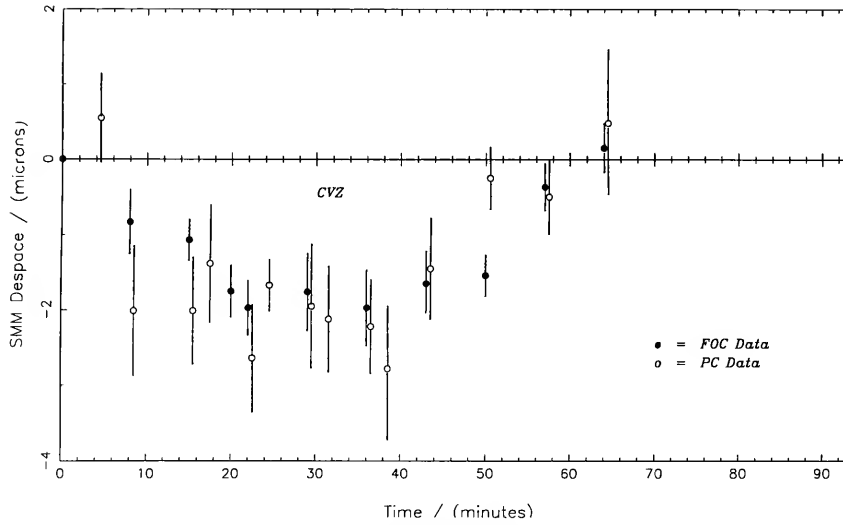


Figure 4. OTA "Breathing". Variations in SMM despace over a period of 2/3 of an orbit, measured using parallel observations with FOC and the PC.

7. Pattern Noise

Although not always seen in low S/N data, there are two patterns (unfortunately not fixed) which are always present. The first, and least obvious, of these arises because of interference between the read beam and an FOC digital timing waveform which has a 4 pixel period. It appears in the data as a vertical striped pattern. Although very coherent with regard to orientation and frequency, this pattern is inconsistent from image to image.

Second, and more often seen, is an approximately sinusoidal pattern which produces a striping orientated at about 45°to the frame axes (see Fig. 3). This pattern has a period of ~3.35 pixels (F/96) and is believed to originate from a Moiré effect between a TV tube grid and the diode array on the target. The peak deviations of this sinusoid can be as much as 10% or more and it seems to intensify when countrates are in the nonlinear regime. In fact, it can be used as a quick way of recognizing the presence of significant nonlinearity in an image since it does not appear to be present in very low count rate images (<0.01 counts/pixel/sec).

One final pattern effect which should be mentioned, although it is not a true pattern, is the effect which we call the "thumbprint". This feature is often seen by observers in the low S/N background of their geometrically corrected data. In appearance, it looks like a fingerprint and has a relatively low spatial frequency, typically with a period of 20–30 pixels. This is not a true pattern since it does not represent background variations. Instead, what is seen are areas of higher and lower variance in the Poissonian noise characteristics of the background. This comes about as a result of the resampling algorithm used in the geometric correction. Essentially, if an output pixel maps directly to the center of a pixel in the input image then little or no smoothing is involved. However, when they map to points between pixels in the input the output value is a weighted average of the set of four input pixels involved (in order to ensure flux conservation). The result then is a mild smoothing, and thus a small reduction in the variance. The pattern is identical in all images which use the same geometric correction file.

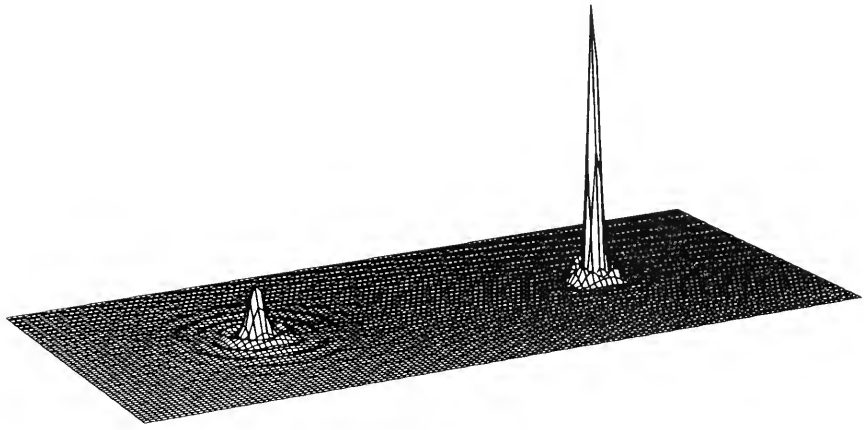


Figure 5. Comparison of the FOC pointspread function before (left) and after (right) the installation of COSTAR.

8. FOC Point Spread Functions

The appearance of the HST aberrated PSF is, by now, so familiar to everyone that I see no point in depressing people with more pictures of it. Many of the problems created by the current PSF have already been addressed. This section is simply a very quick summary and emphasis of these effects, after which we will take a more forward-looking approach. A more detailed discussion of these topics is available in Baxter et al. (1993).

The FOC PSF produced by the aberrated HST comprises a very sharp and narrow core (FWHM ~4 pixels (F/96)), containing approximately 15–20% of the total light. This is surrounded by large, highly structured halo, extending out to radii in excess of 120 pixels (although relatively faint at this radius). (Note: It is estimated that 2–5% (depending on wavelength) of the total light is scattered very thinly to radii greater than this, but we shall ignore this.) With this in mind, the first point to emphasize is that any science data (i.e., the science target itself) which lies within this distance of the edge of the format will have had some of its light scattered out of the format, from where it cannot be recovered. This is what we refer to as "edge effect" and obviously, the closer the target is to the edge, the larger the effect. The second effect related to edge proximity, and already mentioned, is geometric correction. The halos of PSFs close to the frame edges may retain residual distortions, tending to put halo structures out of alignment with equivalent structures in the restoring PSF. Finally, nonlinearity and saturation may significantly effect your science data and/or the restoring PSF. All three of these effects, if present (and they almost certainly are, to some degree), will lead to residuals being left in your restored images.

The second major topic which users should be aware of is that of OTA "breathing" which manifests itself as a pseudo-sinusoidal variation of the OTA focus. A plot of secondary mirror despace (de-focus) against time shows that the effect has a period equal to the HST orbital period, and an amplitude which varies between about 2–6 microns (see Fig. 4). Investigation of this phenomena has indicated that the cause of breathing is primarily thermal (Bely 1993), resulting from the presence of a temperature gradient between the interiors and surfaces of certain HST structural members. As well as the very clear orbital dependency, it appears that there is a significant

dependence on attitude of the spacecraft with respect both the Sun and the bright Earth. The effect of this variation in the secondary mirror despace on the HST point spread function is that it produces a radial movement in the structures in the halo. Also, diffraction rings close to the core of the PSF tend to increase and decrease in radius.

From the point of view of image restoration, this could represent a serious constraint since it has been shown (Baxter et al. 1993) that restoration of perfect data (noise-free, undistorted, linear, and high S/N) with a PSF differing by only 4 microns in focus (otherwise identical) leaves significant residuals in the restored image, and that these are only enhanced and amplified by continued iteration. OTA breathing will continue to be a problem, even after the HST repair mission.

It would be expected that the successful deployment of COSTAR should remove the need for extensive applications of image restoration techniques to FOC data, however there will, no doubt, still be occasions when they will be used. COSTAR will greatly alter the appearance of the PSF, most notably in that the halo should contract in both power and radius. Some 70–80% of the total power of the PSF should reside within a radius of 10 pixels with the peak count increasing by a factor of 4–5 relative to the pre-COSTAR PSF (see Fig. 5). The halo of the post-COSTAR PSF should, for all practical purposes, disappear at a radius of about 30 pixels (0.″4 for F/151), which will also define the new perimeter width for edge effects. In fact, the 400–500% increase in the point source peak count rates, combined with the non-linearity constraints outlined in §6, should act to suppress the PSF halo even further, making it significantly visible only when relatively long exposure times are employed.

Unfortunately, these improvements in the PSF do not come without cost. In particular, COSTAR introduces a position dependence into the FOC point spread function which we have not had to deal with previously (see Fig. 6). The FOC was designed to image the HST focal plane at a distance of 6.′56 from the optical axis (V1). At this distance the focal plane is tilted at an angle of 10° with respect to V1, and it is this plane which the FOC images. Unfortunately the focal surface produced by COSTAR is itself tilted with respect to this plane and introduces field-dependent focus variation and astigmatism (Jedrzejewski et al. 1993). Both of these effects increase linearly with distance from the fully corrected field point. In Fig. 6 we show a simulation which compares the post-COSTAR PSF as it should appear at the center of the photocathode (the fully corrected field point), and at the bottom right corner of the 512×1024 (zoomed) format (the worst case). It can be seen that the effect of this field dependence is small, but is still likely to create problems for stellar photometry and image restoration.

9. Signal-to-Noise Ratio

Considering the count rate limited nature of the FOC, it should be clear that achieving high, or even good, signal-to-noise, while at the same time maintaining acceptable levels of linearity, is rather difficult. Only the smaller formats will accept relatively high count rates and give good linearity response. Unfortunately the data is then ruined by edge effects.

The most important reason why S/N should be taken into account when considering image restoration is that it is the S/N in your data which determines the relative importance of having an accurate and applicable PSF to do the restoration with. If your data consists of faint, diffuse, extended source emission, with low S/N (peak counts of less than, say, 50 counts/pixel, there is no particular need for high S/N PSFs. The results obtained from the data using, for example, a narrow width Gaussian approximation for the PSF, would probably be as valid as if a perfect empirical FOC PSF had been used — it is the data itself which provides the limitation on image restoration. In any case, since the intrinsic image structure is usually not known *a priori*, in general it is the "nicest" reconstruction which is subsequently used.

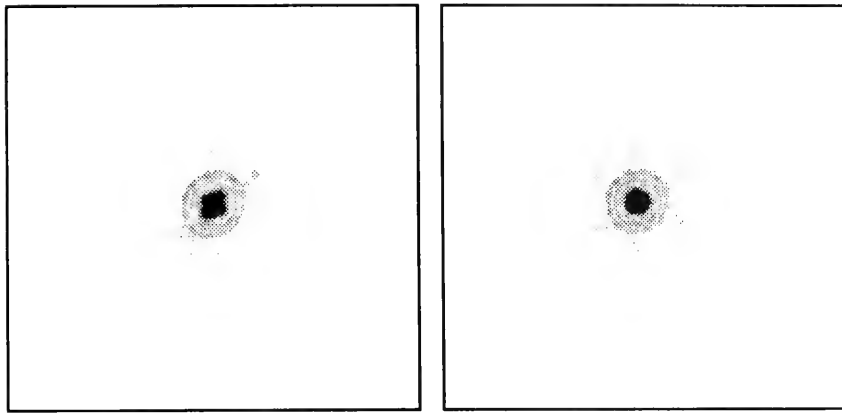


Figure 6. Simulation of the post-COSTAR, field dependent focus variation and astigmatism. On the left is the predicted PSF for the fully corrected field position (intended to be the center of the photocathode), and on the right, the appearance of the PSF when placed in the bottom right corner of the 512×1024 (zoomed) format, the worst case situation.

10. Discussion

In this review I have tried to highlight some of the major factors which must be considered before image restoration work is carried out on FOC data. Although it may seem rather pessimistic, it is not intended to be so since the FOC does what it does very well. It would be fair to say that apart from the limitations imposed by the spherical aberration, the F/96 camera has performed very close to expectations. With the exception of pattern noise and some of the details about detector blemishes, most of the topics presented here were fully expected as being characteristic of this type of detector.

When it was discovered that the HST primary was suffering from spherical aberration, it became clear that image restoration was going play an important part in HST science. Unfortunately, the Faint Object Camera is not the most suitable instrument if image processing is intended or required. This is not to say that image restoration cannot be done on FOC data, only that, as I hope I have shown, great care is required in designing of proposals, preparation of data, and interpretation of results. All too frequently, restorations of FOC data are presented in the literature without any substantial discussions of uncertainties or error analysis. Much more emphasis is required in this particular area.

The significance of most of the topics presented here will be obvious to those members of the community who are familiar with the uses and limitations of the various image restoration algorithms currently available. Many more less experienced users however, are enticed into using these techniques because they are easily available (in IRAF for example); it is only later that the questions begin. To these people I would say: *"A little knowledge is dangerous!"* Before using any image restoration package, take the time to learn 1) What does it do?, 2) How does it do it?, 3) What are its limitations and user parameters?, and of course, the old favorite, 4) When should I stop? Inexperienced users should be skeptical. For example, expressions like "flux conserving" in general mean that the sum of all the counts in an image before and after restoration will be the same. It does not necessarily mean that photons have been put back to their point of origin, only that they are still in there (somewhere).

Although much work is presently being done on error analysis of restored data and errors associated with stellar photometry on restored data, it should be remembered that, in general, image restoration procedures are cosmetic. If they are used only to enhance the visual impact of data then that is fine. However, if you then begin to make measurements, BEWARE.

Acknowledgments. Many people contributed to this work, in particular P. Greenfield, R. I. Jedrzejewski, W. Hack, P. Hodge, A. Nota, and F. Paresce. All images shown were obtained with the Faint Object Camera onboard the NASA/ESA Hubble Space Telescope, obtained at the Space Telescope Science Institute, which is operated by AURA, Inc. under NASA contract NAS 5-26555. D. Baxter acknowledges support from ESA through contract 6500/85/NL/SK.

References

Baxter, D. A., Greenfield, P. E., Hack, W., Nota. A., Jedrzejewski, R. I., & Paresce, F. 1993, in *Space Astronomical Telescopes and Instruments II*, SPIE, Vol. 1945

Baxter, D. A. 1993, Faint Object Camera Instrument Science Report, FOC-073, Space Telescope Science Institute, Baltimore

Greenfield, P. E., Nota. A., Jedrzejewski, R. I., Hack, W., Hasan, H., Hodge, P., Baxter, D. A., Baggett, W., & Paresce, F. 1993, in *Space Astronomical Telescopes and Instruments II*, SPIE, Vol. 1945

Greenfield, P. E. 1993, Faint Object Camera Instrument Science Report, FOC-074, Space Telescope Science Institute, Baltimore

Jedrzejewski, R. I. 1993, Faint Object Camera Instrument Science Report, FOC-062, Space Telescope Science Institute, Baltimore

Jedrzejewski, R. I., Hartig, G. F., Nota. A., Greenfield, P. E., Baxter, D. A., Hack, W., & Paresce, F. 1993, in *Space Astronomical Telescopes and Instruments II*, SPIE, Vol. 1945

Nota, A., Jedrzejewski, R. I., Hack, W. 1993, Faint Object Camera Instrument Handbook, Version 4.0, Space Telescope Science Institute, Baltimore

Paresce, F. 1992, Faint Object Camera Instrument Handbook, Version 3.0, Space Telescope Science Institute, Baltimore

The Restoration of HST Images and Spectra II
Space Telescope Science Institute, 1994
R. J. Hanisch and R. L. White, eds.

FOC Observations of SN 1987A: The Movie

Peter Challis

Harvard-Smithsonian Center for Astrophysics, 60 Garden Street, Cambridge, MA 02138

Abstract. HST has now observed SN 1987A six times over the past 3 years with the FOC as part of the Supernova Intensive Study program. To help visualize the evolution of the debris and circumstellar matter, a movie was created from these observations. The movie provides a unique perspective on the history of SN 1987A covering the epoch 1278 – 2424 days since explosion.

The movie consists of 5 sequences each showing a particular aspect of SN 1987A. For each sequence, 100 images (or one every 11 days) were created by interpolating between the observations, using a cubic spline. The images were then shown in succession at a rate of 66 days/second using the animation software in IDL.

The first sequence shows the raw FOC F/96 F501N [OIII] observations centered on the debris and with a field of view of 2.″2. The fading of the circumstellar ring and debris is quite dramatic. The entire ring fades surprisingly uniformly. The Richardson-Lucy algorithm was used to deconvolve the F501N images to create the next sequence. A nearby bright star in each image was used as the PSF. Since we have low S/N nebular data we stopped the iterations at a relatively low number of 30. The deconvolved images show convincingly that the structure of the ring is clumpy, that the clumps are all about the same size, and that the debris is extended. The third sequence shows the step by step iteration of the deconvolution. This is followed by a comparision between a raw and final deconvolved image. This comparision provides confidence that the deconvolved images restore the raw images without creating false structure.

The fourth and fifth sequences show the raw and deconvolved images of only the central debris from the FOC F/96 F275W observations with a field of view of 0.″56. The expansion of the debris is striking. Both sequences show some asymmetry and possible extension in the N-S direction. It will be interesting to compare the recent deconvolved images of SN 1987A with the COSTAR improved images. This method of interpolating and animating provides a visual tool to aid in the understanding of time sequenced data.

References

Panagia, N., Baxter, D., Gilmozzi, R., Macchetto, F., Adorf, H.-M., Jakobsen, P., & Kirshner, R. P. 1993, in *Science with the Hubble Space Telescope*, P. Benvenuti & E. Schreier, eds., 175

White, R. L 1993, in *Restoration Newsletter*, R. J. Hanisch, ed., Space Telescope Science Institute, 1, 12

The Restoration of HST Images and Spectra II
Space Telescope Science Institute, 1994
R. J. Hanisch and R. L. White, eds.

Superresolution of SN 1987A: A Comparative Deconvolution Study

D. S. Briggs

National Radio Astronomy Observatory, P.O. Box 0, Socorro, NM 87801

Abstract. Supernova 1987A in the Large Magellanic Cloud presents unprecedented opportunity to observe the evolution of a supernova at all wavelengths. While optical observations with the HST Faint Object Camera have obtained resolutions of 0.″1, we are limited in the radio to the resolution obtainable with the Australia Telescope Compact Array (ATCA). At the highest frequency of 8.8 GHz, this corresponds to a synthesized beam width of 0.″9. At this resolution the radio supernova is distinct from a point source, but few physical conclusions can be drawn. We present here superresolved images from the ATCA with an effective resolution of 0.″5. These reveal a spherical shell-like structure with a radius of 0.″6, and an additional component of emission aligned with the circumstellar ring imaged with the FOC. Care is required due to the delicate nature of the imaging problem, and we present evidence that the overall image structure is plausible.

1. Introduction

The Australia Telescope is the premier imaging radio interferometer in the Southern hemisphere, and the only such modern instrument positioned able to observe SN 1987A. The compact array consists of six 22m antennas with a maximum baseline length of six kilometers. Additional Australian antennas can be combined with the Australia Telescope Compact Array (ATCA) for Very Long Baseline observations, but to date there are no detections of the supernova on baselines longer than six km. The data used for these images were taken on 21 October 1992 and 4–5 January 1993. The source was followed for a full 12 hour track each day, with a small amount of data rejected in periods of poor atmospheric phase stability.

Deconvolution in radio interferometry is more usually used to remove the far out side lobes caused by the incomplete sampling of the Fourier plane than for superresolution. While it has been known for many years that modest degrees of superresolution can be successful on simple sources, it is often the case that desired higher resolution can be obtained directly by using data from a different instrument, different array configuration, or different observing frequency. SN 1987A is unusual in that it is a crucially important object where higher resolution radio imaging data is simply not available. While the supernova will expand with time and thus provide a better look at the remnant to come, the existing data are all that will ever be obtained of the supernova at the current epoch of evolution. There is great incentive to extract as much spatial information as possible from the existing data. Even a map of very low dynamic range will be extremely valuable astrophysically, so long as the features interpreted are reliable.

2. Deconvolution

The dirty maps were made with uniform weighting. The slight increase in resolution from superuniform weighting was not deemed important enough to offset the increase in noise and vulnerability to striping. The cell size 0.″1 was selected as adequate to oversample the minimum effective resolution, and most maps were of size 512^2. Iterative algorithms were started with a flat default, as more

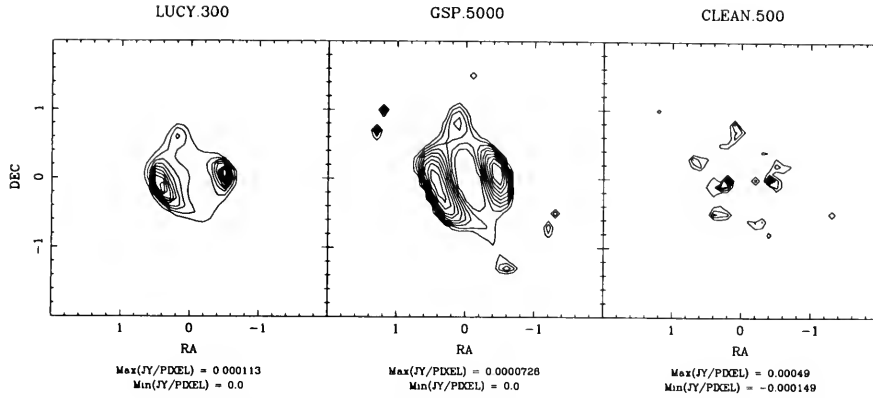


Figure 1. Sky models for different deconvolution algorithms. From the left: Richardson-Lucy after 300 iterations, Gerchberg-Saxon-Papoulis after 5000 iterations, and Högbom CLEAN after 500 iterations. While the character of these models differs substantially, all of these look much the same when smoothed down to the nominal diffraction limit. The position of the hotspots is essentially the same in all of them, and there are hints of the shell even in the CLEAN sky model. All contour intervals in this paper are 10% of the peak value.

sophisticated choices of starting point made little difference. The final maps were verified to be insensitive to the precise values of these and other parameters.

The deconvolution algorithms described below were used for the initial mapping. Some algorithms diverged unless a tight support constraint was used to confine the emission to the immediate region of the supernova. The more robust algorithms were used to verify that this was a reasonable assumption. All algorithms appeared to work better with tight support constraints, although occasionally artifacts were produced at the edge of the box. All algorithms produced similar results at the diffraction limit.

All of these algorithms operate in the image plane, and all use a model of the sky which is measured in Jy/pixel. This model is then convolved with a restoring beam representing the desired final resolution, and the residuals (dirty map − beam ∗ model) are added back to produce the final image normally interpreted by the astronomer. For the purposes of the superresolution study, it is often more productive to examine the sky model directly.

Maximum Entropy (Cornwell & Evans algorithm.) This algorithm produced what appears to be the best reconstruction found. It agrees with the other algorithms on the major features, produces a plausible physical image, performs well on simulated data with a high degree of superresolution, and is robust to support constraints. For this project it is the algorithm of choice.

CLEAN Both the original Högbom algorithm and a positivity constrained version was used. While quite acceptable at the diffraction limit, CLEAN is *not* recommended for superresolution purposes.

Maximum Emptiness This algorithm behaved similarly to MEM, but the emptiness constraint appears somewhat less powerful than the entropy term of MEM. It converges slowly and the degree of superresolution is less than with MEM.

Gerchberg-Saxon-Papoulis GSP produced a final result very similar to MEM, but diverged without support constraint.

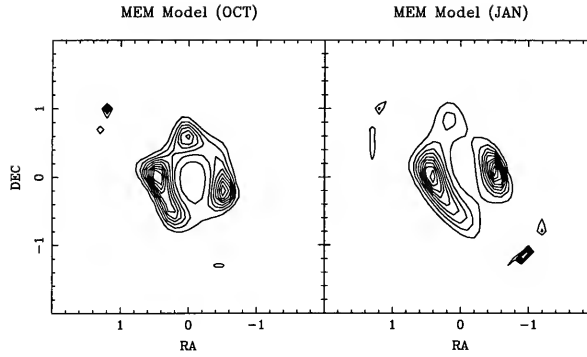


Figure 2. Cross validation: Maximum entropy models for two independent epochs of observation. While the noise level for each model appears appreciably higher than in the models derived from the composite data set below, the essential features are common to both of these model. The isolated pixels are artifacts of the hard support window used.

Richardson-Lucy This algorithm was comparable to Maximum Emptiness in that it produced a physically reasonable model but at a lower resolution than MEM.

Non-Negative Least Squares Direct fit of pixel values to the data is subject to the singular nature of the dirty beam. Constraining the pixel values to be non-negative did regularize the problem, but still produced a very non-physical model.

Direct Algebraic Inversion Direct inversion of the deconvolution equation can also be regularized with a Singular Value decomposition of the beam matrix. With heuristic selection of an appropriate cutoff in singular value a model was obtained that resembled the MEM model, but the model still contained negative flux at an unacceptably high level.

There are a number of consistency checks that can be performed on the observational data itself:

- The result must not depend grossly on minor perturbations of incidental imaging parameters. In particular, check the image as a function of cell size over reasonable ranges of oversampling. The result must not depend too sharply (or oscillate!) in iteration, though many algorithms have a long slow convergent tail. If the algorithm converges, the result must not depend too strongly on the details of the support constraints.

- Cross validate different subsets of the data. In this case, different epochs were imaged separately and compared. Also, random partitions of the composite data set were compared, which smoothed the calibration differences between epochs.

- Sum the deconvolutions from different subsets of the data. If the sum is similar to the global deconvolution of the combined data, algorithm non-linearity isn't the driving force of the deconvolution and the solution is presumably more robust than otherwise.

- Understand the noise insofar as is possible. In this case, there is a problem with confusion as diagnosed by rejecting short spacing visibilities and measuring the change in off source RMS noise. Similarly, median filtering of the data did not degrade the RMS noise by the factor expected if the noise was completely thermal. Fortunately, this confusion noise appeared to be benign for our purposes.

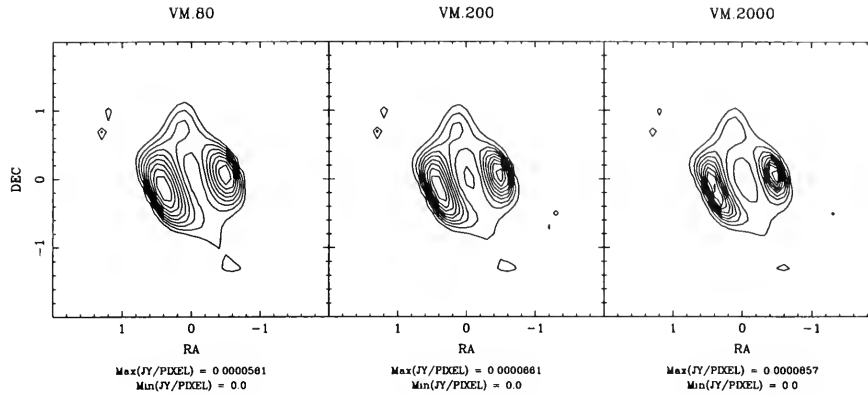


Figure 3. Convergence: Maximum entropy models as a function of iteration. From the left, 80, 200, and 2000 iterations. Notice that the peak model value in Jy/pixel is rising as the the hot spots sharpen. The ratio of peak to minimum value in the central depression is 7.7, 11.6, and 18.9, respectively.

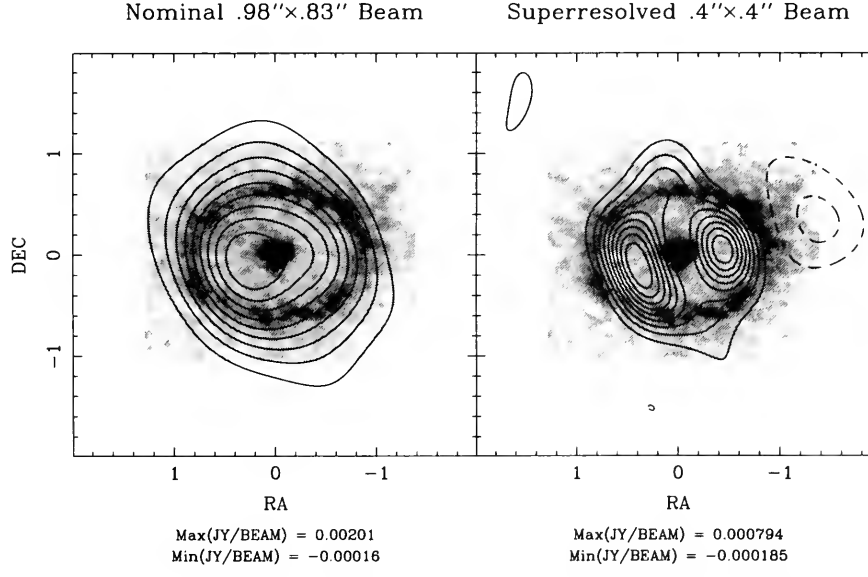


Figure 4. Comparison with other data: Model images convolved with a Gaussian restoring beam and summed with the residuals. The large negative feature in the superresolved image is found only in the January data and can be attributed to thermal noise. Also, no attempt has been made here to renormalize the residual flux scale to that of the final image, which would reduce the residual magnitude by a factor of approximately two.

- Compare the results carefully with other images at other wavelengths. Here we present our image superimposed on the [O III] image of Jakobsen et al. 1991.

3. Modelling

Modelling is the only feasible way of assessing the reliability of the reconstructions. (Modelling here refers to the model used to generate a simulated data set, not the implicit modelling performed by the deconvolution algorithms.)

1. Select plausible model from from initial images.
2. Model-fit for parameters to the original data if possible.
3. Check that simulated observations can recover the significant features of the model reliably. In this case we were concerned that the natural asymmetry in the PSF was driving the position angle of the hot spots.
4. Run plausible models less the features in question though the imaging cycle — can the features be produced spuriously?
5. Vary the parameters of the observation to significantly alter the PSF. Pathological instabilities in the beam can drive all varieties of deconvolution to a similar error.
6. Use simulated observations to determine the effective resolution of the deconvolution sky model. MEM has a bias towards smoothness, and will only make a peak as sharp as needed to agree with the data within the noise. Consequently it has a (space variant) internal resolution in the sky model. The original model was smoothed to maximize agreement with the MEM sky model. In this case, internal resolutions about about 0.″3 to 0.″4 were typical. Consequently when the sky model is smoothed with a restoring beam of 0.″4 the final image will have a resolution of 0.″5 to 0.″55.

4. Results

Image Features

- We believe the central depression and the reality of the hotspots containing a significant portion, possibly 50% of the total flux.
- The position angle of the hotspots agrees well with the FOC image, which is encouraging, but it not particularly well constrained by the radio image.
- The suggestion of emission along the polar axis is not yet judged to be significant, as there is suspicious correlation between the MEM sky model and the residual image.
- Our astrophysical interpretation of these images can be found in Staveley-Smith et al. (1993).

Algorithmic Results

- MEM is the best algorithm found for superresolution in this study. GSP, Richardson-Lucy, and Maximum Emptiness are well behaved but suffer in robustness or resolution. CLEAN, NNLS, and SVD algebraic inversion are not recommended for superresolution.
- When using MEM for superresolution, many more iterations are required than in normal suppression of sidelobes. Conventional wisdom holds that 40 iterations is usually sufficient. In this study, the image was still sharpening at 2000 iterations.

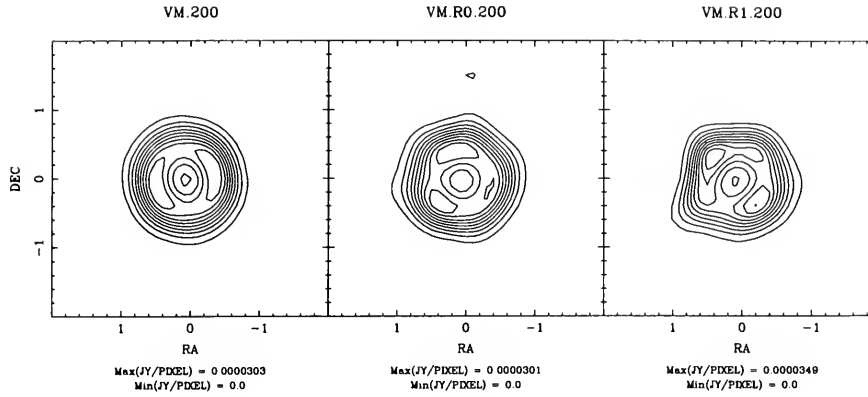


Figure 5. Modelling of spurious features: Maximum entropy models demonstrating that spurious peaks can be produced by the deconvolution process. The simulated data is derived from an optically thin sphere of emission. All peaks along on the ridge are spurious. The leftmost image is noise free, and the other two images differ only in the particular realization of the simulated noise added. The measured flux in such spurious peaks was less than 5% of the total – much less than the ~50% found in the actual observations.

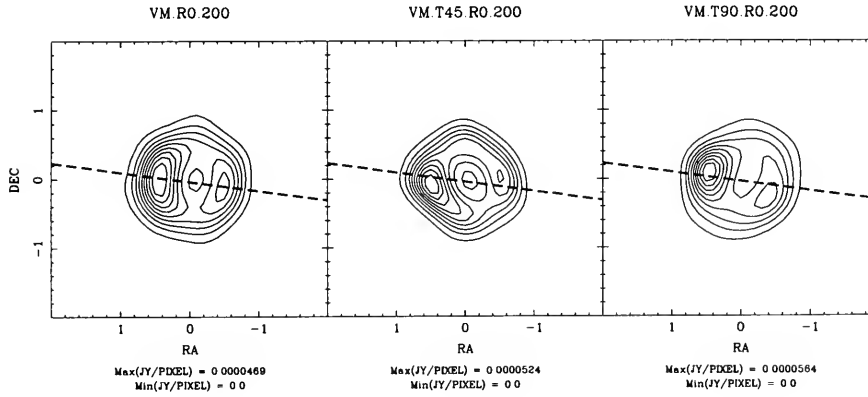


Figure 6. Modelling of feature extraction: Verification of position angle reliability by rotating the PSF while holding the model position angle fixed. The model used here, an optically thin shell plus two point sources has only 16% of the flux in the point sources — probably a conservative lower limit. Thermal noise has been added, so the real reconstruction is probably at least as reliable as this. The dashed line connects the two model point sources. The PSF is that of the original observation and rotated 45°and 90°. While not constrained closer than perhaps ±10°, the position angle of the spots is clearly not driven by the PSF alone.

- Effective resolution of MEM is a strong function of the close-in sidelobes of the PSF. Current work in progress has produced a PSF with sharper peak and better thermal noise than the conventional uniformly weighted PSF, at the expense of close-in sidelobes. The effective resolution in the MEM model gets worse, not better!

References

Staveley-Smith, L., Briggs, D. S., Rowe, A. C. H., Manchester, R. N., Reynolds, J. E., Tzioumis, A. K., Kestevan, M. J. 1993, Nature, 366, 136

Cornwell, T. J., & Evans, K. J. 1985, A&A, 143, 77

Narayan, R., & Nityananda, R. 1986, ARA&A, 24, 127

Högbom, J. A. 1974, A&AS, 15, 417

Jakobsen, P., et al., 1991, ApJ, 369, L63

Gonzalez, R., & Woods, R., 1992, *Digital Image Processing*, Addison-Wesley: Reading, Massachusetts

The Restoration of HST Images and Spectra II
Space Telescope Science Institute, 1994
R. J. Hanisch and R. L. White, eds.

WFPC Image Restoration and Undersampling Problems

S. M. Simkin and P. T. Robinson
Michigan State University, East Lansing, MI 48824-1116

Abstract. We describe two sets of Planetary Camera data for the same object. One was observed in such a way that it is adequately sampled (according to the Nyquist theorem) while the other was not adequately sampled. We discuss a variant of the "roll deconvolution" technique and demonstrate the advantages of this observing strategy for minimizing detector flaws. We show the results of attempting to restore each set of data using both the STSDAS "Lucy" and Weir's "MEM" routines. Our conclusion: **The sampling theorem applies to WF/PC data, Surprise !?**

1. Introduction

Pictor A is the seventh brightest radio galaxy in the sky at 408 MHz. It has a redshift, $z=\Delta\lambda/\lambda_0 \sim 0.036$, and a nuclear, optical spectrum which has broad, permitted emission lines (Hydrogen, Helium), and relatively narrow optical forbidden emission lines ([OI], [SII], [OIII], [NII], etc.) superimposed on a featureless continuum. A ground-based image of the galaxy in the light of Hα is shown in Fig. 1 for comparison with the HST images.

The goal of our observations was to study the physics of the energy generation in this object's nucleus by measuring the size of the narrow-emission-line region (NLR), and searching for morphological features (such as a compact, double nucleus, or an optical jet related to the radio emission) which would provide clues as to the energy's source. At the redshift of Pictor A, $1'' \simeq 500h^{-1}$ pc (where h is the scale of the Hubble constant). Current physical models predicted that the nucleus of this object should be a point source and contain roughly 90% of the total flux observed from the ground within $1''$ while the NLR should be $\sim 0\farcs1$ in size (roughly 2.5 PC pixels at 43 mas pixel^{-1}, the scale of the Planetary Camera).

2. The Choice of Camera and the Undersampling Problem

At wavelengths below 820nm, the Planetary Camera, with pixel size of 43 mas, undersamples the theoretical resolution of the 2.4m HST aperture. In spite of this, we chose to observe Pictor A with the PC because the ability of the CCD detector to accept high photon count rates would minimize observing time and maximize signal to noise in the fainter, defuse regions adjacent to the nucleus. The predicted detectable flux from the point source nucleus was $\sim$100 photons pixel^{-1} s^{-1} while the saturation limit on the FOC, which is not undersampled, is $\sim$0.5 photons pixel^{-1} s^{-1}, requiring a factor of 200 attenuation to avoid saturation and thus a factor of 200 more observing time to obtain the same S/N in the faint, diffuse regions, as the PC.

With the aberrated images, it was necessary to increase the observing time by a factor of 7 over the unaberrated exposures to compensate for the increased photon noise from the scattered light in the diffuse regions surrounding the nucleus. This is roughly the same factor by which the nuclear photon count rate is *attenuated* by the aberration and thus does not render the FOC any more attractive for this project.

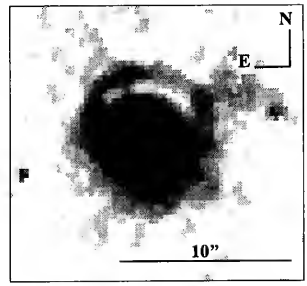


Figure 1. Hα image of Pictor A from the ground (0.″3 resolution).

3. The Data

During Cycle 1, Pictor A was observed through the F517N and F547M filters on Sept. 12, 14, 15, and Oct. 8, 1991. The F517N filter covers the emission lines from [OIII] λλ 4958 and 5007 while the F547M filter includes these emission lines at its blue end but also transmits roughly five times more continuum emission than the F517N filter. Our objective was to separate out the NLR from the continuum emission by subtracting the different filter images.

During Cycle 2, Pictor A was observed through the F648M and F718M filters on Sept. 20, 1992. The F718M filter covers the emission lines from Hα, λ 6563, [NII] λλ 6527, 6548, 6583 and [SII] λλ 6717, 6731, while the F648M filter only includes the [OI] emission lines at λλ 6300, 6363 which appear in the ground-based nuclear spectrum at a factor of $\sim$ 20 less intensity than the Hα emission line. Again our objective was to separate out the Hα + [NII] emission line regions from the continuum emission.

These observations had the obvious drawback that the true, observed PSF for this object would have an "effective filter" which was a function of the relative emission-line strengths at each point and their redshifted wavelength. However, this is a secondary effect compared with the aberration problem.

4. "Dithering" and "Rolling"

4.1. Dithering

To construct a properly sampled image for our Cycle 1 data (The HST primary mirror has a diffraction-limited resolution of 57.7 mas at 550 nm and 37.8 mas at 720nm) we chose to "dither", i.e., observe the object at four slightly shifted positions (offset in Δx and Δy by different pointing combinations of 0.022 mas as shown in Fig. 2.

4.2. Rolling

In fact, because of scheduling constraints, the observations were pointed in the way shown in Fig. 2b with (x, y) offsets of several pixels from the initial pointing position and differences in the spacecraft roll angle ranging between 1.°1 and 24.°2 (Fig. 2b).

A rotated image is the same as a non-rotated image convolved with a PSF of opposite rotation. That is, if I_i is an image at roll angle i and e_i is its exposure time, then:

$$(I_1 * P)_{90} + (I_2 * P)_{90} + (I_3 * P)_{90} + (I_4 * P)_{90} = I_{90} * (\sum P_{-i} \cdot e_i) \qquad (1)$$

where I_{90} and P are the image and PSF at 90°, the subscript 90 outside the parentheses means the image has been rotated back to 90°, and P_{-i} is the PSF at 90° rotated through the negative roll

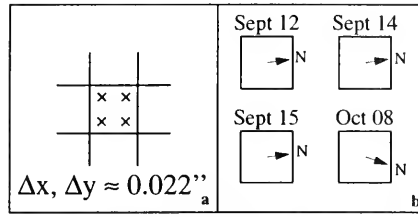


Figure 2. (a) Schematic composite of 4 "dithered" images. (b) Actual shift and orientation of the 4 Cycle 1 images of Pictor A.

angle, $-i$. Thus,

$$P_T \equiv (\sum P_{-i} \cdot e_i) \tag{2}$$

becomes a new, composite PSF for the composite image which is the sum of the individual images shifted to a common center and rotated back to a common angle.

Note that since the rotation operation can be expressed in terms of two orthogonal functions (sine and cosine), as long as two conditions hold: a) the undersampling is less than a factor of 2, and b) for each frame, if Δp is the pixel size, and both the offsets from the common center, Δx and Δy, include a pixel fraction, f_p, such that

$$0.25\Delta p \leq f_p \leq 0.5\Delta p\,, \tag{3}$$

then all of the pixels in the registered images will have sufficient offset spacings to sample a different part of the image. Both these conditions hold for the Cycle 1 data discussed here.

4.3. Advantages of Rolling

The technique described above was applied to our Cycle 1 data which was resampled on a double grid, shifted to a common center, rotated using a sinc interpolation, and then co-added. The composite PSF was generated in the same way using a set of *observed* PC6 PSFs (at 502nm). These were slightly shifted from the PC6 position of the galaxy nucleus but, again, this (along with the slight mis-match in effective wavelength of the PSFs) is a secondary effect.

One immediate advantage of this serendipitously employed technique is shown in Fig. 3 where the patterns of "hot pixels" in the co-added frames have been emphasized. Not only is it easy to spot these artifacts and thus remove them, it is also clear that over a period of 3 to 4 weeks (mid-September to October 8, 1991) some of them have decayed in intensity. This suggests that at least some of the "hot pixels" are transient, possibly induced by energetic cosmic ray events.

Another advantage of the technique (which is more intuitive) is that the composite PSF, P_T, is much more radially symmetric, with the "tendrils" deemphasized (and thus, their propensity to induce artificial, radially extensive features minimized).

5. Cycle 1 Results

The results of restorations using both the STSDAS V1.3 "Lucy" algorithm and the "MEM" interface by N. Weir (1991a, b) is shown in Fig. 4. Both routines give approximately the same normalized peak flux for the nucleus (11,800 for "Lucy" and 13,500 for "MEM"). Both show what appears to be real structure outside the nucleus. The MEM reconstruction gives fluxes for the features labeled 1, 2, and 3 as compared to the nuclear flux, f_n, of $(0.60 \pm 0.05)f_n$, $(0.50 \pm 0.08)f_n$, and $(0.35 \pm 0.04)f_n$, respectively. In particular, the feature labeled 2 appears to be identified with a very long, narrow, radio emission feature which was identified *after* the Cycle 1 data were reduced.

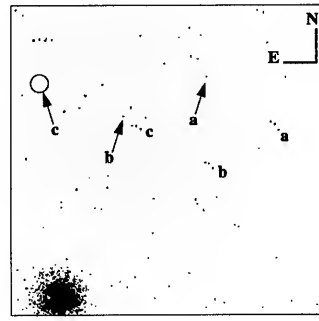


Figure 3. One quadrant of the four Cycle 1 F517N images after rotation, shifting, and co-adding. Note the decay of some of the brighter "hot pixels" over a one month period (marked by the arrows).

6. Cycle 2 Results

The Cycle 2 exposures, scheduled and taken before the Cycle 1 data were fully reduced, were designed to obtain maximum signal-to-noise, without "dithering" in the allotted spacecraft time. The images were all recorded on the same place on the PC 6 chip (to within $0.05\Delta p$) and with the same roll angle. These were also double binned and co-added. the restoration results for the F718M filter (which should show a similar extended structure as the F517N images above), using an observed, composite PSF at the correct wavelength (F718M), are shown in Fig. 5. Fig. 5a is from the STSDAS V1.3 "Lucy" algorithm (converged to a χ^2 of 1.0) and Fig. 5b from the "MEM" interface by Weir (1991a, b).

It is clear that the "Lucy" reconstruction (which converged very rapidly), shows no extended structure beyond that expected from the distortion of the aberrated PSF itself. The MEM restoration does show a hint of the type of extended structure seen in Figs. 4a and b (particularly a small feature corresponding to that marked "2" in 4b), but on the whole it also is disappointing. The most logical cause for this "failure" is simply the fact that the data for Cycle 2 are undersampled.

7. Conclusions

The technique for "roll deconvolution" outlined here is very promising for future work with the HST even after the images have been "fixed" by the first servicing mission. It appears to provide good resolution down to the theoretical limit of the telescope aperture.

Its better to dither than hang tight. But your best bet is to roll with the punches.

Acknowledgments. This work is based on observations with the NASA/ESA Hubble Space Telescope obtained at the Space Telescope Science Institute, which is operated by Association of Universities for Research in Astronomy, Incorporated, under NASA contract NAS5-26555 and with financial support from HST grant GO-245601-87A.

References

Weir, N. 1991a, in *3rd ESO/ST-ECF Data Analysis Workshop*, ESO Conf. & Workshop Proc. Vol. 38, P. J. Grosbøl & R. H. Warmels, eds., European Southern Observatory, 115

Weir, N. 1991b, *10th International Workshop on Maximum Entropy and Bayesian Methods*, W. T. Grandy & L. H. Schick, eds., Kluwer Academic Publishers, Dordrecht, 275

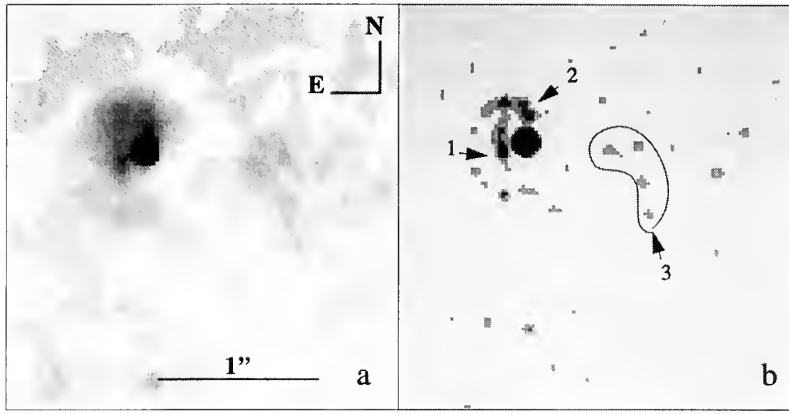


Figure 4. (a) The Cycle 1, composite F517N image restored using the STSDAS "Lucy" algorithm and a composite, observed PSF. (b) The same image restored using N. Weir's "MEM" routine.

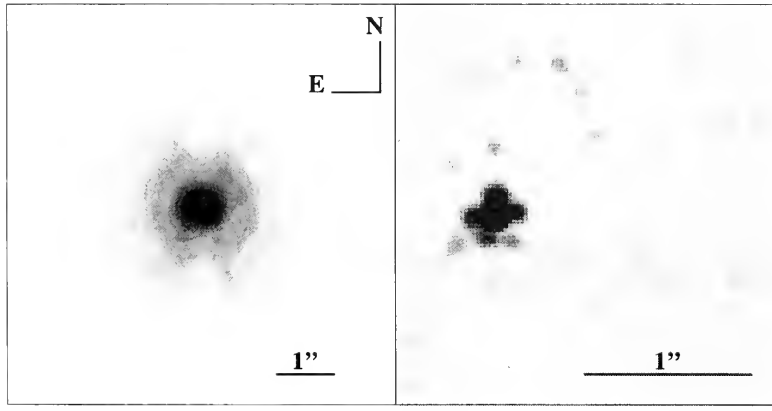


Figure 5. (a) The Cycle 2, composite F718M image (double binned) restored using the STSDAS "Lucy" algorithm and a composite, observed PSF. (b) The same image restored using N. Weir's "MEM" routine.

The Restoration of HST Images and Spectra II
Space Telescope Science Institute, 1994
R. J. Hanisch and R. L. White, eds.

Sub-Stepping as a WFPC Observing Strategy

W. Freudling

Space Telescope – European Coordinating Facility, Karl-Schwarzschild-Strasse 2, D-85748 Garching bei München, Germany

Abstract. WFPC images offer a wide field of view but significantly undersample the PSF. A strategy to recover some of the resolution lost by undersampling is to split the exposure time into several images, which are shifted by a fraction of the pixel size. These images can subsequently be combined with a modified Richard-Lucy algorithm. One of the challenges of this approach is the treatment of cosmic ray hits. A procedure to simultaneously co-add images with different PSFs in the presence of CR hits is presented here. An *IRAF* implementation (Freudling, 1993) was tested on both simulations and real WFPC data. These tests show the viability of the proposed strategy.

1. Richardson-Lucy Co-Addition

Hook and Lucy (1992) have described a method for co-adding images with different PSFs, which is a straight forward generalization of the Richardson-Lucy image restoration method (Richardson 1972, Lucy 1974). In their scheme, a multiplicative correction factor $t(x,y)$ to the current estimate for the image $o^k(x,y)$ is computed in iteration k as

$$t_n(x,y) = \frac{i_n(x,y)}{i_n^k(x,y)}, \tag{1}$$

$$t(x,y) = \sum_n w_n \, t_n(x,y) \star \mathrm{psf}_n(x,y) \tag{2}$$

where n labels the input images $i_n(x,y)$, $i_n^k(x,y)$ is the current estimate of the image $o_k(x,y)$ convolved with the PSF for image $i_n(x,y)$, and w_n is the weight of the input image.

2. Treatment of CR Hits

Once the estimate $o_k(x,y)$ is a good representation of the observed images, CR hits can easily be identified on the correction factor t_n before the convolution and Eq. (1) can be modified to

$$t_n(x,y) = \begin{cases} 0 & \text{if CR hit,} \\ \frac{i_n(x,y)}{i_n^k(x,y)} & \text{otherwise.} \end{cases} \tag{3}$$

The effect will be that there are "holes" in the final correction factor image, which are of the shape of negative PSFs for single pixel events. Those holes can overlap due to close CR hits either on the same image or on different images. The holes in the correction factor from each image $h_n(x,y)$ can easily be modeled by creating masks at the CR removal stage which are unity everywhere but at the position of the CR hit, where they are zero. From those masks, the holes are computed

$$h_n(x,y) = w_n \, \mathrm{mask}_n \circ \mathrm{psf}_n(x,y) \, . \tag{4}$$

Those holes can be filled by locally increasing the weight of the images not affected by the CR hit. If the PSF is well sampled, not all information about the affected pixels is lost because the information contents in neighboring pixels is not independent. This information can also contribute to fill the flux holes. In general, Eq. (2) can be changed to

$$t(x,y) = \sum_n w_n(x,y)\,[t_n(x,y) * \mathrm{psf}_n(x,y)] \tag{5}$$

where the weights $w_n(x,y)$ now depend on the position in the image and are computed from

$$w_n(x,y) = w_n + \frac{h_n(x,y)}{\sum_m h_m(x,y)} \sum_m (w_m - h_m(x,y))\,. \tag{6}$$

In the special case that there are no CR events on any input image, $h_m(x,y)$ equals w_m everywhere and the method reduces to the original co-addition scheme.

The only problem left is to find an initial model of the image good enough to allow easy identification of the CR hits. If more than two input images are available, the original co-addition without CR removal but using the scaled medium rather then the sum of the individual correction factors can be used. If only two images are available, the minimum of the two correction factors should suffice. In either of those cases, the full CR treatment as described above only when the model is a good presentation of the observed images.

3. Summary of CR Treatment

The correction factor computed from each image is set to zero at the positions of CR hits. The contribution to the correction factor which is then missing due to those zeros, $\sum_m (w_m - h_m(x,y))$, is filled with contributions from all images, where each image contributes proportional to its local weight. If for example the PSF is undersampled, a hole $h_1(x,y)$ in image 1 will closely resemble an otherwise flat image that is zero at the affected pixels. In that case, the missing contribution, $w_1 - h_1(x,y)$, will be filled only by the other images which are not affected. On the other hand, for a well sampled PSF, only little information is locally lost, but many pixels are affected. Therefore, the weight of the not affected images is only slightly increased relative to the affected images, but over a large number of pixels. In other words, CR hits do not only destroy information about the flux at the position of the hit, but also some information for neighboring pixels within the PSFs. This is accurately taken care of by this procedure.

4. Tests: Simulations and WFPC Data

The above procedure has been built into IRAF code originally written by R. Hook and was tested both on simulated images (Caulet & Freudling 1993) and on actual WFPC images (Koo 1993). This is shown in Figs. 1 and 2. Additional WFPC observations with sub-pixel shifts between exposures are planned by R. Fosbury and will be used to further test this procedure.

5. Conclusions

Sub-pixel shifting of WFPC image is a promising strategy to recover resolution lost by undersampling. The costs are negligible since multiple exposures of the same field are used in most cases anyway to allow removal of CR hits.

The software discussed in this paper is available from the author.

Acknowledgments. I thank Bob Fosbury for suggesting this project, David Koo for providing the WFPC images to test the algorithm, Richard Hook for his co-adding code and advice, and Leon Lucy for valuable discussions.

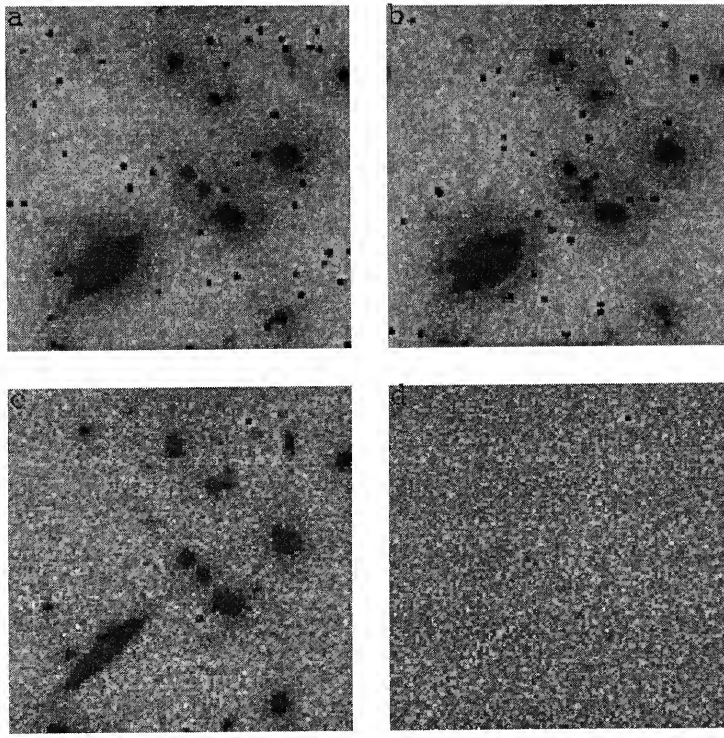


Figure 1. Simulations: Simulated WFPC observations of a cluster of galaxies. Panels a and b show the original observations shifted by 0.5 pixels. Panel c shows the co-addition with CR removal. Panel d shows the difference between panel c and a co-addition of the same images without CR hits.

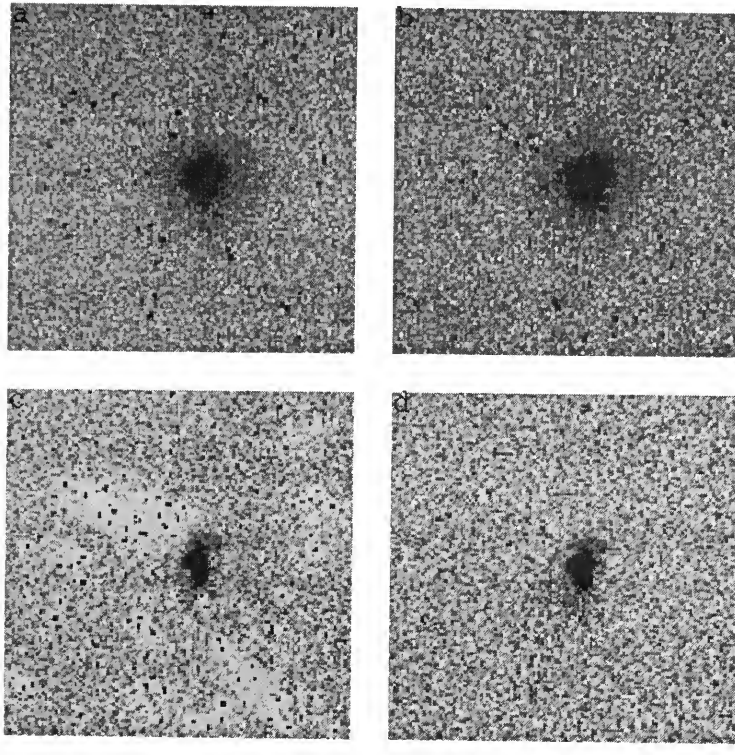


Figure 2. WFPC data: WFPC observations of a galaxy at high redshift. The two observations are shifted by about 0.5 pixels. Panels a and b show the original calibrated images. Panel c and d show co-additions of the two images without and with CR removal. Note the effect of the CR hits on the background if no CR treatment is used.

References

Caulet, A., & Freudling, W. 1993, *ST-ECF Newsletter*, 20, 5
Freudling, W. 1993, *ST-ECF Newsletter*, 20, 8
Hook, R. & Lucy, L. B. 1992, *ST-ECF Newsletter*, 17, 10
Koo, D. 1993, private communication
Lucy, L. B. 1974, AJ, 79, 745
Richardson, W. H. 1972, J. Opt. Soc. Am., 62, 55

The Restoration of HST Images and Spectra II
Space Telescope Science Institute, 1994
R. J. Hanisch and R. L. White, eds.

Optimal Combination of Sub-Stepped Spectra

J. R. Walsh and L. B. Lucy[1]

Space Telescope European Coordinating Facility, European Southern Observatory, Karl-Schwarzschild-Strasse 2, D-85748 Garching bei München, Germany

Abstract. The small aperture of the GHRS projects to just one detector diode. In order to achieve adequate sampling, the data are sub-stepped in half or quarter diode steps. Using the Richardson-Lucy algorithm, optimal combination of these sub-stepped spectra can be achieved in order to regain resolution. The technique, which can be generally applied to spectra which are undersampled by the detector, is described and some examples are shown.

1. Background

The diodes of both GHRS detectors are spaced on 50 μm centers. The limiting resolution of the instrument with the small science aperture (0.″25×0.″25, 0.″22×0.″22 post-COSTAR) is 55.5±6.5 μm (FWHM), and slightly larger for the echelle gratings (GHRS SV Report, §7.1). The spectrum projected on the 500 diodes is therefore not well sampled, with only one diode per resolution element. In order to ensure adequate sampling, the array can be electronically shifted by one half or one quarter of a diode width. The resulting spectra are called half-stepped or quarter-stepped respectively, and can be reconstituted to give two or four samples per resolution element, thus ensuring proper sampling. In imaging such a strategy is called dithering. This procedure, which can also be accomplished together with sampling of the adjacent background, is fast and efficient in terms of observing time.

Depending on the sub-stepping pattern chosen, the PODPS pipeline calibration procedure *calhrs* delivers calibrated spectra where the substep bins have been "merged". This is a simple reconstitution of the data; thus for quarter sub-stepped spectra the merged spectrum consists of
`| D1s1 | D1s2 | D1s3 | D1s4 | D2s1 | D2s2 | D2s3 | D2s4 | D3s1 | D3s2 | ...`,
where `Dn` refers to diode `n` and `sn` to substep `n`. Thus `D2s3` is the second diode from the third sub-stepped spectrum. Experiments show that this method of combination is preferable to those which involve some assumptions about the behavior of the data within the "sub-channels" of a diode, such as linear or spline interpolation onto a finer grid. In addition this merging procedure is fully reversible, allowing the individual spectra at each substep to be reconstituted. However the resulting spectrum still suffers from undersampling.

2. Optimal Combination

The Richardson-Lucy restoration technique can be used to optimally combine sub-stepped GHRS data to produce a spectrum with two or four times finer pixels, with consequent increase in spectral resolution. The technique is a modification of the restoration with increased sampling (Lucy & Baade 1989). The equation of image formation for a pixelated spectrum can be written

$$\phi(x) = \int \psi(\xi)\Pi(x|y)dy$$

[1] Affiliated to Astrophysics Division, Space Science Department, European Space Agency

where

$$\Pi(x|\xi) = \int Q(x|y)P(y|\xi)dy$$

relates the object $\psi(\xi)$ to the pixelated image $\phi(x)$. $\Pi(x|\xi)$ acts as a modified Line Spread Function (LSF) which gives the pixelated image of a point source. $Q(x|y)$ is the pixel function, viz., $1/\Delta x$ for y within a pixel of width Δx containing x, and zero elsewhere. The estimate of the desired solution ψ is produced on a fine grid with the integration performed for each sub-stepped spectrum (which can be weighted other than equally if the sub-stepped spectra have different exposure times). The LSF $\Pi(x|\xi)$ needs to be evaluated on the fine grid. This might require fitting of the LSF by analytical means in order to form the LSF on the sub-diode grid.

In restoration with non-linear algorithms such as the Richardson-Lucy method, overfitting of the data occurs as the maximum likelihood solution is approached. The result shows a fit to noise features and is clearly undesirable. By incorporating a regularization procedure into the iterative restoration, the approach to the maximum likelihood solution can be controlled by penalizing departures from smoothness. A wide choice of regularization functions is possible but two are:

$$S = -\sum \psi log_e \psi$$

which is entropy, or

$$S = -\sum \psi log_e(\psi/\chi)$$

where χ is some default image. χ may be derived from ψ such as by convolution with a Gaussian or other line spread function. The latter regularization term gives more flexibility in the extent of the smoothing and can be matched to the extent of spectral lines. Moreover this regularization greatly diminishes the bias resulting from regularization with the classical form of entropy. A multiplicative factor, α, controls the level of regularization and since this is not known a priori some experiment is required to determine the best value.

A Kolmogorov-Smirnov test can be applied after each complete restoration with a given regularization constant (e.g., Skilling, Strong, and Bennett 1979). The Kolmogorov-Smirnov test determines the greatest distance between two cumulative distributions; the distance is called the Kolmogorov-Smirnov statistic. This statistic can be expressed in terms of the significance level of an observed value of the statistic, giving the probability for the null hypothesis that both data sets are drawn from the same distribution (see Press et al. 1987, p. 472ff for details). Values of the Kolmogorov-Smirnov probability for the null hypothesis of ~0.5 provide a basis on which to choose the value of α such that the fit is neither bad nor too close to the input data.

3. Numerical Experiments

Fig. 1 shows the results of a simulation of sub-stepping and optimal combination of the sub-stepped spectra. A model spectrum with fine channels consisting of an absorption line was generated, then each set of four channels was averaged and random noise added. This set of four sub-stepped spectra were then subjected to optimal combination and the fine channel result was compared to the input model. Fig. 1 shows the four sub-stepped spectra, the "merged" result (as produced by *calhrs*), and the optimally combined spectrum without regularization (i.e., α=0). Over-fitting to the noise is apparent in this combined spectrum.

Fig. 2 shows the comparison of optimally combined quarter sub-stepped GHRS data (dotted line) against the observed data, i.e., as presented by the merged spectrum from *calhrs* (histogram). The data is of the chemically peculiar star χ Lupi and was taken with the G160M grating and the small aperture; only a part of the 2000 channel spectrum is illustrated. The LSF on the fine grid was taken to be a Gaussian whose width was that of the GHRS instrumental resolution (assumed to be 1.0 diodes FWHM). Here 100 iterations were performed without regularization. The code with the $-\sum \psi log_e(\psi/\chi)$ regularization was applied to the same data and Fig. 3 shows the result. Here the

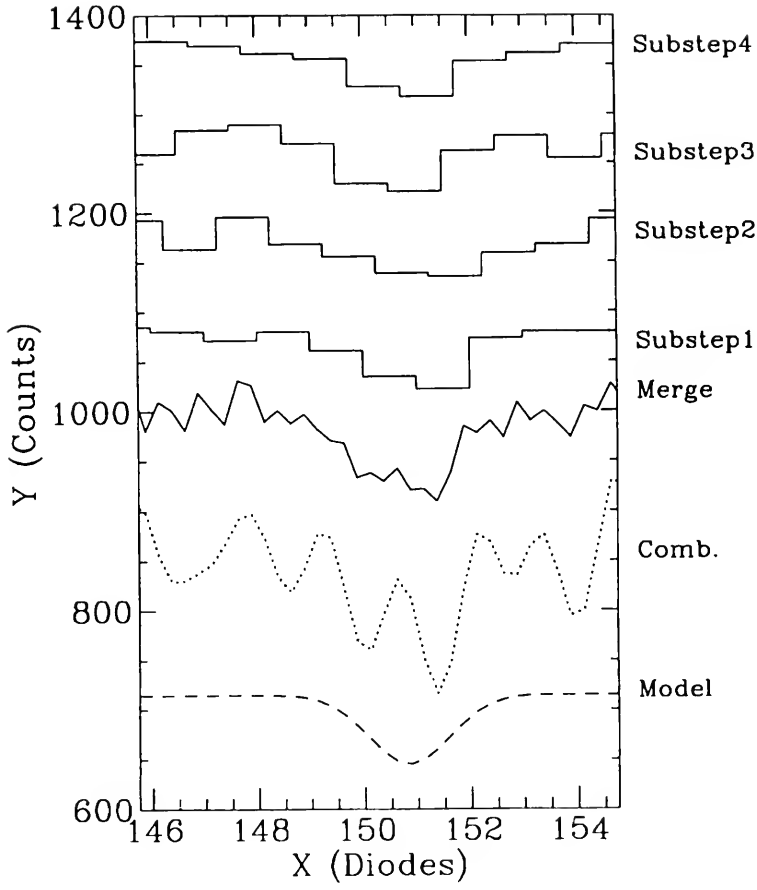


Figure 1. A simulation of sub-stepped spectra and their optimal combination with the Richardson-Lucy method. Seven spectra are shown: the noiseless model profile (labelled "Model") from which were generated four quarter-stepped spectra with added random noise (labelled "Substep"); the spectrum labelled "Merge" is produced by simply reconstituting the four substepped spectra as described in the text; finally the optimally combined spectrum is labelled "Comb".

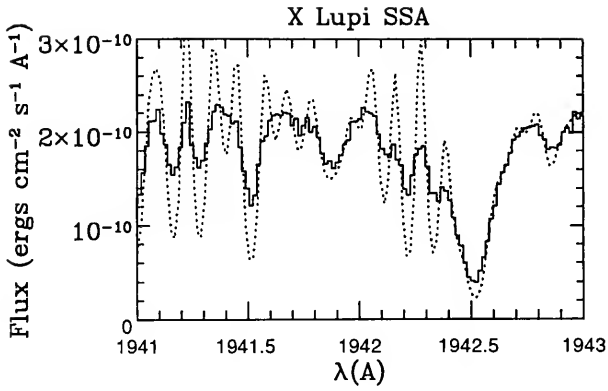


Figure 2. The spectrum formed by merging the GHRS small aperture, quarter substepped, spectra of the chemically peculiar star χ Lupi is shown as a bold line. The data set was z0g7010gx. This is the output which *calhrs* delivers. The Richardson-Lucy combination of the data (100 iterations) is shown by a dotted line. No regularization has been applied to the spectra.

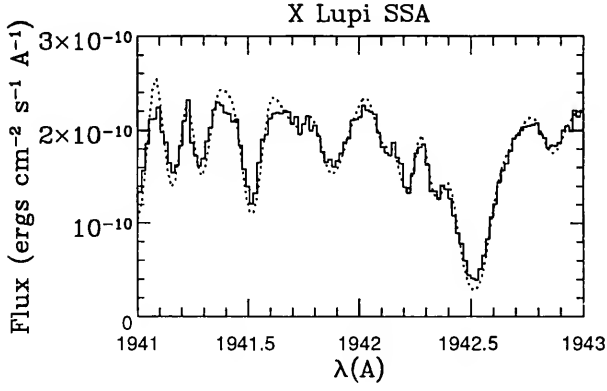


Figure 3. As in Fig. 2 except that a regularization has been applied to the Richardson-Lucy combination of the data with regularization constant $\alpha = 0.3$.

regularization constant, α, was set at 0.3 after a series of runs varying the value of α. The effect of the regularization in controlling the fit to the noise in the combined spectrum is evident. GHRS fluxed data must be converted to equivalent counts in order to perform the necessary statistical tests. After combination the data can be renormalized to flux. This operation is faciliated by multiplication by a numerical constant, which could be determined from the statistical errors delivered by the PODPS pipeline.

4. Developments

This method of combination can be generalized to spectra with non-constant steps if the separation between the steps is known, or can be computed from cross-correlation. Thus, for ground-based spectra, strategies for obtaining data with small shifts in wavelength and position on the detector are common in assisting with reduction of pixel-to-pixel noise. To this data the Richardson-Lucy method could be applied to combine the seperate spectra and improve the spectral resolution. The resolution can thus be improved beyond that imposed by the diode width, provided that the achievable instrumental resolution is less than one diode, which is usually the case in modern spectrographs with current CCD's when the slit is narrowed. Thus the pixel size of the CCD need not be seen as a limiting factor in design of spectrographs. The absolute size of the CCD limits the simultaneous wavelength coverage, whilst pixel size can be effectively reduced by sub-stepping strategies and optimal combination of the data.

References

Ball Aerospace Group 1992, *Final Report, GHRS Science Verification Program for the Hubble Space Telescope*

Lucy, L. B. 1974, AJ, 79, 745

Lucy, L. B., & Baade, D. 1989, in *1st ESO/ST-ECF Data Analysis Workshop*, P. J. Grosbøl, F. Murtagh, & R. H. Warmels, eds., European Southern Observatory, Garching, 219

Press, W. H., Flannery, B. P., Teukolsky, S. A., & Vetterling, W. T. 1987, *Numerical Recipes — The Art of Scientific Computing*, Cambridge University Press, Cambridge

Skilling, J., Strong, A. W., & Bennett, K. 1979, MNRAS, 187, 145

The Restoration of HST Images and Spectra II
Space Telescope Science Institute, 1994
R. J. Hanisch and R. L. White, eds.

Application of the Pixon Based Restoration to HST Spectra and Comparison to the Richardson-Lucy and Jansson Algorithms: Restoration of Absorption Lines

Athanassios Diplas, Edward A. Beaver, Phillip R. Blanco, Robert K. Piña, and Richard C. Puetter

Center for Astrophysics and Space Sciences, University of California, San Diego, CA 92093-0111

Abstract. We discuss the application of the Pixon Based Image Restoration Method on spectroscopic HST data and compare the results with those obtained from the Richardson-Lucy and Jansson algorithms. In order to better evaluate the performance of the various algorithms we have also used artificial data.

1. Introduction

The HST spherical aberration is responsible for the large point spread function (PSF) already demonstrated in various papers in this volume. Because of the extended PSF, the two HST spectrographs (GHRS and FOS) suffered a significant throughput loss. As a result, the use of the larger apertures was often necessary, with obviously degraded spectral resolution. Our particular emphasis in this work is the restoration of absorption lines blurred by the large line spread function (LSF). An added complication to the problem is the fact that the LSF is very poorly known. The LSFs that are available to us are derived from the PSFs, which in turn are either deduced from observations (e.g., from FOC), or from models (e.g., TinyTIM). The real LSF of course changes with time and focus position, so the LSFs used for our analysis are only approximate.

2. The Methods

We have applied the Pixon Based Image Restoration Method (Puetter, this volume; Piña & Puetter 1993, 1992; Puetter & Piña 1993) to real and simulated HST Spectra. Both the Richardson-Lucy (Richardson 1972, Lucy 1974) and Jansson (1984) algorithms have already been used extensively for spectroscopic applications (Gilliland 1990, Morris et al. 1991). All three methods are nonlinear. An undesirable characteristic of most restoration techniques is the noise amplification, over-resolution and ringing. In spectra, these problems are manifested as narrow features. Ringing is especially bad in the vicinity of emission lines, where the slope of the continuum changes significantly. Since our emphasis is on the restoration of absorption lines, we have worked only with normalized spectra. This approach can be useful when the continuum does not change rapidly over the length of the LSF. The advantage of this technique is that we can impose upper and lower bounds, namely positivity and continuum. We can therefore restrict the algorithm from unnecessarily amplifying the noise in the spectrum. Our selection of the proper continuum will be a factor of error in our analysis and caution has to be used to minimize its effects. In all cases, the inclusion of the two bounds improves the restored spectrum quality significantly, compared to the unconstrained implementation.

3. Results

We demonstrate the method for a variety of simulated cases similar to those encountered in existing spectra. Cases of interest include weak absorption lines located in the vicinity of stronger lines,

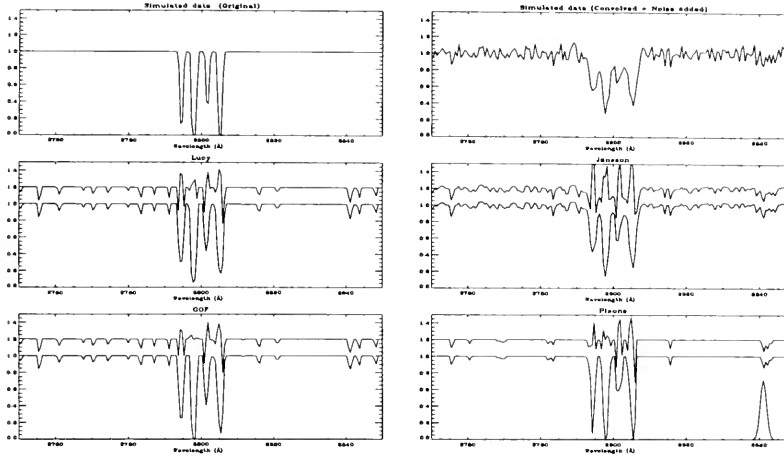


Figure 1. Restoration of a simulated Mg II interstellar line, with a high velocity feature at -375 km s^{-1}

isolated weak lines in relatively noisy spectra, and severely blended, strong saturated lines. Since most of the objects in our program were obtained with the FOS (including the 2.″0×0.″25 slit, the 1.″0 circular, and the 4.″3 aperture), we have used LSFs appropriate for these apertures for our simulations. Fig. 1 shows a simulated Mg II doublet with a high velocity component approximately 1.5 diodes (375 km s^{-1}) away from the main absorption. The line parameters used are $N_1 = 3\times10^{14}$ cm^{-2}, $b_1 = 50$ km s^{-1}, and $N_2 = 5\times10^{13}$ cm^{-2}, $b_2 = 50$ km s^{-1}. The top left panel illustrates the input spectrum. The top right panel shows the input spectrum after convolution with the LSF of the 2.″0×0.″25 slit and the G270H grating. Gaussian noise of the order of 8% is also added. The rest of the panels illustrate each one of the three methods. The GOF panel shows the restoration produced by a straight goodness of fit (GOF) criterion, without the Pixon algorithm. For all figures, the LSF used is illustrated at the bottom of the Pixon panel. Fig. 2 shows the same components, but separated only by approximately one diode element (200 km s^{-1}). The residuals with respect to the original input spectrum are plotted on top of each restored spectrum.

Table 1 contains a summary of the 3 methods examined here, for the different spectra. Overall the Pixon Based Image Restoration Method reproduces the original with higher accuracy and with fewer spurious features.

Table I. Comparison of Spectral Restoration Algorithms. The χ^2 per degree of freedom is given for each of the restorations shown in Figs. 1 and 2.

Method	Fig. 1	Fig. 2
Pixon	0.60	0.86
Richardson-Lucy	0.73	1.25
Jansson	1.27	1.95

Figs. 3, 4, and 5 show examples of restored FOS spectra towards PKS 2251+11, 3C273, and 3C286, respectively. For all three figures, the top left panel illustrates the original spectrum and the

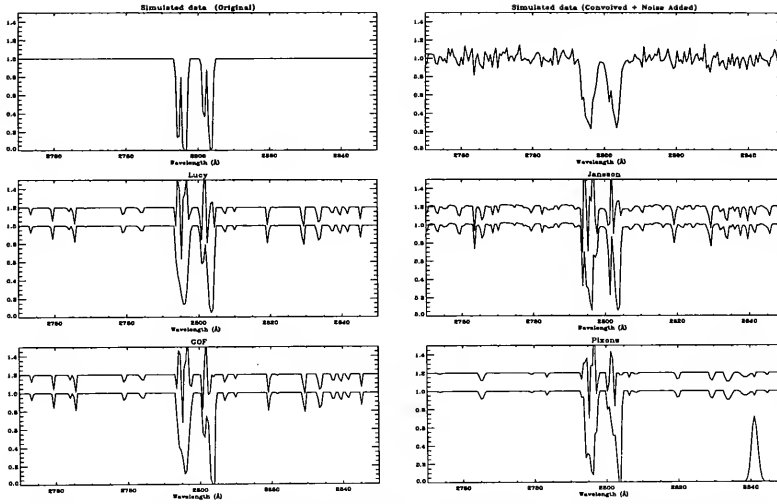


Figure 2. Restoration of a simulated Mg II interstellar line, with a high velocity feature at -200 km s^{-1} (significantly blended).

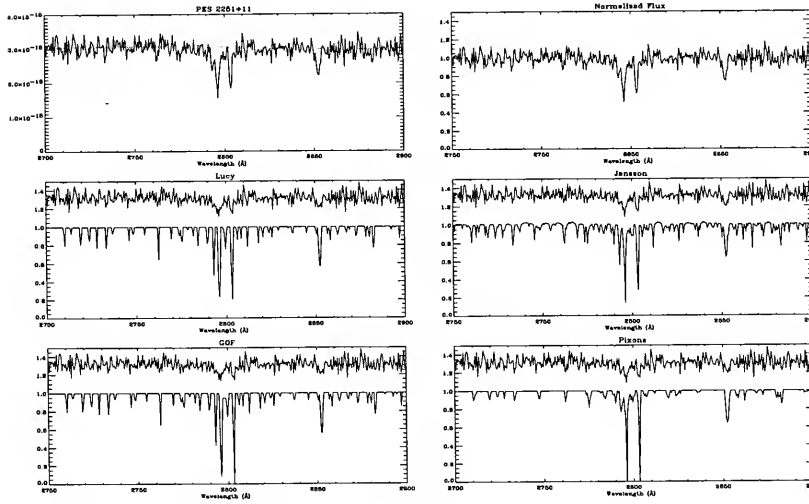


Figure 3. The FOS Spectrum of PKS 2251+11.

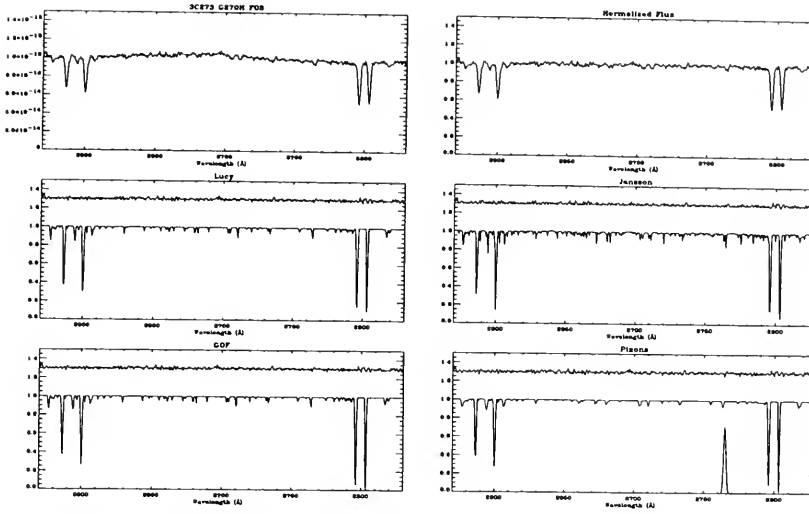


Figure 4. The FOS Spectrum of 3C273.

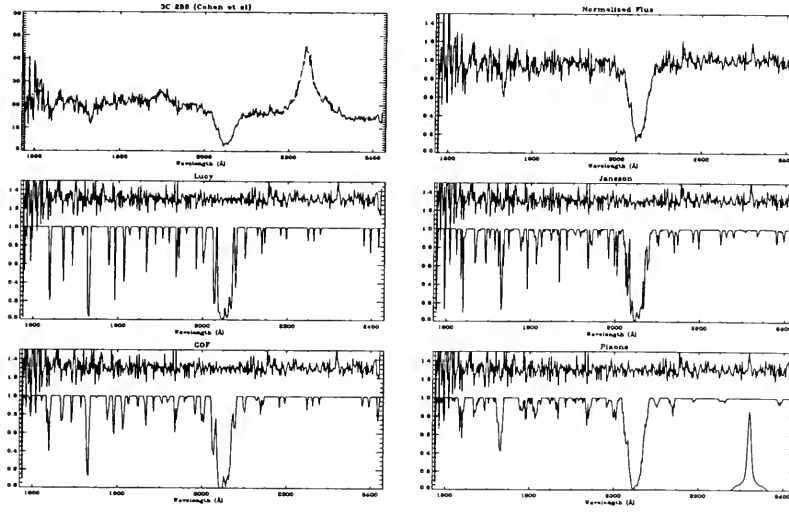


Figure 5. The FOS Spectrum of 3C286.

continuum used for the normalization, the top right the normalized spectrum, and the rest of the panels each method respectively. The residuals are plotted above the restored spectra.

A high velocity Mg II component associated with the Magellanic Stream was identified in the original spectrum, towards PKS 2251+11, by Bahcall et al. (1993) and Savage et al. (1993). The feature is clearly seen in the restored spectra. The 3C273 results from Fig. 4 can be compared with the higher resolution GHRS observations of 3C273 (Brandt et al. 1993, Savage et al. 1993), obtained with the G270M grating, through the Large Science Aperture (LSA). The Milky Way interstellar lines of Mn II, Fe II, Mg II and Mg I are clearly identified.

Fig. 5 shows the restored spectrum of 3C286 (Cohen et al. 1993). The object was observed with the low resolution grating G160L of the FOS and with the large 4.″3 aperture. The Pixon based restoration has eliminated most of the spurious features will it has restored quite well the real lines. The two most prominent lines at the left of the damped Lyman α at $z = 0.692$ are the Lyman β and Lyman γ lines at the same z.

4. Summary

The Pixon Based Image Restoration Method provides us with a new tool for spectral restoration. This work, still in progress, indicates that the Pixon algorithm can recreate the original spectral features with higher accuracy compared with the Richardson-Lucy and Jansson methods. At the same time, it seems to be more effective in suppressing spurious features.

Acknowledgments. We would like to the thank R. Puetter and R. Piña for providing us with early versions of their software, as well as for numerous valuable discussions. This work was supported by NASA FOS Science Grant NAG5-1630, and NASA GHRS Science grant NAG5-1858.

References

Bahcall, J. N., et al. 1993, ApJS, 87, 1

Brandt, J. C., et al. 1993, AJ, 105, 831

Cohen, R. D., Barlow, T. A., Beaver, E. A., Junkkarinen, V. T., Lyons, R. W., & Smith, H. E. 1993, ApJ, in press

Gilliland, R. L. 1990, in *The Restoration of HST Images and Spectra*, R. L. White & R. J. Allen, eds., Space Telescope Science Institute, Baltimore, 7

Jansson, P. A. 1984, in *Deconvolution with Applications in Spectroscopy*, Academic Press, New York

Lucy, L. B. 1974, AJ, 79, 745

Morris, S. L., Weymann, R. J., Savage, B. D., & Gilliland, R. L. 1991, ApJ, 377, L21

Puetter, R. C. 1993, this volume

Puetter, R. C. & Piña, R. K. 1993, in *Science with High Spatial Resolution Far-Infrared Data*, Infrared Processing and Analysis Center, Pasadena

Piña, R. K., & Puetter, R. C. 1992, PASP, 104, 1096

Piña, R. K., & Puetter, R. C. 1993, PASP, 105, 630

Richardson, W. H. 1972, J. Opt. Soc. Am., 62, 55

Savage, B. D., Lu, L., Weymann R. J., Morris, S. L., & Gilliland, R. L. 1993, ApJ, 404

Savage, B. D. et al., 1993, ApJ, in press (QSO Absorption Lines Key project, Paper III)

Photometry, Analysis, and Model Fitting

The Restoration of HST Images and Spectra II
Space Telescope Science Institute, 1994
R. J. Hanisch and R. L. White, eds.

Evaluation of Image Restoration Algorithms Applied to HST Images

I. C. Busko
Space Telescope Science Institute, Baltimore, MD, and
Astrophysics Division, INPE, S. J. dos Campos, Brazil.

Abstract. This work reports results on intercomparison of image restoration algorithms, when used in the specific context of stellar fields imaged by the HST WFPC. Properties as fidelity to the original image and photometric linearity, as well as computation performance, were evaluated.

1. Introduction

In the Image Restoration Project carried out at STScI, the restoration quality and scientific usefulness of several restoration algorithms were evaluated in a comparative way by numerical experiments in which a "perfect" (or "truth") image was artificially degraded, next restored by a given technique, and finally the output was compared with the original. Previous work on this line in the astronomical literature (e.g. Cohen 1991, Thomson et al. 1992) mostly restricted their efforts to the MEMSYS implementation of the Maximum Entropy method. With the availability of other image restoration techniques in public domain format (as in the Space Telescope Science Data Analysis System, STSDAS), as well as a suite of simulated data sets available on the STScI Electronic Information Service (STEIS), more systematic and complete studies were possible.

2. Methodology

This study focussed on pure stellar images, since they are probably the kind of object most often imaged by HST. Also, they present the most difficult challenge to any restoration algorithm, since the intensity gradient and dynamic range in such an image are very large when compared to usual "terrestrial" scenes, to which restoration algorithms are usually designed.

The main data set used in this study is a simulated WFPC I observation of a "star cluster", similar to the ones available in STEIS. The simulated image was built on a 8×8 oversampled data grid, later block-averaged to the actual instrument resolution. The PSF was computed by TinyTIM (Krist 1991) at $\lambda = 5500$ Å, at center of CCD #1. Finally, the appropriate noise model for WF CCDs, including Poisson and readout (Gaussian) components, was added to the image.

Algorithms studied include current STSDAS implementations of the Richardson-Lucy iteration (Richardson 1972, Lucy 1974), Maximum Entropy (Wu 1993), the Wiener filter (Andrews & Hunt 1977), σ-CLEAN (Keel 1991) and a standard Iterative Least Squares algorithm (e.g. Katsaggelos 1991). Also, an independent implementation of the Iterative/Recursive least squares algorithm (Coggins et al. 1993, Fullton et al. 1993) was tested.

Criteria to evaluate restoration quality include both generic and astronomical-specific ones. An often used goodness-of-fit criterion in image restoration work is a measure of the "distance" between the restored image $\hat{f}(i,j)$ and the "truth" image $f(i,j)$ (the one without any degradation)

$$\eta_2 = 10 \log \frac{\sum_{i,j}(g(i,j) - f(i,j))^2}{\sum_{i,j}(\hat{f}(i,j) - f(i,j))^2}$$

where $g(i,j)$ is the observation at pixel i,j. The measure is expressed in dB, and increases as long as the restored image becomes "closer" to the truth image. We also used an absolute-value distance

$$\eta_1 = 10 \log \frac{\sum_{i,j} |g(i,j) - f(i,j)|}{\sum_{i,j} |\hat{f}(i,j) - f(i,j)|}$$

which correlates better than η_2 with visual quality evaluation. Both η_1 and η_2 are independent of image content.

Astronomical criteria include photometric linearity, precision, and sky background statistics. Photometric evaluation was performed on star images using standard aperture techniques.

Algorithm sensitivity to PSF errors was evaluated by restoring the test image using both the original PSF, as well as another TinyTIM PSF, computed for a bluer star and situated $\sim$ 200 pixels away from the the CCD center.

3. Results

Table 1 summarizes some distance measure results. These measures were computed both for the full image and for a 50 pixel sq. region centered on the star cluster, dominated by crowded star images. Thus, they are sensitive to different image properties. The full image measures asses the global goodness-of-fit, but in the particular image under study they are strongly sensitive to the sky background fit. The crowded region measures, on the other hand, are more sensitive to how well star images were fitted. The highest peak value is a measure on how close the brightest star's peak approached the truth value, and so it is also a goodness-of-fit measure for star images.

Algorithms are arranged in Table 1 in increasing order of execution speed. All algorithms were iterated up to the point where at least one of η_1 or η_2 just started to decrease. Most often this happened for the full image's η_1, which is sensitive to spurious noise peaks introduced in the sky background. The exception is the Iterative/Recursive algorithm, which was run arbitrarily for 5 levels of recursion and 3 iterations at each level. σ-CLEAN was iterated to the 2.6σ level.

Visual inspection alone showed that the three algorithms: R-L, MEM0, and ILS+$\mathcal{P}$ are capable of very similar results. This is confirmed by the similar η_1 values for them. The solution obtained by the ILS+$\mathcal{P}$ algorithm, being a least-squares one, had the highest η_2, as expected. R-L produced a somewhat better solution than MEM0, but at a considerable speed penalty. σ-CLEAN produced not so good η_1 results as the methods above, mostly because of the large PSF residuals left on the CLEAN+residual map. The RILS results show that inclusion of a spatially-adaptive smoothness constraint in the least-squares iteration actually *decreased* restoration quality.

Results also suggest that the ILS+$\mathcal{P}$ iteration is the most robust against PSF mismatches, followed by R-L and MEM0, which showed very similar degradations. σ-CLEAN was the most sensitive to PSF errors.

Figure 1 depict typical results from photometry with a small aperture. Results were also obtained for a subset of isolated (non-crowded) star images, using larger apertures up to $\phi = 8$ pixel. This large-aperture data show that systematic zero-point shifts are a common feature of restored stellar images. Only the R-L iteration showed negligible zero-point shifts; all other algorithms showed $\sim$ 0.05 mag shifts to either side of the zero-residual locus.

Restoration with PSF error led to an increased scatter and more confusion. Besides, large-aperture measurements showed measurable (~ -0.02) slopes in R-L and MEM0 residuals. ILS+$\mathcal{P}$ showed only a slight increase in zero-point shift.

Comparison of results in Figure 1 with the ones in Figure 2 points to the existence of a trade-off between a particular algorithm's linearity, and its ability to control high-frequency noise buildup. Extreme cases are the Iterative Inverse and IR algorithms, which are very linear but do not control sky background noise, and RILS+$\mathcal{P}$, which produces very smooth backgrounds but at the cost of losing $\sim$ 0.5 mag of light at the faint levels in this particular test image. R-L, MEM0 and ILS+$\mathcal{P}$ seem all to offer a reasonable compromise among these extremes.

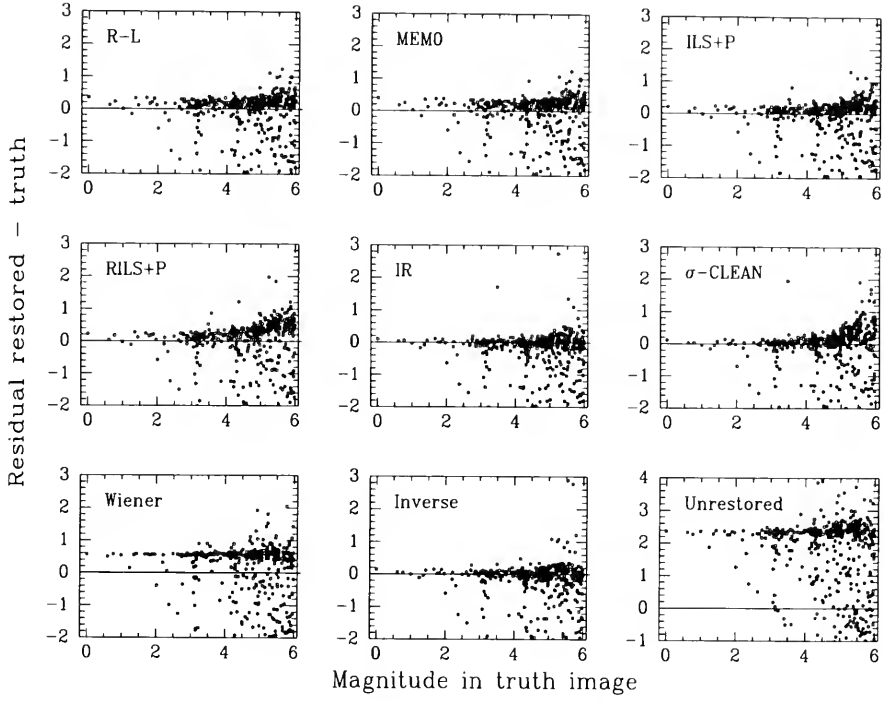


Figure 1. Aperture photometry in restored and unrestored images. A 2.2 pixel diameter aperture was used to measure each one of the 500 stars in the simulation input list. Abcissa is magnitude in truth image, ordinate is residual in the sense restored image minus truth. A negative residual indicates excess light in the aperture, probably due to crowding. A positive residual means some light is missing from the aperture, either because it was not recovered by the restoration algorithm, or because the restored star profile does not fit inside the aperture. Sky background is constant and set to zero in the aperture measurements.

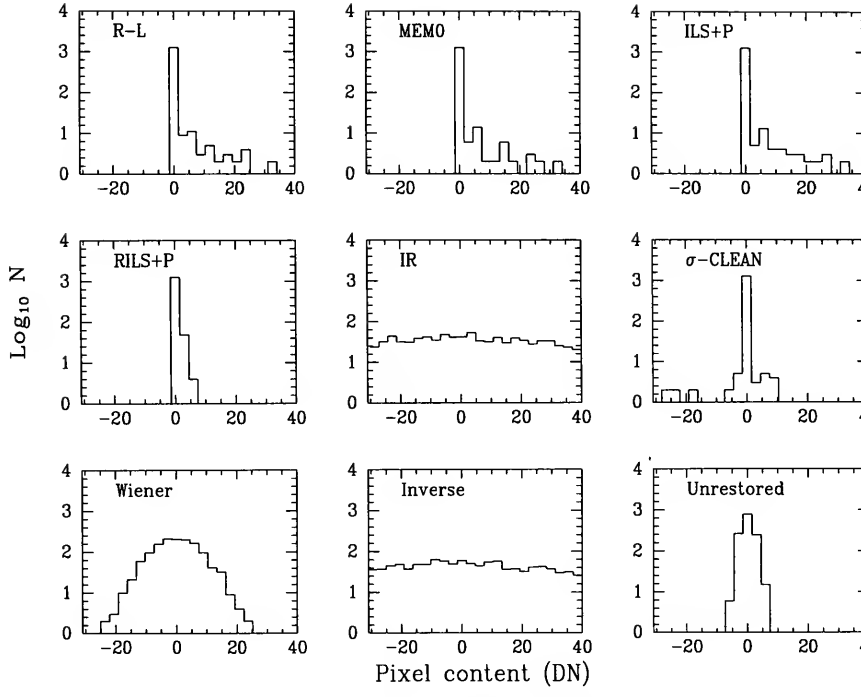


Figure 2. Sky background intensity distribution (log scale) in a 35 sq. pixel area devoid of any stars.

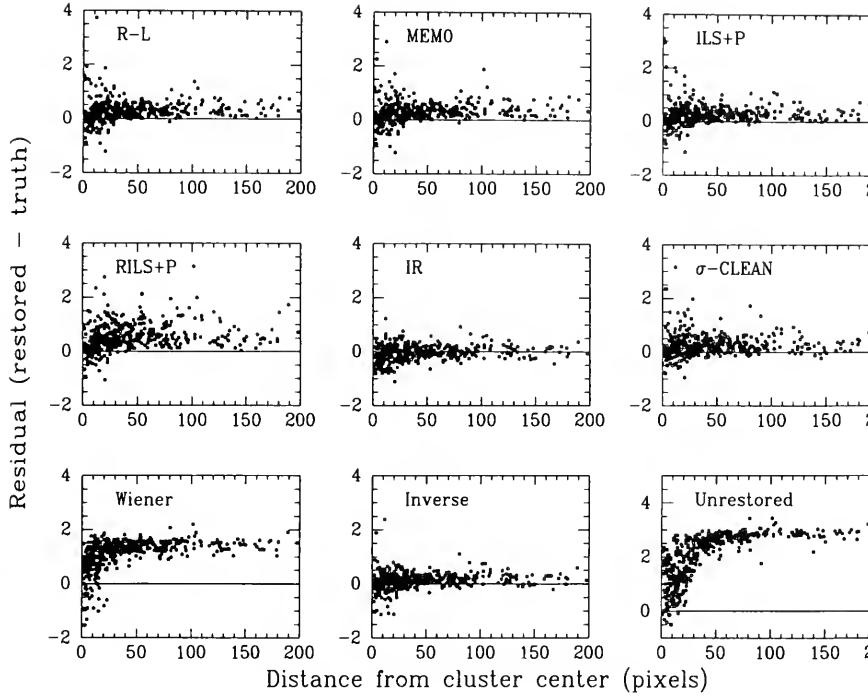


Figure 3. Effectiveness in minimizing neighbor star confusion. Residuals (in mag scale) are from star's intensity in *central pixel* only. They are plotted as a function of star's distance to cluster center. Unrestored image shows the largest degree of confusion; both iterative inverse and IR restorations the smallest. An algorithm's ability in reducing confusion is directly related to its resolution enhancement power.

Table 1. Goodness-of-fit measures. Meaning of symbols is as follows: R-L: Richardson-Lucy; MEM0: Maximum Entropy; ILS: Iterative Least Squares (no regularization); RILS: ILS with Miller regularization and spatial adaptivity; IR: Iterative/Recursive; $\mathcal{P}$: projection (positivity).

PSF errors	Algorithm	Number of iterations	CPU time[a] (minutes)	Full image η_2	Full image η_1	Crowded region η_2	Crowded region η_1	Highest peak[b] (%)
NO	R-L	3000	132	6.90	5.08	10.61	5.83	52
	MEM0	300	19	5.98	4.87	9.88	5.68	46
	σ-CLEAN[c]	40000	13	9.32	3.89	10.16	5.01	87
	ILS + $\mathcal{P}$	300	12	9.48	5.43	11.82	5.88	69
	RILS+$\mathcal{P}$[d]	150	12	9.03	4.88	11.94	5.70	65
	RILS[e]	100	3	8.44	0.82	10.31	3.18	78
	Iterative inverse	80	1	7.68	-2.90	9.90	2.61	74
	IR	3 X 5	[f]	7.41	-3.91	8.27	1.79	89
	Wiener		6 sec.	1.30	-0.52	1.39	0.74	29
YES	R-L	1500	65	5.92	4.49	9.64	5.18	46
	MEM0	300	19	4.81	4.21	8.76	4.97	39
	σ-CLEAN	30300	10	7.94	2.84	8.31	3.73	90
	ILS + $\mathcal{P}$	200	8	8.79	5.03	11.06	5.34	66

[a] SPARC10, 256 sq. image.

[b] Ratio of brighlest star's highest peak in restored and truth images.

[c] Loop gain = 0.05.

[d] $\alpha = 0.2, \mu = 0.5$.

[e] $\alpha = 0.03, \mu = 0.5$.

[f] Algorithm wasn't run in a SPARC10, but code analysis suggests that it is roughly twice as fast as the iterative inverse.

Preliminary results from an ongoing study of photometric errors from aperture photometry in restored images suggests that restoration by R-L, MEM0 and ILS+$\mathcal{P}$ does not degrade the random error level in large-aperture photometry of isolated stars. There is marginal evidence that restoration may even *decrease* the random error in faint stars. This point needs more study, and of course a detailed comparison with sophisticated PSF-fitting techniques (Stetson 1993) is fundamental to answer the question: should stellar photometry be performed in restored or unrestored images ?

References

Andrews, H. C., & Hunt, B. R. 1977, *Digital Image Restoration*, Prentice-Hall, New Jersey

Cohen, J. G. 1991, AJ, 101, 734

Coggins, J. M., Fullton, L. K., & Carney, B. W. 1993, CVGIP: Graphical Models and Image Processing, submitted.

Fullton, L. K., Carney, B. W., Coggins, J. M., Janes, K. A., Heasley, J. N., & Seitzer, P. 1994, in *Astronomical Data Analysis Software and Systems III*, D. Crabtree, R. J. Hanisch, & J. Barnes, eds., ASP Conference Series, in press

Katsaggelos, A. K. 1991, *Digital Image Restoration*, Springer-Verlag

Keel, W. C. 1991, PASP, 103, 723

Krist, J. 1991, The Tiny TIM User's Manual, Space Telescope Science Institute

Lucy, L. B. 1974, AJ, 79, 745

Richardson, W. H. 1972, J. Opt. Soc. Am., 62, 55

Stetson, P. 1993, this volume

Thomson, R. C., Schade, D. J., Elson, R. A. W., Mackay, C. D., & Wilkins, T. N. 1992, MNRAS, 259, 104

Wu, N. 1993, *Astronomical Data Analysis Software and Systems II*, R. J. Hanisch, R. J. V. Brissenden, & J. Barnes, eds., ASP Conference Series, 52, 520

The Restoration of HST Images and Spectra II
Space Telescope Science Institute, 1994
R. J. Hanisch and R. L. White, eds.

Star Detection, Astrometry, and Photometry in Restored PC Images

Don Lindler

Advanced Computer Concepts, Inc., Code 681, Goddard Space Flight Center Greenbelt, MD 20771

Sara Heap

NASA/Goddard Space Flight Center, Code 681, Greenbelt, MD 20771

Jarita Holbrook

University Space Research Association, Code 681, Greenbelt, MD 20771

Eliot Malumuth

Computer Sciences Corporation, Code 681, Goddard Space Flight Center, Greenbelt, MD 20771

Dara Norman and Patricia C. Vener-Saavedra

Advanced Computer Concepts, Inc., Code 681, Goddard Space Flight Center Greenbelt, MD 20771

Abstract. We have evaluated various image restoration techniques (both linear and non-linear) applied to a simulated crowded star field observed with the HST Planetary Camera (PC). Evaluation criteria included star detection, astrometry, and photometry. Numerous restoration artifacts made star detection difficult for the linear restorations. The non-linear methods give much better star detection and slightly better astrometry results. Aperture photometry measurements in images restored with the non-linear restoration methods show significant non-linearities. The flux of fainter stars are systematically underestimated. This problem is not seen in the images restored with the linear methods. However, we see significantly increased RMS scatter in the photometry results for images restored with a linear method. We found that a hybrid approach which combines the result of a non-linear method with a linear method will give linear photometry results with lower RMS scatter.

1. Simulated Test Data

Fig. 1 is an 800×800 simulated PC image used to test the restoration algorithms. It contains 2500 stars with a magnitude distribution shown in Fig. 2. The image was constructed using a Tiny TIM PSF at a wavelength of 4000 Å and a pixel size of 1/5 the size of the PC pixels. This sub-sampled PSF allows us to simulate stars with positions that are not required to be at the center of a pixel. We added both Poisson and read-out noise to the image to simulate the noise statistics expected when two PC images with 25 sec exposures (filter F439W) are co-added.

2. Restoration Methods Evaluated

The following restoration methods were evaluated. The central 200×200 region of the restored images are shown in Fig. 3 for selected results.

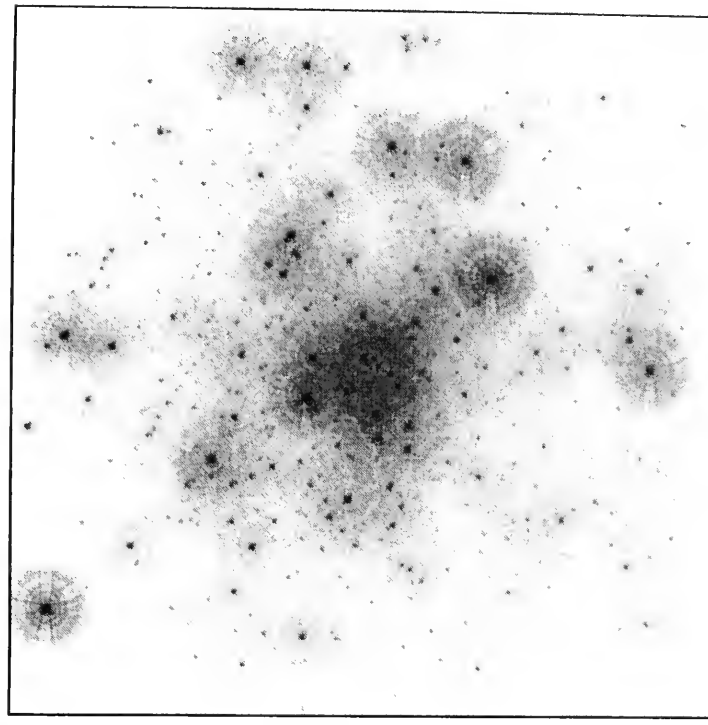


Figure 1. Simulated PC observation of a cluster containing 2500 stars.

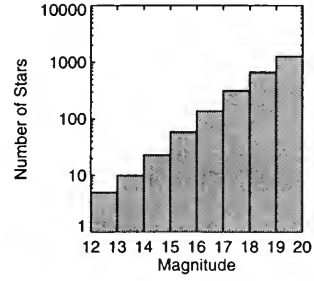


Figure 2. Histogram of the magnitude distribution of the stars in the simulated test image.

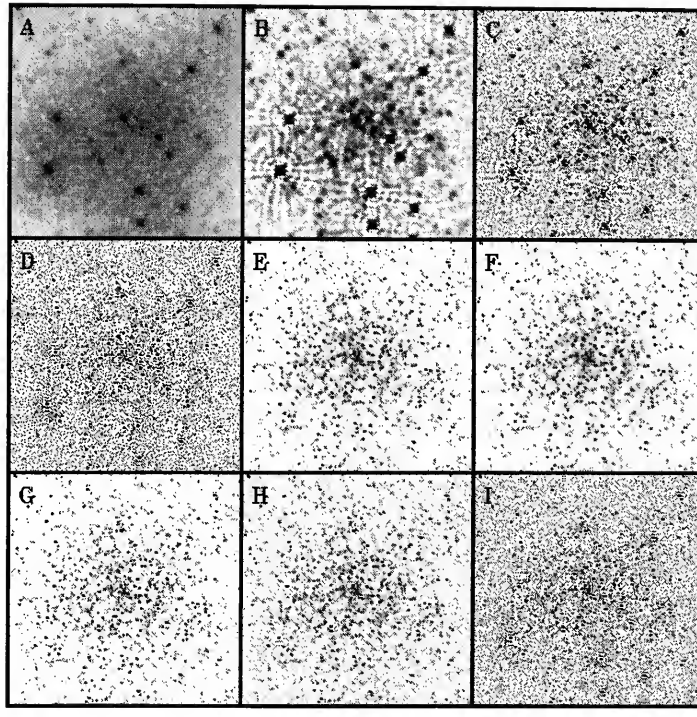


Figure 3. Selected Restorations of the central 200×200 region of Fig. 1. (A) Raw image. (B) Minimum norm, $\gamma = 0.01$. (C) Minimum norm, $\gamma = 0.001$. (D) Five iteration minimized difference from the previous iteration, $\gamma = 0.001$. (E) 1000 iteration R-L. (F) 1000 iteration R-L/Snyder. (G) 400 iteration MEM. (H) Hybrid method, $\gamma = 0.01$. (I) Hybrid method, $\gamma = 0.001$.

1. *Minimum norm.* This technique is a linear constrained least squares method which minimizes the norm of the solution. It is solved using the method of Lagrange multipliers (Andrews and Hunt 1977). A reciprocal Lagrangian multiplier, γ, can be adjusted to control the smoothness of the solution or to satisfy chi-squared test for the solution. Results are evaluated with $\gamma = 0.1$, 0.01, and 0.001.

2. *Five iteration minimized difference from the previous iteration.* This is a constrained least-squares method which minimizes the norm of the difference of the solution from a trial solution (Twomey 1963). A reciprocal Lagrangian multiplier, γ, controls the amount of constraint. In our solution we have used the raw blurred image as the initial trial solution and performed 5 iterations. After each iteration, a new trial solution was set to the results of the previous iteration with a positivity constraint applied. Results are evaluated with $\gamma = 0.1$, 0.01, and 0.001.

3. *Richardson–Lucy method.* This technique is the Richardson-Lucy (R-L) maximum likelihood solution for Poisson statistics (Richardson 1972, Lucy 1974). Results were evaluated after 100, 300 and 1000 iterations.

4. *Richardson–Lucy/Snyder.* This method is the R-L method with the Snyder modification for non-Poisson read-out noise (Snyder 1990; Snyder, Hammoud, & White 1993). Results were evaluated after 100, 300 and 1000 iterations.

5. *Maximum entropy method.* This is the Image Reduction and Analysis Facility (IRAF) maximum entropy method (MEM) developed by Wu (1994). The results were evaluated after 100, 200 and 400 iterations.

6. *Hybrid method.* This approach minimizes the norm of difference of the restored image from a trial solution (Twomey 1963). The trial solution was selected as the results of 1000 iterations of the R-L/Snyder method. Results are evaluated with the reciprocal Lagrangian multiplier $\gamma = 0.1$, 0.01 and 0.001.

3. Star Detection

The first criterion used to evaluate the restoration techniques is the ability to detect stars in the restored image without "detecting" noise and restoration artifacts. We used a simple detection algorithm which finds the local maximum in the restored image above a user-supplied threshold. Fig. 4 shows how the number of stars detected (solid line) increases as the threshold decreases. At the same time, the number of false detections (dashed line) also increases. As the threshold is decreased below 100, more than half of the additional detections are spurious.

Table 1 gives the results for various percentages of false detections. The first two methods (both linear) show very poor results. Numerous artifacts around bright stars make star detection very difficult in the linear restorations. The next three methods (all non-linear) give significantly better results. The R-L/Snyder method gives slightly better results than the R-L and MEM methods. Some improvement in star detection is seen in the hybrid method when γ is large (0.1). However, as γ decreases, artifacts begin to show up in the solution.

4. Astrometry

The second criterion used to evaluate the restoration methods is the accuracy in measuring stellar positions in the restored image. Accurate positions of the stars are important for aperture photometry in crowded star fields where small apertures must be used. The RMS errors (in pixels) in position measurements are given in Table 2 for the various restoration methods. Results are shown for different combinations of stellar magnitude ranges and separation of the stars from their nearest

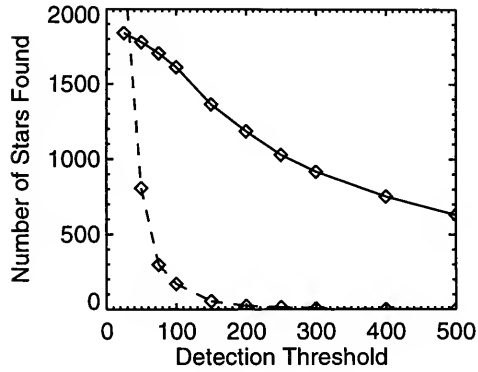


Figure 4. Number of stars detected in the 1000 iteration R-L/Snyder restoration versus detection threshold. Solid line shows the actual stars detected; dashed line shows false detections.

Table 1. Number of stars detected for various rates of false detections.

False Detections (%):	1.0	2.0	5.0	10.0
Minimum Norm				
$\gamma = 0.1$	138	157	175	205
$\gamma = 0.01$	280	312	398	519
$\gamma = 0.001$	87	104	197	404
Five iter. min. diff. from prev. iter.				
$\gamma = 0.1$	59	50	161	365
$\gamma = 0.01$	131	195	448	780
$\gamma = 0.001$	54	100	144	251
R-L Method				
100 iter	865	1034	1231	1440
300 iter	886	1090	1372	1646
1000 iter	950	1113	1450	1728
R-L/Snyder				
100 iter	737	886	1240	1831
300 iter	883	1113	1439	1707
1000 iter	968	1211	1492	1803
MEM				
100 iter	684	731	922	1213
200 iter	853	1005	1287	1756
400 iter	831	1038	1363	1677
Hybrid				
$\gamma = 0.1$	973	1217	1498	1811
$\gamma = 0.01$	958	1218	1512	1778
$\gamma = 0.001$	315	419	678	1403

neighbor. Positions were measured using a 3×3 centroid centered on the brightest pixel in the star. Centroids using areas larger than 3×3 pixels gave slightly better results for well separated stars but gave significantly worse results in crowded regions.

Table 2. Astrometry errors in 3×3 centroid, RMS errors (pixels).

Magnitude:	12 to 17.5			> 17.5 to 18.5			> 18.5		
Distance:	< 3	3 – 6	> 6	< 3	3 – 6	> 6	< 3	3 – 6	> 6
Minimum Norm									
$\gamma = 0.1$	0.57	0.40	0.21	0.79	0.63	0.37	0.96	0.84	0.63
$\gamma = 0.01$	0.43	0.32	0.17	0.70	0.50	0.26	0.87	0.72	0.56
$\gamma = 0.001$	0.28	0.21	0.16	0.51	0.36	0.23	0.77	0.59	0.52
Five iter. min. diff. from prev. iter.									
$\gamma = 0.1$	0.49	0.44	0.37	0.76	0.59	0.35	0.89	0.77	0.59
$\gamma = 0.01$	0.36	0.23	0.16	0.55	0.36	0.24	0.77	0.58	0.49
$\gamma = 0.001$	0.31	0.17	0.14	0.46	0.31	0.23	0.76	0.55	0.51
R-L Method									
100 iter	0.38	0.23	0.11	0.59	0.41	0.19	0.84	0.66	0.46
300 iter	0.31	0.18	0.11	0.51	0.29	0.19	0.77	0.57	0.44
1000 iter	0.30	0.14	0.11	0.38	0.24	0.19	0.72	0.50	0.45
R-L/Snyder									
100 iter	0.38	0.23	0.10	0.62	0.43	0.19	0.85	0.71	0.50
300 iter	0.33	0.18	0.10	0.51	0.31	0.18	0.79	0.61	0.44
1000 iter	0.30	0.13	0.10	0.40	0.24	0.18	0.72	0.51	0.42
MEM									
100 iter	0.35	0.22	0.09	0.59	0.40	0.21	0.83	0.73	0.51
200 iter	0.33	0.17	0.09	0.48	0.30	0.18	0.75	0.62	0.45
400 iter	0.31	0.17	0.10	0.41	0.24	0.18	0.72	0.51	0.41
Hybrid									
$\gamma = 0.1$	0.31	0.17	0.10	0.40	0.24	0.18	0.73	0.51	0.42
$\gamma = 0.01$	0.31	0.18	0.10	0.39	0.22	0.18	0.73	0.51	0.43
$\gamma = 0.001$	0.31	0.13	0.10	0.45	0.24	0.19	0.74	0.54	0.47

The best astrometry results were obtained using the non-linear restoration and the hybrid methods, although none of the methods give very good results for stars fainter than 18.5 magnitudes. Stars between 17.5 and 18.5 magnitudes also gave significantly poorer results when the star's closest neighbor was within 3 pixels. RMS errors of 0.1 pixel were obtained for well separated stars with a magnitude brighter than 17.5.

Fig. 5 shows the astrometry errors for well separated, bright stars in the 1000-iteration R-L/Snyder restoration. The errors are plotted for various positions of the star with respect to the center of a pixel. The diamonds give the true star positions. Vectors are drawn to the measured position. Stars not centered on a pixel show systematic errors in the position measurements. These errors are partly due to the use of only a 3×3 centroid centered at the brightest pixel. When a 7×7 pixel centroid is used, some of the systematic errors remain with a magnitude on the order of 0.05 pixels. The systematic errors occur in all of the restoration methods and are slightly larger for the linear methods.

5. Photometry

We performed aperture photometry on the restored images using a 2-pixel radius circular aperture. The small aperture size was selected to avoid confusion in the crowded regions. We evaluated both linearity of the measured stellar flux values and the RMS scatter in the results. Figs. 6 and 7 show the photometry results for one of the non-linear methods (1000 R-L/Snyder iterations) and one of the linear methods (minimum norm solution). The ratio of the measured counts divided by the actual counts in each star is plotted versus the magnitude of the star. The R-L/Snyder method shows significant non-linearity in the results for the fainter stars.

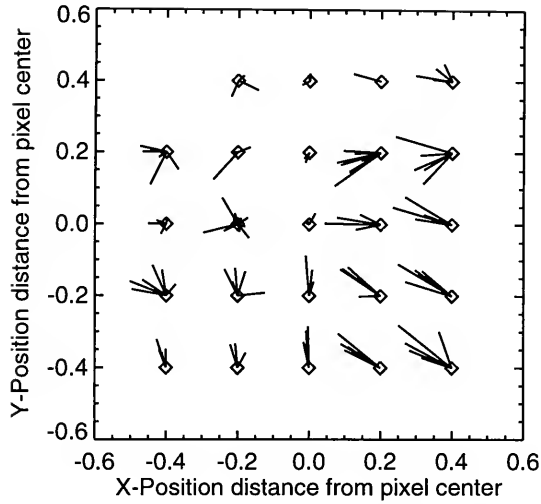


Figure 5. Astrometry errors for well separated bright stars in the R-L/Snyder restoration plotted versus the distance of the star from the center of a pixel. Diamonds give true star positions. Vectors are drawn to the measured positions.

Table 3 shows the linearity of the results for each of the restoration methods for various ranges of magnitude and distance of the stars to their nearest neighbor. All methods overestimate the flux of stars having neighbors within 3 pixels. The measured flux values are significantly contaminated by the neighboring stars. For stars with separations greater than 3 pixels, all of the non-linear restoration methods show a non-linear behavior and give smaller median ratios for the fainter stars.

Table 4 shows the RMS scatter about the median ratios in Table 3. None of the methods give very good results for fainter stars with neighbors closer than 3 pixels. The MEM solution has slightly smaller RMS values than the other non-linear methods, particularly for the brighter stars with a neighbor 3 to 6 pixels away. The hybrid method with $\gamma = 0.001$ has a lower RMS error than the linear methods while maintaining the linearity of the measured flux values.

6. Conclusions and Recommendations

Star detection is much more successful in images restored with the non-linear methods. Numerous artifacts in the linear restorations make detection of fainter stars difficult. The non-linear methods also give better astrometry results than the linear methods. The linear methods give more linear photometry results while the non-linear methods give photometry results with less RMS scatter. The hybrid approach is a good compromise giving linear flux measurements with a RMS scatter less then the linear methods.

The following approach is recommended for performing aperture photometry in restored PC images:

1. Restore the image with a non-linear restoration method. Use the result for star detection and position measurement.
2. Use the hybrid method restoration for photometry using the star list and positions from Step 1.
3. Use photometry results from a linear technique (e.g., minimum norm) to check the linearity of the results from Step 2.

Acknowledgments. We would like to acknowledge the suggestions and expert advice of the Space Telescope Science Institute's Image Restoration Project staff. This work has been supported by NASA contract NAS5-32076 funded by the Ultraviolet, Visible and Gravitational Astrophysics Research and Analysis Program.

References

Andrews, H. C., & Hunt B. R. 1977, *Digital Image Restoration*, Prentice-Hall, Englewood Cliffs, New Jersey, 149

Lucy, L. B. 1974, AJ, 79, 745

Snyder, D. L. 1990, in *The Restoration of HST Images and Spectra*, R. L. White & R. J. Allen, eds., Space Telescope Science Institute, Baltimore, 56

Snyder, D. L., Hammoud, A. M., & White, R. L. 1993, J. Opt. Soc. Am., 10, 1014

Twomey, S. 1963, J. ACM, 10, 97

Wu, N. 1994, this volume

Table 3. Photometry linearity, median ratio measured counts/actual counts.

Magnitude:	12 to 17.5			> 17.5 to 18.5			> 18.5		
Distance:	< 3	3 – 6	> 6	< 3	3 – 6	> 6	< 3	3 – 6	> 6
Minimum Norm									
$\gamma = 0.1$	0.28	0.15	0.12	0.71	0.28	0.14	2.03	0.55	0.19
$\gamma = 0.01$	0.61	0.51	0.49	0.88	0.55	0.50	2.00	0.67	0.41
$\gamma = 0.001$	1.08	0.98	0.97	1.28	0.98	0.97	2.29	1.03	0.99
Five iter. min. diff. from prev. iter.									
$\gamma = 0.1$	0.57	0.47	0.45	0.81	0.49	0.44	1.87	0.59	0.41
$\gamma = 0.01$	1.09	1.00	0.99	1.25	0.96	0.96	2.14	0.91	0.90
$\gamma = 0.001$	1.17	1.08	1.09	1.33	1.06	1.07	2.11	1.08	1.07
R-L Method									
100 iter	0.95	0.89	0.88	0.96	0.72	0.75	1.74	0.55	0.49
300 iter	1.06	0.96	0.91	1.13	0.84	0.79	1.87	0.67	0.61
1000 iter	1.09	0.97	0.91	1.20	0.88	0.80	1.86	0.74	0.63
R-L/Snyder									
100 iter	0.94	0.88	0.89	0.90	0.63	0.71	1.57	0.44	0.23
300 iter	1.06	0.97	0.96	1.12	0.88	0.89	1.83	0.63	0.67
1000 iter	1.09	0.98	0.96	1.21	0.93	0.90	1.90	0.78	0.74
MEM									
100 iter	0.96	0.87	0.86	0.94	0.61	0.57	1.71	0.42	0.19
200 iter	1.09	0.97	0.96	1.18	0.90	0.89	1.83	0.63	0.67
400 iter	1.09	0.98	0.97	1.24	0.95	0.92	1.96	0.80	0.76
Hybrid									
$\gamma = 0.1$	1.10	0.99	0.97	1.23	0.94	0.91	1.93	0.80	0.77
$\gamma = 0.01$	1.11	1.01	1.00	1.27	0.98	0.96	2.03	0.90	0.87
$\gamma = 0.001$	1.13	1.04	1.34	1.04	1.04	1.34	2.17	1.06	1.03

Table 4. Photometry errors, RMS errors from median ratio (percent).

Magnitude:	12 to 17.5			> 17.5 to 18.5			> 18.5		
Distance:	< 3	3 – 6	> 6	< 3	3 – 6	> 6	< 3	3 – 6	> 6
Minimum Norm									
$\gamma = 0.1$	157.7	243.2	21.5	197.7	176.5	63.4	211.9	459.7	244.8
$\gamma = 0.01$	173.0	77.7	3.8	402.3	76.8	11.8	467.1	607.4	53.2
$\gamma = 0.001$	157.6	6.6	3.2	402.4	15.5	9.3	548.6	79.9	31.1
Five iter. min. diff. from prev. iter.									
$\gamma = 0.1$	187.0	120.3	4.4	439.3	110.4	12.6	518.3	799.8	66.2
$\gamma = 0.01$	157.1	7.1	3.2	422.5	22.4	9.7	610.8	116.0	32.1
$\gamma = 0.001$	178.3	5.0	4.2	433.7	15.4	9.8	666.7	116.7	30.2
R-L Method									
100 iter	196.9	12.9	7.4	536.2	27.7	18.3	135.7	92.6	44.1
300 iter	195.0	9.8	8.1	470.3	17.7	15.3	728.9	37.6	35.8
1000 iter	195.3	9.4	8.3	448.2	14.7	15.5	749.2	37.4	33.8
R-L/Snyder									
100 iter	198.0	14.7	6.1	569.6	31.5	22.1	813.5	129.1	66.8
300 iter	194.6	9.3	2.8	475.0	19.1	8.9	747.8	41.1	38.3
1000 iter	194.3	8.4	3.1	445.0	12.9	7.7	735.4	36.0	28.2
MEM									
100 iter	183.5	9.2	6.6	541.5	25.4	23.2	737.5	142.5	57.4
200 iter	179.8	6.2	2.7	450.1	14.8	8.5	690.0	43.1	42.9
400 iter	186.8	3.6	2.7	430.9	11.3	7.7	692.6	37.6	28.2
Hybrid									
$\gamma = 0.1$	194.3	8.0	3.0	440.8	12.9	7.6	725.1	36.4	27.1
$\gamma = 0.01$	192.9	6.9	2.7	427.1	12.7	7.1	692.2	46.6	23.8
$\gamma = 0.001$	193.4	4.3	3.3	414.3	12.0	8.4	652.5	76.0	25.8

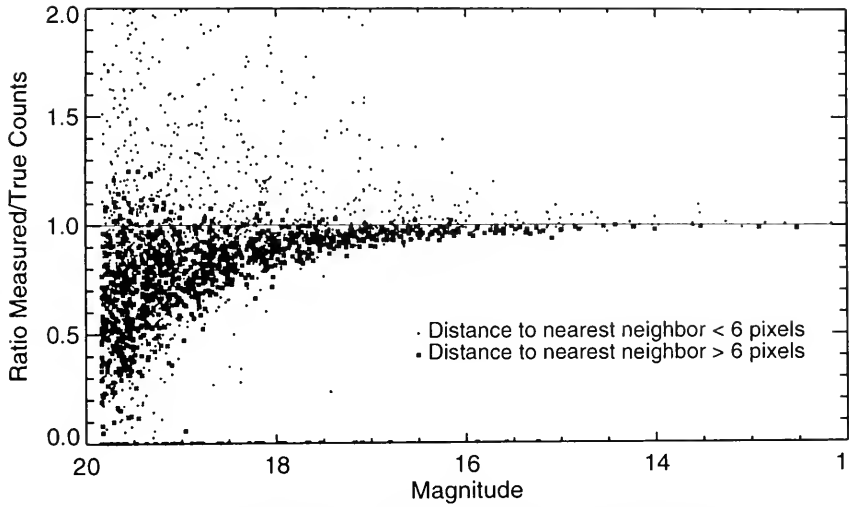


Figure 6. Photometry errors for the 1000 iteration R-L/Snyder restoration. The ratios of observed counts divided by actual counts are plotted versus the stars' magnitudes.

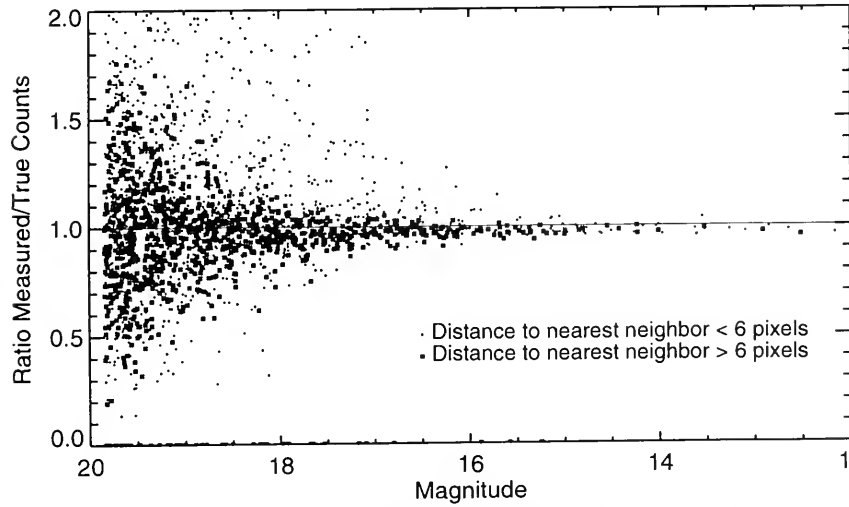


Figure 7. Photometry errors for the minimum norm restoration with $\gamma = 0.001$. The ratios of observed counts divided by actual counts are plotted versus the stars' magnitudes.

The Restoration of HST Images and Spectra II
Space Telescope Science Institute, 1994
R. J. Hanisch and R. L. White, eds.

Improved Photometry of HST Data With Iterative/Recursive Deconvolution Techniques

L. K. Fullton and B. W. Carney

Department of Physics and Astronomy, CB 3255, Phillips Hall, University of North Carolina at Chapel Hill, Chapel Hill, NC 27599-3255

K. A. Janes

Department of Astronomy, Boston University, 725 Commonwealth Ave., Boston, MA 02215

J. M. Coggins

Department of Computer Science, CB 3175, Sitterson Hall, University of North Carolina at Chapel Hill, Chapel Hill, NC 27599-3175

P. Seitzer

Department of Astronomy, 830 Dennison Bldg., University of Michigan, Ann Arbor, MI 48109-1090

Abstract. Image restoration results are presented using a new iterative/recursive method for removing a linear, spatially-invariant blur from an image. We have implemented this algorithm and used it to restore Hubble Space Telescope (HST) Planetary Camera images of the globular cluster NGC 6352. The resulting color-magnitude diagram illustrates the photometric accuracy which can be obtained from images deconvolved using our technique with the PSFs currently available from the Space Telescope Science Institute. If better PSFs become available, we believe the color-magnitude diagram could improve significantly. For comparison, we have analyzed the unrestored images with PSF-fitting photometry.

1. Algorithm

Iterative/recursive deconvolution is a new image restoration technique developed at the University of North Carolina. The iterative/recursive algorithm is linear, converges rapidly, conserves flux, produces no artifacts and is relatively insensitive to image noise or error in the PSF. It is described in Coggins, Fullton, & Carney (1994a) and in Coggins, Fullton and Carney (1994b, hereafter CFC). In simulations, images deconvolved with this algorithm produced linear photometric results. The intensity of point sources in the restored images was conserved to within 1–2% over a luminosity range of 6 dex (CFC; see also Busko 1994).

2. Observations

We deconvolved two pairs of HST Planetary Camera F555W and F785LP filter images of the globular cluster NGC 6352. The image scale is approximately 0.04 arcsec/pixel. We preprocessed the images applying the A/D correction, bias and preflash subtraction, dark current and flat field corrections. For both chips #7 and #8, three 300-second F555W exposures and four 1000-second F785LP images were averaged using the IRAF *imcombine* task with cosmic ray rejection to produce separate F555W and F785LP master frames for each chip. These master frames were deconvolved with the iterative/recursive algorithm invoked with 3 recursion levels and 2 iterations at each level.

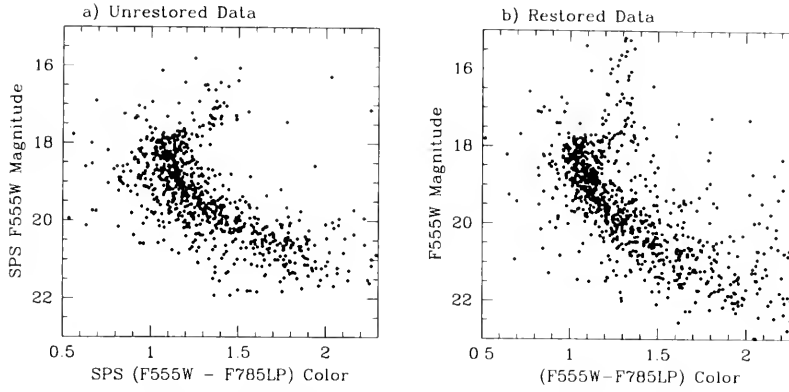


Figure 1. Color-magnitude diagram of NGC 6352. a) SPS Photometry of the unrestored images. b) Aperture photometry of the restored images.

Variation of the PSF across the images was accounted for by dividing the approximately 750×750 images into nine 300×300 overlapping sections and restoring each section independently with a PSF generated by the Tiny TIM software (Krist 1992) for the section center. The signal-to-noise ratio of the fainter stars in the deconvolved images was much improved over those in the originals. There were no artifacts in the images, and the residuals produced by the slight mismatch between the actual PSF and the PSF used in the restorations were smaller compared to the central stellar fluxes than the diffuse halos in the original data. As shown below, these residuals have only a minimal effect on photometry of the images.

3. Photometry

We performed photometry on both the original and deconvolved images of NGC 6352. The Stellar Photometry Software (SPS) developed by Janes and Heasley (1993) was used to perform PSF-fitting photometry on the original (blurry) data. For each master frame, a single spatially-constant PSF was generated from a large sample of bright stars scattered uniformly about the frame. This PSF was used to measure the magnitudes of all stars found using the SPS automatic star-finding algorithm. After this first pass, a few stars were added to the list from visual inspection, and the fits were repeated. The resulting color-magnitude diagram (CMD) is shown in Fig. 1a. The photometric zero-points have been adjusted to correspond to the F555W and F785LP magnitudes determined from the restored images from CCD #7.

Aperture photometry was performed on all stars identified in the deconvolved master frames with aperture radii of 4 and 5 pixels for the F555W and F785LP images respectively. Due to the larger diffraction-limited core radius, more power in the first diffraction ring and a poorer match between the Tiny TIM and observed PSFs at the longer wavelength, the stellar cores in the restored F785LP filter images were slightly larger than in the F555W images, and we were able to use a smaller aperture radius in the F555W images to help minimize crowding effects. Aperture corrections were determined from the brightest uncrowded stars in each image subsection and applied to the magnitudes of all the measured stars. The corrections for the CCD #7 images were 0.059 and 0.036 mag, F555W and F785LP, respectively. For the CCD# 8 images, the equivalent corrections were 0.044 and 0.080 mag.

The CMD for all stars whose nearest neighbor was greater than 5 pixels away is plotted Fig. 1b. Comparison with the CMD from the PSF fits to the unrestored images indicates that the main-

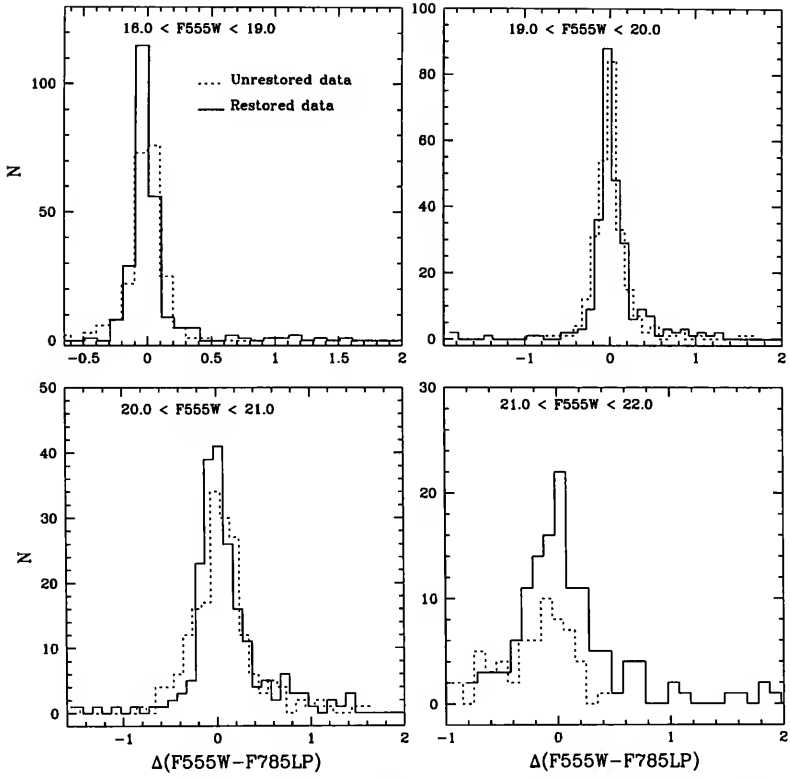


Figure 2. Histograms of the color difference between a star's position in the CMD and the mean main-sequence. The four histograms represent the data for stars in decreasing magnitude bins 1 magnitude wide.

sequence is better defined in the deconvolved data, especially at the faintest limit. Comparison of these results with those from our simulation (CFC) indicate that the precision of the NGC 6352 data is limited mainly by the mismatch between the actual PSF and that used in the deconvolution. We believe we could obtain even better photometric results from a PSF which matched the observed PSF more closely.

Fig. 2 illustrates the relative quality of the two different photometric reductions. We have plotted histograms of the color difference between a star's position in the CMD and the mean main-sequence defined by a parabolic fit to the data below the level of the main-sequence turn-off. Note that our deconvolution procedure allowed for the spatial variation of the PSF while the PSF-fitting software did not. Four histograms are plotted, which represent the data for main-sequence stars of decreasing brightness in 1-magnitude bins. The dotted lines represent the scatter about the mean sequence determined from the PSF fits to the unrestored data, and the solid lines indicate the scatter in the aperture photometry on the restored images. Since the largest aperture used was 5 pixels in radius, only data from stars whose nearest neighbor was more than 5 pixels distant are included in the figures to minimize scatter in the diagram due to the use of aperture photometry in crowded

fields. This scatter is obviously not due to the effects of the deconvolution algorithm but rather depends on the density of stars in the original image. Since the measurement aperture used was 5 pixels in radius, a nearest neighbor limit of 10 pixels would be necessary to completely remove the effects of crowding on the aperture photometry. In our experiments, we found that using a nearest neighbor criterion this strict removed an unacceptably large number of stars from our sample and did not produce a sequence noticably different from that shown in Fig. 2 otherwise.

With the profile fitting software, we were able to measure 727 stars fainter than 18th magnitude. Because of the higher signal-to-noise ratio in the deconvolved images and because the brightest stars and their environs were not measurable with SPS in the blurry data due to the large wings of the PSF, more stars were measured in both filters of the deconvolved data than in the original images, especially at the faint end. Thus, even after omitting stars with close neighbors from the histograms (about 15 percent of the total sample), there are a total of 834 stars represented in the histograms of the deconvolved data. Fig. 2 indicates that the scatter about the mean faint star sequence is much smaller for the aperture photometry of the deconvolved images than for the PSF fits to the original images. This effect is small at the brightest and faintest magnitudes but is particularly evident in the magnitude range $20 \leq \mathrm{F555W} \leq 21$.

Since NGC 6352 lies toward the Galactic center ($l = 341°, b = -7°$), even perfect photometry of the cluster images will produce CMDs with residual scatter due to the large field star contamination from the Galactic Bulge. Thus while some of the scatter in Figs. 1 and 2 is obviously attributable to photometric error, many of the larger deviations from the cluster sequence are likely due to this field contamination. Quantification of the extent of contamination in the CMD awaits analysis of our off-cluster field images. While a more extensive comparison between the photometric precision possible from the restored vs. unrestored cluster images in the absense of field star contamination would have been possible with simulated data, we believe the comparison between the results for actual data is more relevant to evaluation of the usefulness of the restoration procedure.

4. Conclusions

Comparison of the scatter in the CMDs of the unrestored and restored images as illustrated in Figs. 1 and 2 indicates that iterative/recursive deconvolution followed by aperture photometry produces photometrically superior results to those of the SPS PSF-fitting software for data from HST in its current configuration. From comparison of the scatter in the diagrams for the brightest stars, it appears that there is little to be gained by deconvolution where the signal-to-noise ratio in the unrestored data is high. However, this conclusion is based on the comparison between aperture and PSF-fitting photometry. The application of least-squares type PSF-fitting algorithms to the restored images should further improve the photometric scatter (analagous to the improvement in ground-based data between aperture and PSF-fitting techniques). We believe that further work is necessary to develop least-squares fitting techniques which are appropriate for the image and noise properties of deconvolved HST images.

References

Busko, I. 1994, this volume

Coggins, J. M., Fullton, L. K., & Carney, B. W. 1994a, this volume

Coggins, J. M., Fullton, L. K., & Carney, B. W. 1994b, Comp. Vis. Graph. & Image Proc., submitted

Janes, K. A., & Heasley, J. N. 1993, PASP, 105, 527

Krist, J. 1992, "The Tiny TIM User's Manual," Space Telescope Science Institute, Baltimore

The Restoration of HST Images and Spectra II
Space Telescope Science Institute, 1994
R. J. Hanisch and R. L. White, eds.

Assessment of Methods Used for HST Image Reconstruction

M. Bertero

Dipartimento di Fisica dell'Universitá di Genova, Via Dodecaneso 33, I-16146 Genova, Italy

F. Maggio

CRS4, Via N. Sauro, 10, I-09123 Cagliari, Italy

E.R. Pike and D. Fish

Department of Physics, King's College, London WC2R 2LS, England

Abstract. An assessment of the performance of various inversion algorithms is presented and compared with the Richardson-Lucy method, the currently favored approach used for the restoration of Hubble Space Telescope images. Numerical validations are made in the case of two simulated images.

1. Introduction

As is well known the Hubble Space Telescope (HST) suffers from, amongst other defects, severe spherical aberration which takes light away from the central core of the Point Spread Function (PSF). A partial correction of the aberration effects can be attempted since the PSF of the instrument is known. One knows the linear integral relationship between the object and the image and therefore one can try to restore the object by solving this equation. This is a Fredholm equation of the first kind, which is a classical example of an ill-posed problem, namely a problem whose solution is affected by numerical instability and therefore is strongly noise dependent. Approximate and stable solutions can be obtained using suitable techniques (regularized inversion methods). The accuracy of these solutions depends essentially on the signal-to-noise ratio. For the HST, this is not very high.

The isoplanatic (deconvolution) problem, assuming the scalar diffraction theory which applies well to these low-aperture systems, reduces to one of inverting the first-kind Fredholm convolution integral equation

$$g(\mathbf{x}) = (Af)(\mathbf{y}) = \int f(\mathbf{y})\, H(\mathbf{x}-\mathbf{y})\, d\mathbf{y}\,, \tag{1}$$

where g is the image from some distribution f via the integral operator A. H represents the PSF of the telescope. The problem is to estimate f for a given g, assuming H is known.

Current efforts at reconstructing aberration-degraded images are, in the main, based upon conventional Fourier methods such as iterative deconvolution (here we use the term deconvolution in this more restricted sense rather than the more general term "reconstruction" which it sometimes erroneously replaces) as well as other techniques such as Maximum Entropy (ME) (Gull and Skilling, 1984) and so on. The favored data inversion method currently employed by the astronomy community is that of the Richardson-Lucy iterative deconvolution where the f_{k+1}'th iteration approximating the solution is given by (Richardson 1972, Lucy 1974)

$$f_{k+1} = f_k \left(H^T * \frac{g}{H * f_k} \right)\,, \tag{2}$$

where the symbol $*$ denotes convolution. It may be seen that this method retains the positivity expected of incoherent-image data since $g \geq 0$ and also $H \geq 0$. This is not necessarily true of other matrix and projection methods which can give negative values also in the case of a noiseless image. In the case of iterative methods, positivity may be imposed at each iteration stage or simply at the end of the reconstruction, depending on the method used and, in our work, was not found to limit the efficiency of such routines. Deconvolution techniques become suspect when the transfer function (TF) (i.e., the Fourier transform of the PSF) $\hat{H}(\omega)$ becomes small, as noise on the image in this region would result in anomalous reconstruction, exaggerated by the vanishing modulus of the TF. Thus, for conventional techniques that do not have an adequate stopping criterion, small amounts of error on the data manifests itself as a much greater error on the solution. This is the source of the well known ill-conditioning associated with the inversion of such integral equations. In §2 we briefly describe the inversion methods we have used and in §3 we give the numerical results we have obtained.

2. Description of the Reconstruction Methods

A quite natural way of discretizing Eq. (1) is as follows:

$$g(x_{ij}) = \sum_{lm} H(x_{ij} - x'_{lm}) f(x'_{lm}) \Delta_{lm} \,, \tag{3}$$

where $g(x_{ij})$ $(i, j = 0, ..., N-1)$ and $f(x'_{lm})$ $(l, m = 0, ..., N-1)$ are the $N \times N$ image and object, respectively; the Δ_{lm} are quadrature weights. Eq. (3) can also be written in the form $g = Af$ where now A is a $N^2 \times N^2$ matrix. It can be diagonalized by the Discrete Fourier Transform (DFT) and becomes

$$\hat{g}(\omega_{ij}) = \hat{H}(\omega_{ij}) \hat{f}(\omega_{ij}) \tag{4}$$

where $\hat{g}$, $\hat{H}$, and $\hat{f}$ are the DFT's of g, H, and f, respectively. The matrix A is, in general, ill-conditioned and therefore the solution of (3) $f = A^{-1}g$, as well as the solution of (4),

$$\hat{f}(\omega_{ij}) = \frac{\hat{g}(\omega_{ij})}{\hat{H}(\omega_{ij})} \,,$$

are affected by numerical instability. Then one has to use regularized inversion methods in order to obtain approximate and sensible solutions. The methods we have considered are the following:

- (a) Tikhonov regularization (TIKH),
- (b) Landweber iteration (LW),
- (c) Landweber iteration with imposed positivity (LWPOS),
- (d) Conjugate gradients (CG),
- (e) Conjugate gradients with imposed positivity (CGPOS), and
- (f) Richardson-Lucy method (R-L).

All these methods contain a free parameter which is the so-called regularization parameter; in the case of method (a) this is a smoothing of the higher spatial frequencies; for the methods $(b) - (f)$, it is the number of iterations. In the case of a simulated experiment, the typical behavior of the approximate solutions as a function of the free parameter is the following: if we compute the distance between the approximate solution and the known true solution (for instance the Euclidean norm of their difference) and we consider its behavior as a function of the free parameter, then this distance will exhibit a minimum. The value of the parameter corresponding to this minimum can be considered as the optimum one since it provides the best approximation to the solution of the problem. Obviously, in the case of real data, the solution of the problem is not known and therefore

this method can not be used for the estimation of the free parameter. For this reason several criteria have been investigated for its choice (Bertero 1988). However, in general they are not very practical and very often the choice they provide is not satisfactory, so that one has to use some empirical criterion which depends on the problem one is investigating. In the case of simulated data, we will choose the value of the parameter corresponding to the minimum of the distance between the approximate and the true solution.

For the convenience of the reader we recall that the approximate solution given by method (*a*) is

$$f_\mu = (A^* A + \mu I)^{-1} A^* g \tag{5}$$

where μ is the regularization parameter. More explicitly in the case of Eq. (4) we have

$$\hat{f}_\mu(\omega_{ij}) = \frac{\hat{H}^*(\omega_{ij})}{|\hat{H}(\omega_{ij})|^2 + \mu} \hat{g}(\omega_{ij}) \,. \tag{6}$$

In the case of method (*b*) the iteration scheme is given by

$$f_0 = 0; \quad f_{n+1} = f_n - \tau r_n \tag{7}$$

where $0 < \tau < 2 \|A\|^{-2}$ is the stationary relaxation parameter and

$$r_n = A^* A f_n - A^* g \,. \tag{8}$$

We point out that this iterative method is equivalent to a filtering since the final result of N iterations can be explicitly written and in Fourier space it is given by

$$\hat{f}_n(\omega_{ij}) = \left[1 - \left(1 - \tau |\hat{H}(\omega_{ij})|^2\right)^n\right] \frac{\hat{g}(\omega_{ij})}{\hat{H}(\omega_{ij})} \,. \tag{9}$$

This expression cannot be used in method (*c*) because the negative part of the reconstruction of f_n obtained from (7) must be zeroed at the end of each iteration stage. Since r_n must be computed in Fourier space at each iteration, two FFTs are required: the first for computing $\hat{r}_n$; the second one for coming back to the physical space in order to impose positivity. On the practical side, for both methods (*b*) and (*c*) the relaxation parameter τ can be set as follows:

$$\tau = \frac{1}{\| A \|^2} = \frac{1}{|\hat{H}|^2_{max}} \,, \tag{10}$$

where $|\hat{H}|_{max} = \max_{i,j} |\hat{H}(\omega_{ij})|$.

Finally, the conjugate gradients iteration scheme is

$$f_0 = 0; \quad f_{n+1} = f_n - \tau_n p_n \tag{11}$$

where

$$p_n = r_n + \sigma_{n-1} p_{n-1} \,. \tag{12}$$

with $p_0 = r_0 = -A^* g$, $\sigma_{n-1} = \|r_n\|^2 / \|r_{n-1}\|^2$ and the non-stationary relaxation parameter $\tau_n = (r_n, p_n) / \|A p_n\|^2$. This iteration scheme can also be easily written in Fourier space.

3. Simulated Results with the HST

In our numerical experiments we have mainly considered two test objects provided by R. Hanisch and characterized by different types of brightness distribution; one of the objects f simulates an elliptical galaxy and the other one a star cluster with size 512×512 and 256×256 pixels, respectively.

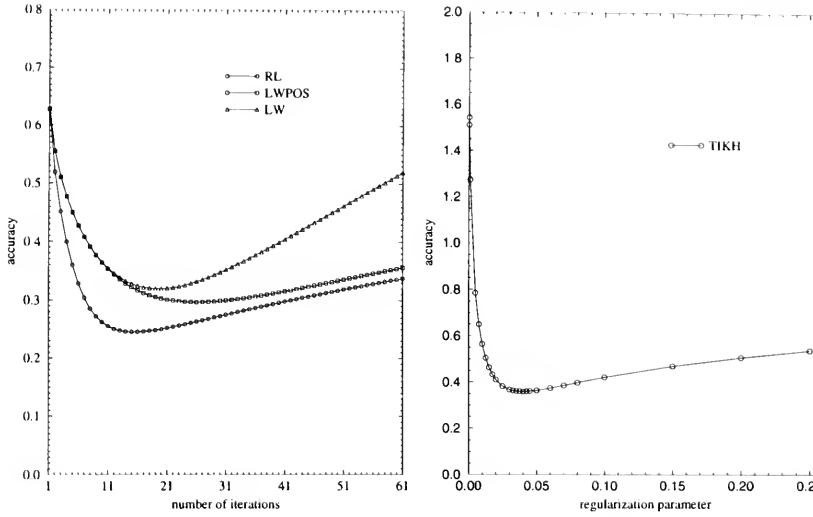


Figure 1. Accuracies (ϵ from Eq. 13) for the elliptical galaxy achieved with iterative methods (left) and with the Tikhonov method (right).

Simulated blurred images are obtained from the object f convolved with a computed PSF, with the addition of Poissonian noise and readout noise (Hanisch 1993). The discrepancy between the reconstructed solution $f^{\rm rec}$ and the known solution f may be measured by the RMS norm ϵ of their difference divided by the norm of f (here $f_{ji} = f(x_{ij})$)

$$\epsilon = \sqrt{\frac{\sum_{ij} |f_{ij}^{\rm rec} - f_{ij}|^2}{\sum_{ij} |f_{ij}|^2}} \ . \tag{13}$$

Another possible way to evaluate algorithm performance is the so called Mean Square Error Improvement Factor (MSEIF) which reads (Busko 1993)

$$\eta = 10 \log_{10} \frac{\sum_{ij} |g_{ij} - f_{ij}|^2}{\sum_{ij} |f_{ij}^{\rm rec} - f_{ij}|^2} \ ; \tag{14}$$

in the sequel we use both. By varying the regularization parameter, it is hoped to determine a minimum of ϵ (Eq. 13) or a maximum of η (Eq. 14). An example is given in Fig. 1, where this has been done for the case of the elliptical galaxy. We now give our results in the comparison of the various methods and we consider first the case of the star cluster.

It is worthwhile to introduce here the condition number of the matrix A, which gives an useful estimate of the error propagation:

$$\frac{\| \delta f \|}{\| f \|} \leq {\rm cond}(A) \frac{\| \delta g \|}{\| g \|} \ , \tag{15}$$

where δg is the error on the blurred image (Poisson + read-out) and δf is the consequent error on the restored image (this simple two-norm condition number doesn't take into account the error on

the PSF). The condition number (Eq. 15) can be simply evaluated as follows:

$$\text{cond}(A) = \frac{|\widehat{H}(\omega)|_{max}}{|\widehat{H}(\omega)|_{min}}\,, \tag{16}$$

where

$$|\widehat{H}(\omega)|_{max} = \max_{\omega_{ij}}\{|\widehat{H}(\omega)|\} \qquad \text{and} \qquad |\widehat{H}(\omega)|_{min} = \min_{\omega_{ij}}\{|\widehat{H}(\omega)|\}\,.$$

For the PSF of the star cluster we have

$$\text{cond}(A) = 23.2764 \tag{17}$$

that sounds exceedingly low for people working in the field of inverse and ill-posed problems; consequently one expects the discrepancy (13) to be rather small at the minimum. The best performances in the restoration of the star cluster image are provided by the R-L algorithm and the LWPOS one; they give accuracies of 6% and 5%, respectively (see Table 1). This is clearly an excellent mathematical result but not exceedingly satisfactory on the practical side as many false stars appear when the cluster's core is examined in details; this is shown in Fig. 2. This is partially due to the fact that the solution itself is represented by discrete points, whereas the image and reconstructed solution are blurred representations thereof. In fact the choice of such solutions are particularly demanding for the restoration routines, since they are delta-function-like objects with sharp cut-offs.

Table 1. Algorithm comparison. Analysis performed on star cluster (256×256).

Algorithm	# Iterations	RMS	MSEIF	CPU Time [a]
R-L	1006	0.0606	23.5642	21 min 18 sec
LWPOS	1038	0.0484	25.5218	12 min 58 sec
LW	direct method	0.1020	19.0445	2 sec
CG	110	0.1355	16.5777	2 min 5 sec
CGPOS	100	0.1027	18.9851	1 min 52 sec
TIKH	direct method	0.1115	18.2710	1 sec

[a] IBM RISC 6000/560 with ESSL

As a matter of fact, the LWPOS method seems to perform better. This is highlighted from the analysis of the computational efforts (see last column of Table 1): we recall that LWPOS just needs two FFTs per iteration while R-L needs four. Other methods examined seem to be very fast at the expenses of a rather poor accuracy (all the RMSs are greater than 10%).

Let us turn our attention to the case of the elliptical galaxy. Once more LWPOS gets the best performance among the regularization methods but this time the R-L algorithm seems to work better, both in terms of convergence rate and accuracy (see Fig. 1 and Table 2).

For the elliptical galaxy we have

$$\text{cond}(A) = 18.2002\,, \tag{18}$$

a value not too different from (17). This means that the lower accuracy, with respect to the star cluster, my be imputed to a lower signal-to-noise ratio

$$S/N = \frac{\| k * f \|}{\| g - k * f \|} \tag{19}$$

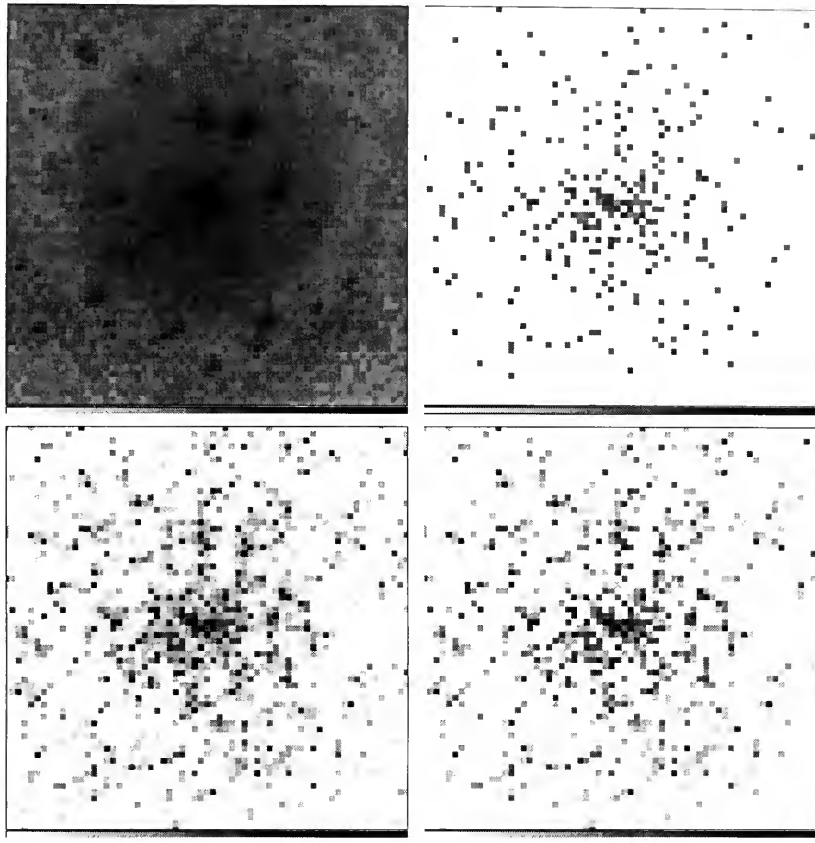


Figure 2. Detailed views of the star cluster. The simulated data (top left) is compared with the true image (top right) and with the two restored images: the R-L solution (bottom left) and the LWPOS solution (bottom right). Many false stars can be observed in both the restored images. The grey inverted scale is logarithmic in intensity.

Table 2. Algorithm comparison. Analysis performed on elliptical galaxy (512×512).

Algorithm	# Iterations	RMS	MSEIF	CPU Time [a]
R-L	15	0.2453	14.6068	1 min 30 sec
LWPOS	26	0.2967	10.8007	1 min 34 sec
LW	direct method	0.3146	9.6263	2 sec
CG	1	0.4879	0.8508	15 sec
CGPOS	1	0.4879	0.8508	15 sec
TIKH	direct method	0.3592	6.9753	6 sec

[a] IBM RISC 6000/560 with ESSL

rather than to a worse conditioning of the problem. Indeed we found, in the two cases, that

$$S/N_{\text{star cluster}} = 14.2857 \quad \text{and} \quad S/N_{\text{elliptical galaxy}} = 3.9841 \ .$$

Furthermore it is not clear how important the main features of these images (brightness, shape, etc.) are in restoring our noisy data. It seems evident that for this test object the R-L algorithm provides the best restoration. Fig. 3 shows the comparison between R-L and LWPOS results.

4. Conclusions and Perspectives

The mathematical problem arising from HST image restoration provides an example of an inverse problem which can be fruitfully treated by regularization techniques. Among these methods, LWPOS seems to provide the best compromise between accuracy and computational effort. On the other hand it is not evident, from the numerical experiments performed in this work, which is the best choice between LWPOS and the well-known R-L method: a larger set of test problems, covering a wide range of astronomical objects, has to be taken into account.

The forthcoming correction of the imaging disorders with the installation of COSTAR and WFPC II should restore near perfect imaging to the HST. However, the possibility of reconstructing even diffraction limited images is attractive since some super-resolution will be possible (Bertero et al. 1989). For this reason, restoration techniques able to provide good results coupled with a tolerable computational cost will be greatly appreciated even after the first servicing mission of December 1993. Non-isoplanatic corrections will be highly desirable.

In this context we cannot forget that the R-L method, the most popular algorithm used for the restoration up to now, is efficient at the further expenses of a computational overload since there are four FFT operations per iteration, that makes it easily the most expensive in computer time of all the methods apart from the Maximum Entropy method.

Acknowledgments. The work of F. Maggio has been supported by Sardinian Regional Authorities. The work at King's College has been supported by US Army grant DAAL03–92–G–0142.

References

Bertero, M., 1988, in *Advances in Electronics and Electron Physics*, Academic Press, New York, 75, 1

Bertero, M., Boccacci, P., Davies, R.E. & Pike, E.R. 1989, Inverse Problems, 5, 441

Busko, I. 1993, in *Restoration Newsletter*, R. Hanisch, ed., Space Telescope Science Institute, Baltimore, 1, 45

Engineering and Scientific Subroutine Library Version 2, IBM

Gull, S. F. & Skilling, J. 1984, IEEE Proc., 131, 646

Hanisch, R. 1993, in *Restoration Newsletter*, R. Hanisch, ed., Space Telescope Science Institute, Baltimore, 1, 76

Lucy, L. B. 1974, AJ, 79, 745

Richardson, W. H. 1972, J. Opt. Soc. Am., 62, 55

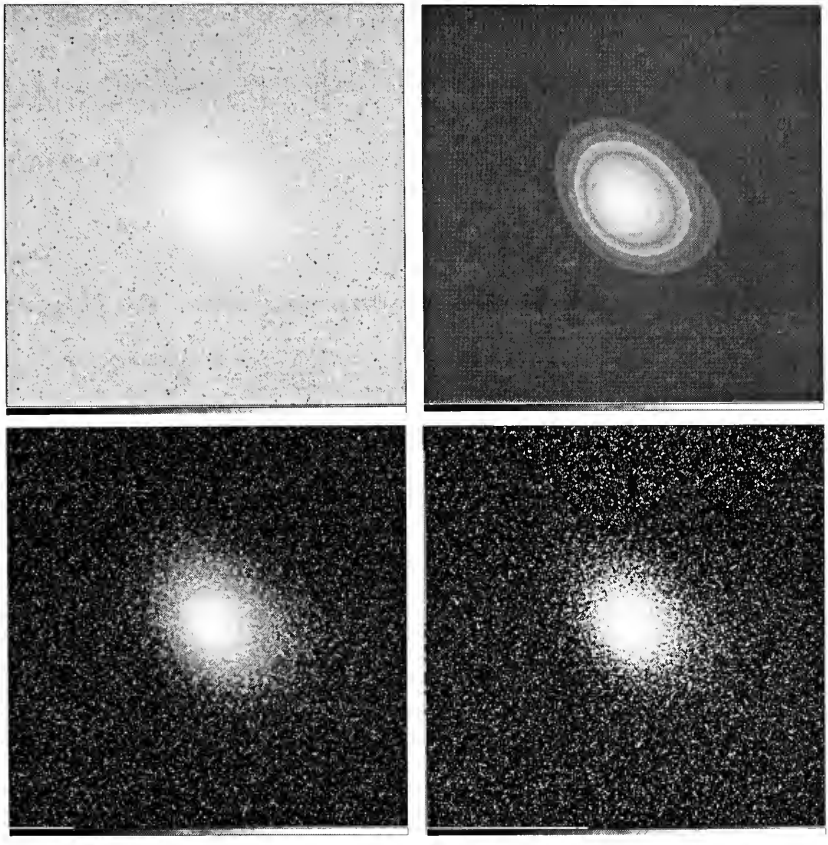


Figure 3. Restoration results for the elliptical galaxy. The simulated data (top left) is compared with the true image (top right) and with the two restored images: the R-L solution (bottom left) and the LWPOS solution (bottom right). The grey scale is logarithmic in intensity.

The Restoration of HST Images and Spectra II
Space Telescope Science Institute, 1994
R. J. Hanisch and R. L. White, eds.

DAOPHOT Reductions of the Simulated Cluster Field

Peter B. Stetson
Dominion Astrophysical Observatory, Herzberg Institute of Astrophysics, National Research Council of Canada, 5071 West Saanich Road, Victoria, British Columbia V8X 4M6

Abstract.
I present the results of DAOPHOT reductions of the simulated star cluster images provided by ST ScI's Image Restoration Project.

1. Introduction

Bob Hanisch called me on the phone and asked me to drop whatever I was doing and reduce his simulated cluster field. My reward would be an opportunity to speak at a workshop held at the Space Telescope Science Institute. Second prize was two workshops.

2. Conditions of the Test

Since I had only one image to analyze, reductions were performed with the DAOPHOT II generation of ALLSTAR (see Stetson & Harris 1988; Stetson, Davis, & Crabtree 1990; Stetson 1994a). I tried reducing the field with five different model point-spread functions (PSFs): (1) a spatially invariant PSF derived from ten bright stars in the test image itself; (2) a quadratically varying PSF derived from the NGC 1850 frames discussed by Stetson (1994b); and (3) spatially invariant, (4) quadratically varying, and (5) cubically varying PSFs derived from grids of model profiles generated with Tiny TIM by Shaun M. Hughes. These Tiny TIM simulations were for a different focus setting and jitter history than assumed in the simulation. For each of (2)–(5), I added the 256×256 simulation back into the middle of an 800×800 blank image, so that the variations in the PSF would be correctly mapped into the frame.

Actually, there were two tests: one where the stars were exactly centered in their respective central pixels, and one where stars could effectively fall anywhere with respect to the centers of pixels. I analyzed these separately. I used the "truth" files to provide initial estimates of the stars' positions, but *not* of their magnitudes. This is fair because in most cases of interest the astronomer would have multiple images of a given star field, and would be able to generate a more complete list of detections than could be derived from any one frame alone.

3. Figure of Merit

Hanisch provided me with the "right" answers for the stars in the simulated cluster, so it's easy to determine the error of each magnitude determination: find the star in the truth file that falls nearest to the position of each star in my photometric output, and compute the difference between their magnitudes. The question of how to encapsulate these magnitude differences in an overall figure of merit for the reduction procedure deserves some thought.

The root-mean-square residual is the theoretician's dream for parameterizing the width of a probability distribution. It can be shown from first principles that, subject to a few assumptions (such as that the total errors consist of an infinite number of infinitesimal contributions), the error

distribution is driven in the limit to the Gaussian distribution characterized by two parameters: the mean (which I set identically to zero by forcing my magnitude zero-point to agree with that in the simulations), and the standard deviation. The root-mean-square deviation can be shown from first principles to be the most "efficient" estimator of the standard deviation.

That doesn't necessarily mean that the root-mean-square residual is the best figure of merit for our present purposes. In real life, photometric errors are not always made up of an infinite number of infinitesimal contributions: detector blemishes, cosmic rays, satellite trails, and any number of other problems can represent single, discrete, macroscopic error sources. Every observational astronomer knows that three-sigma residuals occur far more often than the Gaussian formula predicts. Even in the simulated images, two stars in adjacent pixels will probably be perceived by the software as a single, brighter star lying between the pixels. Then, when the output photometry is compared with the "truth," the measured magnitude will be compared with one of the two input magnitudes, and a large photometric error will be inferred. Is this a fault of the photometric algorithm? Or of the subsequent cross-identification procedure? Or is it simply a fact of life that sometimes astronomers have to be satisfied with composite photometry for unresolved doubles? Since these occasional major mistakes are squared before they go into the root-mean-square estimator of the standard deviation, they overwhelm the majority of smaller residuals. In a case like the present one, all the root-mean-square residual will tell you is whether in this particular data set your largest screw-up is a six-sigma blunder or a seven-sigma howler — it won't tell you anything about the bulk of the data.

Examining the formula for the Gaussian distribution, one can come up with several other ways of estimating the standard deviation:

- The mean absolute deviation can be calculated:

$$\begin{aligned} \langle |x| \rangle &= \frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^{+\infty} |x| \exp(-x^2/2\sigma^2)\, dx \\ &= \sqrt{2/\pi}\,\sigma \end{aligned}$$

  whence $\sigma = 1.2533 \times$ (mean absolute residual).

- In the cumulative Gaussian function, $N(-\sigma) = 0.1587$ and $N(+\sigma) = 0.8413$, whence $\sigma \approx$ half the difference between the 84$^{\text{th}}$ and 16$^{\text{th}}$ percentiles.

- In the cumulative Gaussian function, $N(-0.6745\sigma) = 0.25$ and $N(+0.6745\sigma) = 0.75$, whence $\sigma = 0.7418$ times the interquartile range.

- In the cumulative Gaussian function, $N(-\sigma/2) = 0.3085$ and $N(+\sigma/2) = 0.6915$, whence $\sigma \approx$ the difference between the 69$^{\text{th}}$ and 31$^{\text{st}}$ percentiles.

Obviously, there exists an infinity of such statistics, each one capable of providing an estimate of exactly the same Gaussian σ that the root-mean-square deviation is intended to measure. These alternatives are less efficient than the root-mean-square when operating on the theoretician's idealized continuous distribution, but they may be preferable when it comes to dealing with real data. Often when an astronomer is trying to determine the location of the principal sequences of some star cluster in the color-magnitude diagram, the goal is to obtain the sharpest possible sequence for the bulk of the stars, regardless of the few oddballs that may be scattered by photometric blunders to remote regions of the CMD. The five alternative σ estimators that I have listed above, beginning with the r.m.s., proceed in order from those that are most influenced by the far tail of the error distribution, to those that measure primarily the width of the core of that distribution. Lest I later be accused of fraud, let me say right now that for the remainder of this paper I use the last estimator: $\sigma \approx$ the difference between the 69$^{\text{th}}$ and 31$^{\text{st}}$ weighted percentiles.

What do I mean by a *weighted percentile*? Imagine that for each observation I take a stick of wood and cut it to a length proportional to the weight of that observation (weight = the inverse

square of the standard error as estimated from the readout noise, the photon noise, and the goodness of the profile fit). On each stick I write the value of that observation. Then I sort the sticks in increasing order of the values written on them and arrange them into a continuous bar. Then I point at positions 31% and 69% of the way along the total length of the bar and read off the two observational values that I'm pointing at. The difference between them is my estimate of σ.

This figure of merit is obviously dominated by the highest-weight observations, i.e., the brightest stars. I think that is legitimate. The purpose of this exercise is to test the quality of the photometric software. The errors of the faintest stars are dominated by their Poisson and read noise, not by any failures of the algorithm. If I were to generate an unweighted estimate of σ from the sample as a whole, I could practically make my results come out as good as I want: all I'd have to do is manage to find fewer and fewer of the faint stars, and the σ generated from the stars that I did find would get better and better.

The figure of merit used here is reasonable, in that it tests the quality of the photometric algorithm by seeing what it can do with the best data rather than measuring the amount of synthetic noise added to the many faint stars; and it is impartial, in that it is difficult for me to produce a better or poorer figure of merit by changing which stars I want to consider "recovered" as opposed to "lost." In short, it measures how well the algorithm does with a typical, well-exposed star.

4. Results

As mentioned above, two simulated images (stars centered on pixels, stars not necessarily centered on pixels) were reduced with five point-spread functions each (a PSF was obtained from each of the simulated "science frames," one was obtained from actual WFC images, and three were obtained from Tiny TIM simulations). In each case I found that the PSFs obtained from the simulations themselves yielded the most accurate photometry. Tables 1 and 2 present actual values of σ_{69-31} derived using the PSFs from the synthetic frames themselves, for four subsets: (a) all stars; (b) all stars except the faintest two magnitudes (true < 2.00); (c) all stars more than 28 pixels from the center of the cluster; and (d) all stars brighter than true = 2.00 inside 28 pixels from the center of the cluster.

Sample	σ_{69-31}	Fraction recovered
All	0.062	357/470
true < 2.00	0.054	118/130
r > 28 px	0.060	246/247
r < 28 px, true < 2.00	0.179	54/67

Table 1. Stars centered in pixels.

Sample	σ_{69-31}	Fraction recovered
All	0.051	299/500
true < 2.00	0.038	79/108
r > 28 px	0.048	246/250
r < 28 px, true < 2.00	0.306	26/56

Table 2. Stars not necessarily centered in pixels.

ALLSTAR does not seem to have a big problem with undersampling: with one exception, the results are not worse when the stars are allowed to lie anywhere within their pixels. And I believe

that exception may be more an artifact of the simulation than of the reduction. In the process of generating the first test image, whenever two artificial stars had their centers in the same pixel, one was thrown away and only one was kept. In the latter simulation, any number of stars falling in the same pixel were all retained. Subsequently, the reduction software was unable to distinguish the numerous stars lying in a single pixel, and reduced the blend as one object. The cross-identification software then associated this object with only one of the artificial stars in the "truth" list, thereby implying a lower recovery rate and a higher typical photometric error.

References

Stetson, P. B. 1994a, submitted to PASP

Stetson, P. B. 1994b, in *Proceedings of HST Calibration Workshop*, C. Blades, ed., Space Telescope Science Institute, Baltimore

Stetson, P. B., Davis, L. E., & Crabtree, D. R. 1990, in *CCDs in Astronomy*, G. H. Jacoby, ed., ASP Conference Series, 8, 289

Stetson, P. B., & Harris, W. E. 1988, AJ, 96, 909

The Restoration of HST Images and Spectra II
Space Telescope Science Institute, 1994
R. J. Hanisch and R. L. White, eds.

Precision Photometry at the LMC Center: Simulating Post-COSTAR HST Observations

P. Linde, R. Snel, and S. Spännare

Lund Observatory, Box 43, S-221 00, Lund, Sweden

Abstract. We intend to use the restored HST to study age structure and chemical evolution of the stellar population at the center of the LMC. To obtain the necessary observational parameters, Strömgren photometry is planned using the Planetary Camera. To study the reduction techniques needed, a set of simulated PC images have been made. The stellar density in the images is based on an extrapolated luminosity function and show a high degree of crowding. Two of the generally available photometric software packages, ROMAFOT and DAOPHOT, have been used to analyse the images. Results show that the photometric accuracy reached with these packages is not enough for high precision photometry in this configuration. We show that the two major contributors to the errors are undersampling and the unusual shape of the point spread function. Adaptation of locally developed software is currently in progress to deal with these problems.

1. Introduction

Large scale processes driving and controlling star formation and nucleosynthesis are fundamental but, in many cases, particularly for ordinary stars, poorly known and understood. Due to its relative proximity, the Large Magellanic Cloud (LMC) offers unique possibilities of detailed observations providing that sufficient image resolution is available. A group of Scandinavian astronomers intends to use the Hubble Space Telescope (HST) to make a detailed study of stellar populations down to magnitude V = 22 (F–dwarfs), at the LMC center. Age structure and chemical composition will be determined by performing Strömgren photometry using the Planetary Camera (PC).

To assist in the selection of suitable target areas, ESO Schmidt plates have been scanned. Fig. 1 shows an 11×11 arcmin region at the center of the LMC. The square in the figure marks a 35×35 arcsecond area that has been chosen as one HST target.

2. A PC Simulation

To thoroughly study the reduction problems involved, simulated PC images have been made. The point spread functions (PSFs) expected after optical correction were derived for each filter using version 2.4 of the Tiny Tim software (Krist, 1993). The images were initially created with four times oversampling, then scaled down to 800×800 pixels, and finally background and noise were added. These were estimated, following the recommendations in the WFPC II Handbook, to correspond to a signal to noise ratio of about 100 at magnitude V = 21.0. The star density was derived using an extrapolated luminosity function determined from earlier observations (Ardeberg et al. 1985), where a field in the LMC bar was studied in detail. An example of a simulated image is given in Fig. 2, which represents the field seen through the Strömgren y filter (F547M). This image contains about 40,000 stars down to magnitude V = 29. Fig. 3 shows an expanded part from Fig. 2, and in Fig. 4 the Tiny Tim PSF is seen, both with and without oversampling. The FWHM of this y filter PSF is 0.9 pixels. The irregular shape of the poorly sampled stellar images is evident as well as "fixed pattern noise," resulting from the uneven, patchy, wings of the PSF.

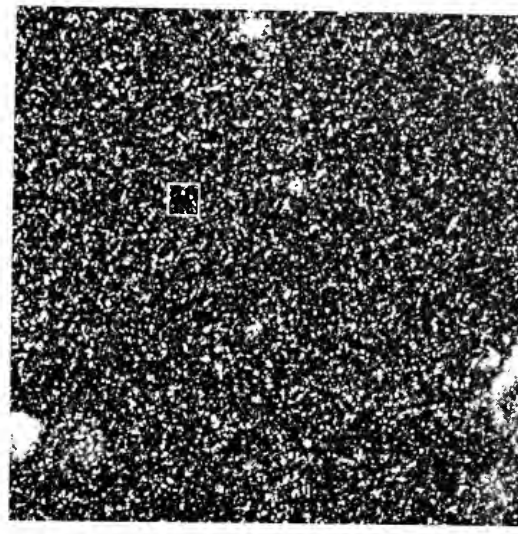


Figure 1. A 1024×1024 pixel scan from an ESO Schmidt plate, showing an area of 11×11 arcmin at the center of the LMC. The square marks the region selected for the HST study.

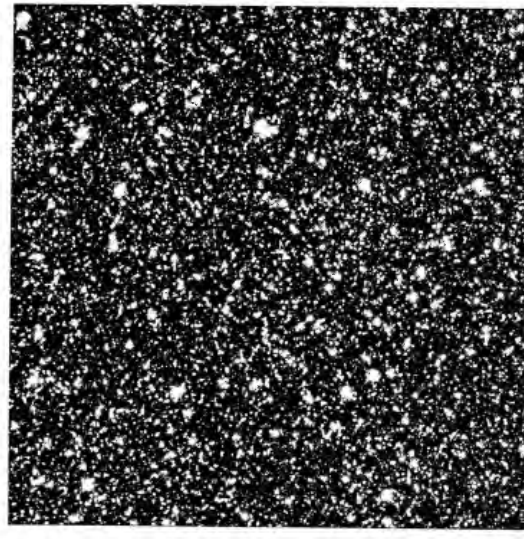


Figure 2. A simulation of the selected area as seen with the PC through the Strömgren y filter. About 40,000 stars down to V = 29 are included.

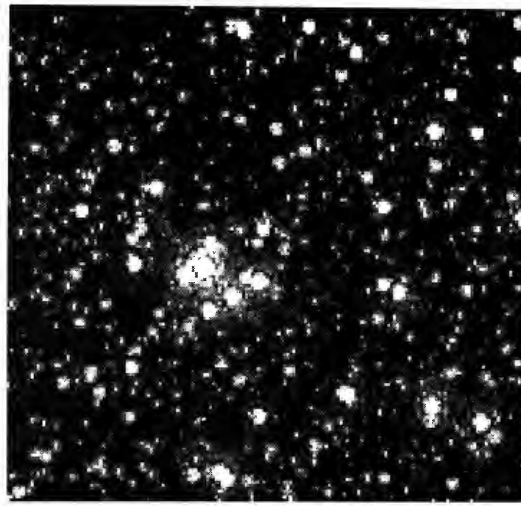


Figure 3. Expanded view of a region from Fig. 2.

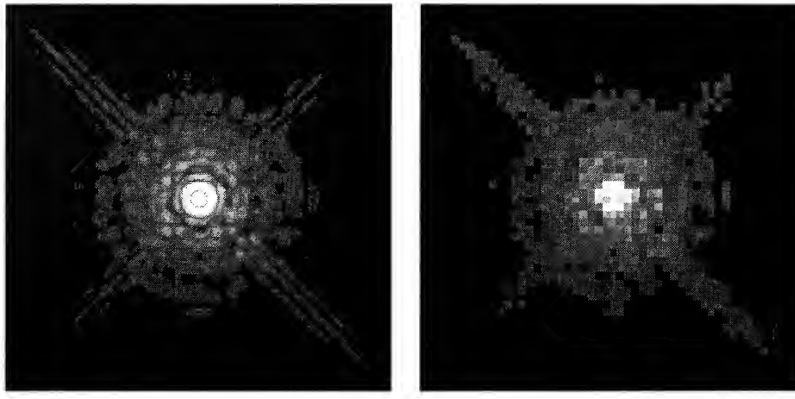


Figure 4. The Tiny Tim PSF (y filter, F547M) used for generating the image in Fig. 2. To the left (a) the PSF is shown with four times oversampling, to the right (b) with normal sampling. The intensity scale is logarithmic, in order to also show the outer parts of the PSF.

In addition to the original images, several modified versions were made to test the response of the photometric programs:

- To test for PSF shape dependence, images were generated identically to the original images, except that a Moffat function with FWHM = 1.4 pixels was used instead of the Tiny Tim PSF.
- To test for the effect of the undersampling, a rebinned and expanded image was made with twice the pixel resolution of the original image (using pixel replication) and convolved with a Moffat function with a FWHM of 2 pixels. This increased the FWHM of the PSF from 0.9 pixels to 4.3 pixels.

3. Photometric Analysis

A photometric analysis was made of the simulated images using two of the widely available reduction programs, ROMAFOT (Buonanno et al. 1983) and DAOPHOT (Stetson 1987). It should be noted that both packages have several tunable parameters which need careful attention in order to produce optimal results. Default values were initially used, followed by a range of parameter changes, in an attempt to optimize the functioning of the programs.

The output data from the photometric programs were compared to the original positions and magnitudes used to create the simulated images. A measured star was considered successfully detected if its position differed less than one pixel from its nearest neighbor with known position. If no neighbors were found within the one pixel radius the star was discarded.

It should be noted that the results from the programs were not edited, except that magnitude errors larger than two standard deviations were considered false identifications.

3.1. ROMAFOT

The version of ROMAFOT available under MIDAS 92NOV was used in our tests. The PSF was extracted from the image and the reduction followed the standard procedure for ROMAFOT. Analysis of the image generated using the Tiny Tim PSF gave substantial errors (σ = 0.1 mag) even for bright stars which also showed some instability. Fig. 5a shows the results with magnitude error as function of magnitude. The error bars represent $\pm 1\sigma$ spread in each magnitude interval. Measurements on the rebinned, expanded image showed only marginal improvements (Fig. 5b), the bright stars were not detected at all. In contrast, the (non-realistic) image generated with the Moffat PSF showed a marked improvement in accuracy (Fig. 6a). Here, the error for bright stars remained within σ = 0.02 mag. This clearly implies a strong dependence on PSF shape, which is not unexpected since ROMAFOT uses analytical Moffat functions to fit the stellar images. A remarkable feature is the systematic underestimation of the intensities for the faint stars. The investigation also gave information regarding detectability of sources. This is seen in Fig. 6, which shows the fraction of found stars as function of magnitude. The total number of detected stars was about 17,000. Thus, the majority of very faint stars remained undetected.

3.2. DAOPHOT

The DAOPHOT version used was DAOPHOT II, which included enhancements up to October 1993. As for ROMAFOT, the PSF was derived from the image. Again, measurements made on the Tiny Tim PSF generated image gave unsatisfactory results. Bright stars showed errors of about σ = 0.05 mag and a zero point offset was seen for fainter stars (Fig. 7a). The rebinned and expanded image gave better results (Fig. 7b), with the smallest errors in the order of σ = 0.04. However, the brightest stars showed some anomaly. Similarly to ROMAFOT, the image with the Moffat PSF gave significantly better results than the image generated with the Tiny Tim PSF; the errors in the magnitudes of the brighter stars were around σ = 0.02. These results are slightly better than those obtained with the rebinned image. For the fainter stars however, the results are significantly better, with nearly twice as many detected stars, and about half the spread in the magnitudes of those stars.

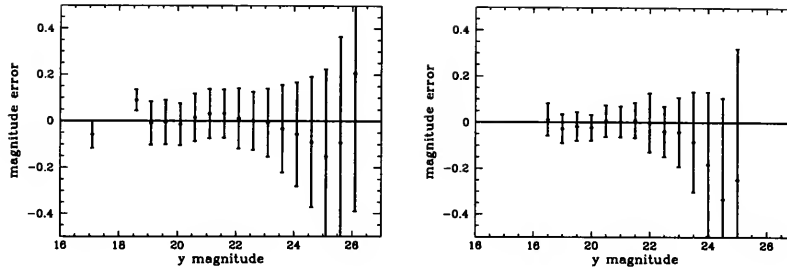


Figure 5. Magnitude error as function of y magnitude as given by ROMAFOT. The error bars represent a deviation of $\pm 1\sigma$. To the left (a) are shown the results from the y image generated with the Tiny Tim PSF. To the right (b) are given the corresponding results for the image after a magnification by a factor of two and a convolution with a Moffat function with FWHM = 2.0 pixels.

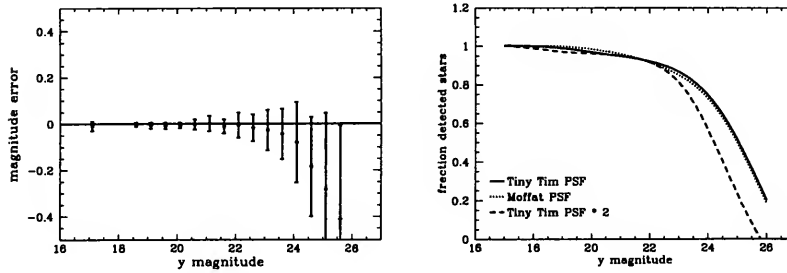


Figure 6. To the left (a) are seen the results for ROMAFOT after analyzing an image generated with a Moffat PSF. To the right (b) is seen the fraction stars found as function of y magnitude. The solid line represents the result from the Tiny Tim PSF generated image, the dashed line the result from the magnified and convolved Tiny Tim PSF, and the dotted line the result from the Moffat PSF generated image.

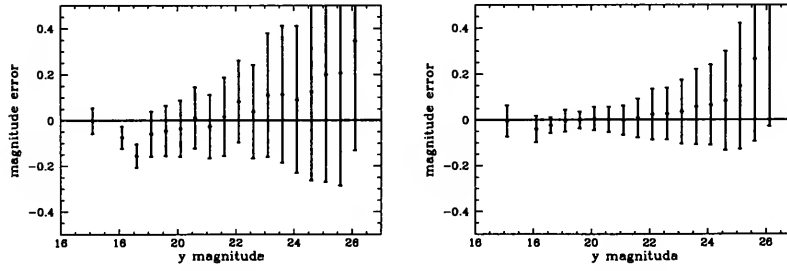


Figure 7. The same as in Fig. 5, but with results derived using DAOPHOT.

The number of detected stars increased from 15,000 to 17,000 when measuring the rebinned and expanded image (Fig. 8a).

Positions were determined equally well in either the original or the rebinned image; slightly better in the image with the Moffat PSF. Fig. 8b shows the spread (with $\sigma = 0.18$ pixels) of position determinations on the original image.

Intrinsic DAOPHOT Errors A simple test was done to see if DAOPHOT could produce consistent results with the complex given input data to determine if it was at all possible to reach the limiting S/N ratio. The results of DAOPHOT's *allstar* were fed as input to a new run of *allstar*. Ideally this should provide identical results. This was, however, not the case for the original image with the Tiny Tim PSF. Only 15% of the stars had identical positions and magnitudes after this second run. For the remaining stars the differences in position and magnitude were determined (Fig. 9a). A positional accuracy of better than 0.02 pixels is required to reach a photometric reproducibility of 5%. The fraction of stars that has positions that can not be determined sufficiently accurate includes a relatively large fraction of the brightest stars. The fact that the two runs gave different results probably depends on the size and shape of the PSF. In the tested case the centers of the stars could not be determined with sufficient accuracy, but it should be emphasized that the FWHM of the PSF is only 0.9 pixels and the fitting radius is only 3 pixels in this case, so the number of pixels with astrometric and photometric information is small.

3.3. Achieved vs. Possible Accuracy

In Fig. 9b is shown the theoretical signal to noise ratio limit for the simulation. This limit, valid for a 45 min exposure with filter F547M (Strömgren y), was determined using Poisson noise, sky background, dark noise, and read out noise, and assumed that the PSF deposited about 30% of the total stellar intensity in one pixel. The entire observation was assumed to be on a single exposure. Also shown are the measurements from ROMAFOT (dotted line) and DAOPHOT (dashed line) from the rebinned image, interpreted as signal to noise ratio.

4. Deconvolution

Some preliminary tests with deconvolution using the Richardson-Lucy algorithm were performed on the simulated images. However, so far no improved results have been seen, either in terms of photometric accuracy or in number of detected stars. In contrast with results from deconvolution of ground-based images (Linde et al. 1993), this is not surprising considering the poor sampling of the images.

5. Summary

- Neither ROMAFOT nor DAOPHOT managed to reach the limiting S/N ratio.
- The shape of the PSF has a large impact on the performance of the photometric packages in the case of undersampled PSFs.
- Rebinning the image to double the original pixel resolution showed no large difference in accuracy for ROMAFOT but increased both limiting magnitude and accuracy for DAOPHOT.
- Deconvolving the test images did not increase accuracy, either for ROMAFOT or for DAOPHOT.
- In view of the above shortcomings, local software (Linde 1989) is currently being adapted to HST PC images in order to better approach the theoretical photometric limit.

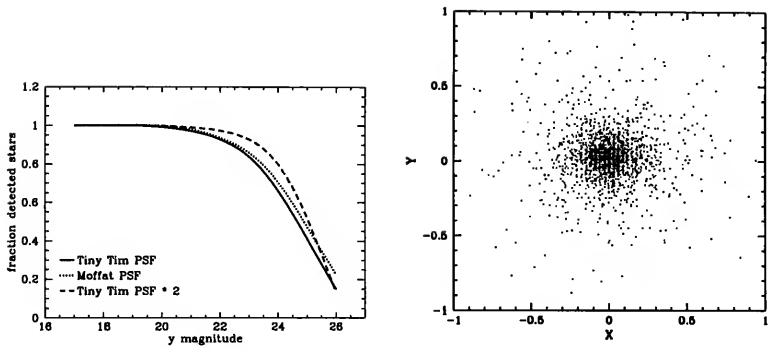


Figure 8. To the left (a) is shown the fraction stars found by DAOPHOT as function of magnitude. To the right (b) are shown positional errors as derived by DAOPHOT from the original Tiny Tim PSF generated image.

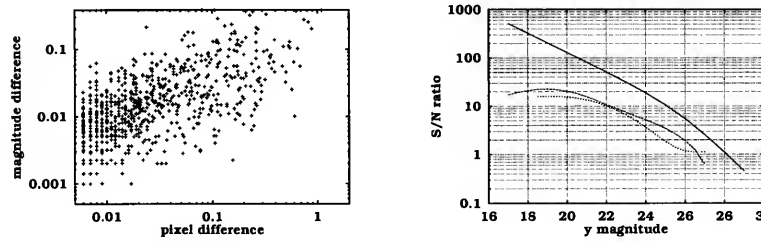


Figure 9. To the left (a) are seen differences between two DAOPHOT runs on the same image. To the right (b) is shown signal to noise ratio as function of magnitude for three cases. The solid line represents the theoretically reachable limit, the dashed line and dotted lines represent actually achieved performance from DAOPHOT and ROMAFOT, respectively.

References

Ardeberg, A., Linde, P., Lindgren, H., Lyngå, G. 1985, A&A, 148, 263

Buonanno, R., Buscema, G., Corsi, C. E., Ferarro, I., & Iannicola, G. 1983, A&A, 126, 278

Krist, J., 1993, *Tiny Tim User's Manual*, Space Telescope Science Institute, Baltimore

Linde, P. 1989, in *1st ESO/ST-ECF Data Analysis Workshop*, P. J. Grosbøl, F. Murtagh, & R. H. Warmels, eds., European Southern Observatory, Garching, 201

Linde, P., Spännare, S. 1993, in *5th ESO/ST-ECF Data Analysis Workshop*, P. J. Grosbøl & R. C. E. Ruijsscher, eds., European Southern Observatory, Garching, 131

Stetson, P. B. 1987, PASP, 99, 191

The Restoration of HST Images and Spectra II
Space Telescope Science Institute, 1994
R. J. Hanisch and R. L. White, eds.

Photometric and Surface Brightness Measurements of Simulated HST Restored Images

F. Fusi Pecci, L. Federici, G. Parmeggiani, F. R. Ferraro, and C. Cacciari

Osservatorio Astronomico, Bologna, Italy

G. Iannicola, C. E. Corsi, and R. Buonanno

Osservatorio Astronomico di Monte Mario, Roma, Italy

F. Zavatti and O. Bendinelli

Dipartimento di Astronomia, Università di Bologna, Bologna, Italy

Abstract. Stellar photometry of simulated HST uncorrected and corrected images of a crowded star cluster has been carried out before and after deconvolution (with RMG, R-L, MEM, etc.) to optimize the reduction procedures and compare the degree of completeness and photometric accuracy actually achievable in the various cases. Experiments are also presented on the description of the surface brightness profiles of simulated clusters with different deconvolution techniques and sizes of the adopted PSF and data matrices.

1. Introduction

The experiments reported here have been made in connection with a general program, aiming at the accurate and detailed photometric study of the Milky Way and M31 Globular Clusters, that our group is currently carrying on by obtaining precise color-magnitude diagrams and complete and unbiased luminosity functions.

Six HST proposals have been approved over the length of the 4 cycles of HST operations, and the specific aim of these experiments is to improve the capabilities of our data analysis and, in particular, of the stellar photometry, which is mostly performed with the ROMAFOT and IRAF packages.

Here we report on the analysis of HST simulated images, with special emphasis on the photometry of individual stars in crowded fields (before and after image restoration) and on the description of the surface brightness profiles of simulated clusters with different deconvolution techniques and sizes of the adopted PSF and data matrices.

2. Stellar Photometry

In order to obtain a quantitative estimate of the accuracy and reliability of our data reduction procedures and results, we have performed a data reduction pipeline based on the following steps: i) define a set of simulated HST-like images with synthetic globular cluster stellar distributions, which provide the "true" data for comparison; ii) convolve these images with appropriate assumptions of PSF and noise in order to construct simulated "observed" data; iii) deconvolve the "observed" images using some of the most widely adopted techniques and assumptions; iv) reduce the "deconvolved" frames in a totally independent way, and compare with the initial "true" data. Specifically, for steps i) and ii) we used the set of HST-simulated images prepared by Hanisch (1993) (namely the files in the directory `star_cluster`), in order to be able to compare our results also with the work of other

researchers in the field, while steps iii) and iv) were performed by two different teams of our group who ignored the assumptions made by the other team in order to assure the complete independence of each step. Since this work was started in September 1993, we adopted the version of Hanisch files then available, which is different from the revised version released on October 6, 1993. The characteristics of our files are then: (i) intensity-normalized PSF peak; (ii) stars centered on the pixels; and (iii) zero-point = 9.0 mag.

In summary, using a special version of the package ROMAFOT (Buonanno et al. 1983, Buonanno & Iannicola 1988), we have then reduced the following set of files:

a) "Truth": input star cluster with a globular cluster-like luminosity function (LF) and spatial distribution (Hanisch 1993). It contains 470 stars.

b) "Observed": HST-simulated "star_cluster" (Hanisch 1993) with 470 stars, old WFC (sim2.fit), new WF2 (sim2c.fit), size = 256×256, space invariant PSFs (psf12.fit, psfc12.fit), filter F555W, center of image, where "sim2" stands for the HST images uncorrected for spherical aberration, and "sim2c" are the HST-corrected frames, and hereafter label the two separate families of adopted simulations.

c) "R-L deconvolved": deconvolved with the Richardson and Lucy method (R-L) (Richardson 1972, Lucy 1974, 1992), using a variable number of accelerated iterations (50, 150, 250).

d) "MEM deconvolved": deconvolved with the Maximum Entropy Method (MEM), adopting two different versions, i.e., MEM-IDL (with 50, 150, 250 iterations; this routine was written by F. Varosi based on Hollis et al. 1992) and MEM/MemSys5 (with ICF = 0, 1; Weir 1991).

e) "Wiener": deconvolved with a Wiener-based routine (IRAF, package *restore*; Busko 1993).

In the reduction of all frames the star detection threshold has been set always at 5σ, the value calculated on the background empirically determined in the "observed" frames (i.e., 8 counts). Note that, due to the combination of the systematic centering of the stars in the pixel centers and the crowding effects, only 373 stars are actually detectable as independent objects on the "observed" frame listed in Table 1. The second column, labelled "Detected", contains the total number of identified peaks, including those which could not be fitted by the ROMAFOT stellar profile fitting routine and are interpreted as noise. Finally, note also that the two original "observed" images (Hanisch 1993) were PSF peak-intensity normalized, therefore the data obtained from the two families (Table 1 and Table 2) are not directly comparable as the corresponding stars have actually different S/N values.

Tables 1 and 2 and Figs. 1 and 2 report the summary of the results for the uncorrected and corrected images "sim2"and "sim2c", respectively.

The main conclusions one can draw from the tables and figures are:

1. With "sim2" (uncorrected HST), the number of detected "true" stars in the "observed" frame is about 20% smaller than the number of stars that could in principle be detected (302 vs. 373), and the number of "fitted" stars includes about 10% of spurious identifications. Overall, the completeness factor (i.e., "true"/470) is about 65%, and it would presumably be somewhat larger if the stars were not assumed to be centered in the pixels (further tests are in progress using a revised version of the simulation where the stars are not centered in the pixel matrix). These numbers increase in the deconvolved images and with increasing the number of iterations, however, the number of "spurious" objects also increases steeply, correspondingly. The behavior of these quantities as a function of magnitude can be seen in Fig. 1.

2. With "sim2c" (corrected HST), the number of stars detected in the "observed" frame (before deconvolution) is almost coincident with the maximum attainable (369 vs. 373). The search on the deconvolved images (at a fixed threshold of 8 counts, as derived from the "observed"

Table 1. Reductions of "sim2" cluster simulation.

File: "Observed" = sim2.fit (470 stars, 373 stars actually detectable)
- WFC, F555W, zero-point = 9.0 mag
- psf12.fit, peak intensity-normalized
- detection threshold = 5σ

mag =	0.5	1.5	2.5	3.5	4.5	5.5	6.5
S/N =	455	272	152	75	40	27	7

File	Number of stars			
	Detected	Fitted	"True"	"Spurious"
Observed	457	341	302	39
R-L 50	322	316	308	8
R-L 150	386	365	339	26
R-L 250	481	473	362	111
MEM 50	363	362	309	53
MEM 150	499	485	333	152
MEM 250	561	544	362	182
MEMSYS/1	386	334	281	53
MEMSYS/0	428	399	338	61
Wiener	251	251	216	35

Table 2. Reductions of "sim2c" cluster simulation.

File: "Observed" = sim2c.fit (470 stars, 373 stars actually detectable)
- WF2, F555W, zero-point = 9.0 mag
- psfc12.fit, peak intensity-normalized
- detection threshold = 5σ

mag =	0.5	1.5	2.5	3.5	4.5	5.5	6.5
S/N =	176	103	65	35	22	12	4

File	Number of stars			
	Detected	Fitted	"True"	"Spurious"
Observed	369	369	369	0
R-L 50	335	335	335	0
R-L 150	356	356	342	13
R-L 250	361	361	344	17
MEM 50	338	338	336	2
MEM 150	343	343	340	3
MEM 250	346	346	345	1
MEMSYS/1	114	114	114	0
MEMSYS/0	365	365	365	0
Wiener	334	334	323	11

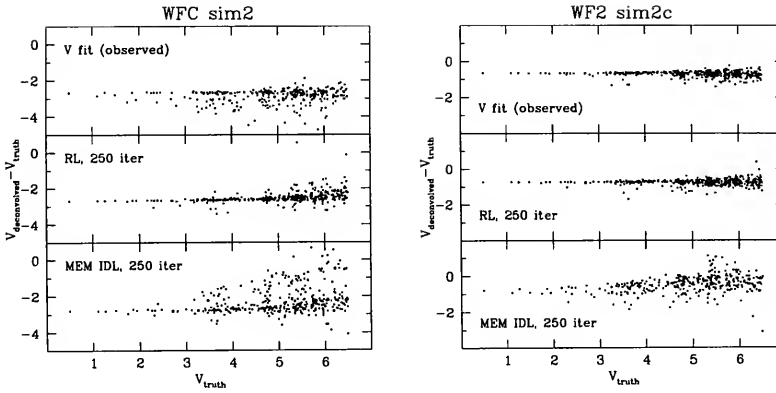


Figure 1. Stellar photometry on star cluster restored images. Residuals are deconvolved minus true magnitudes.

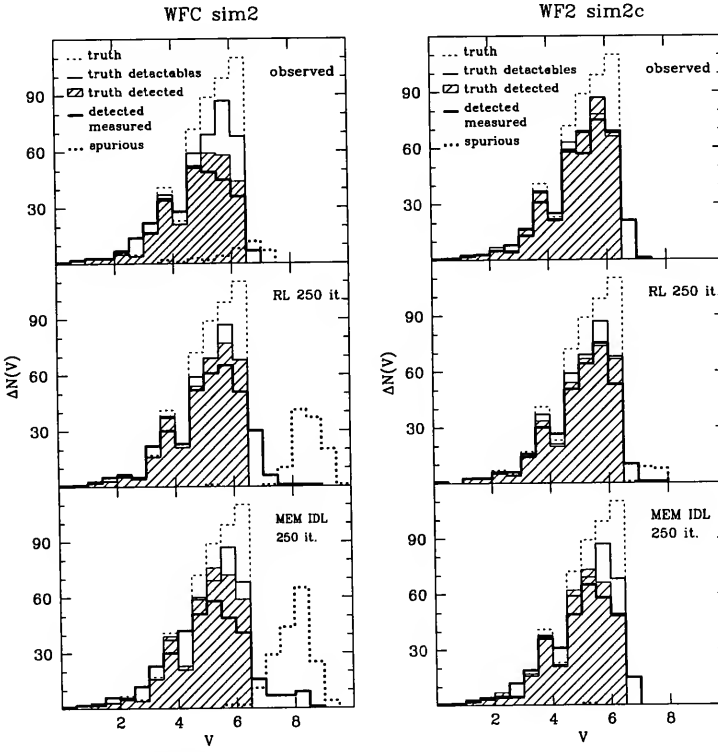


Figure 2. The luminosity functions obtained in the various cases, compared to the initial true distribution

frame) yields a smaller number of detected objects. The stars lost are faint objects with peak intensities just over the threshold in the "observed" frame, which fall just below the threshold after deconvolution. This is due to the fact that all deconvolution methods tend to enhance the contrast, namely to increase slightly the peak intensity of the brightest stars at the expense of the weakest peaks which are damped as noise. Finally, it is remarkable to note the difference in number of detected objects obtained adopting ICF = 1 or 0 with MEMSYS.

3. The impact of these effects on the LFs consists in a "deformation" of the differential LF, with a shift of some objects to the adjacent brighter bin and a lack of stars close to the detection limit. The detection limit is of course defined by the adopted threshold, which was selected in order to optimize the number of detected "true" stars with respect to the spurious identifications.

4. Concerning the linearity, the internal photometric accuracy, and the zero-point checks, additional tests are in progress. In general, the photometry of safely uncrowded stars yields magnitudes accurate to better than 0.1 mag, including any systematic zero-point shift. In particular, the use of a Gaussian fit to the PSF plus a look-up table to map the systematic residuals already yields satisfactory results.

3. Surface Brightness Profiles

From the data taken during Cycle 1, surface brightness profiles have been obtained for 13 globular clusters in M31 (Fusi Pecci et al. 1993), leading also to the detection of a central cusp in a candidate post-core-collapse cluster (Bendinelli et al. 1993). Detailed tests and simulations have been carried out to verify the reliability of the quoted results obtained by applying various deconvolution methods to the observed data.

To further explore the influence on the various procedures of any variation in the size of the adopted PSF compared to the data matrix and in the adopted background, we have carried out here a new set of simulations based on the following assumptions:

a) Adopted cluster model: King model, r_c = 0.2 arcsec, c = 1.75 at the M31 distance (700 kpc), matrix size: 512×512. The simulated cluster has then been convolved with the FOC/96 + F430W observed PSF (star GD248, observed on 12/1/92), with no further noise added (i.e., background = 0).

b) Deconvolutions with: i) Regularized Multi-Gaussians (RMG) (Bendinelli 1991), ii) Richardson-Lucy (R-L) (Richardson 1972, Lucy 1974, 1992), iii) MEM-IDL (Hollis et al. 1992), iv) MEM/MemSys5 (Weir 1991). The deconvolutions have been performed on data matrices of various size, down to 64×64, as specified below.

c) Surface brightness profiles obtained with VISTA.

As shown in Figs. 3 and 4, the following results have been obtained:

1. Both cluster data and PSF matrices have size 512×512 and sky = 0 (Fig. 3a): the derived profiles are substantially coincident with the input model independent of the used deconvolution method.

2. PSF 512×512, data matrix smaller and smaller (256×256, 128×128, 64×64), sky = 0, RMG (Fig. 3b): the derived profiles remain unaffected a part from the reduction in radial extension.

3. PSF and data matrices smaller and smaller (256×256, 128×128, 64×64), RMG (Fig. 3c): there is an increasing shift, parallel to the model, which could be easily calibrated as a function of the matrix size reduction factor if sky = 0, while an increasing divergence with respect to the model occurs if the "local" sky in the data matrix is adopted (Fig. 3d). For "local" sky we intend the average background estimated over the 4 corners of each frame.

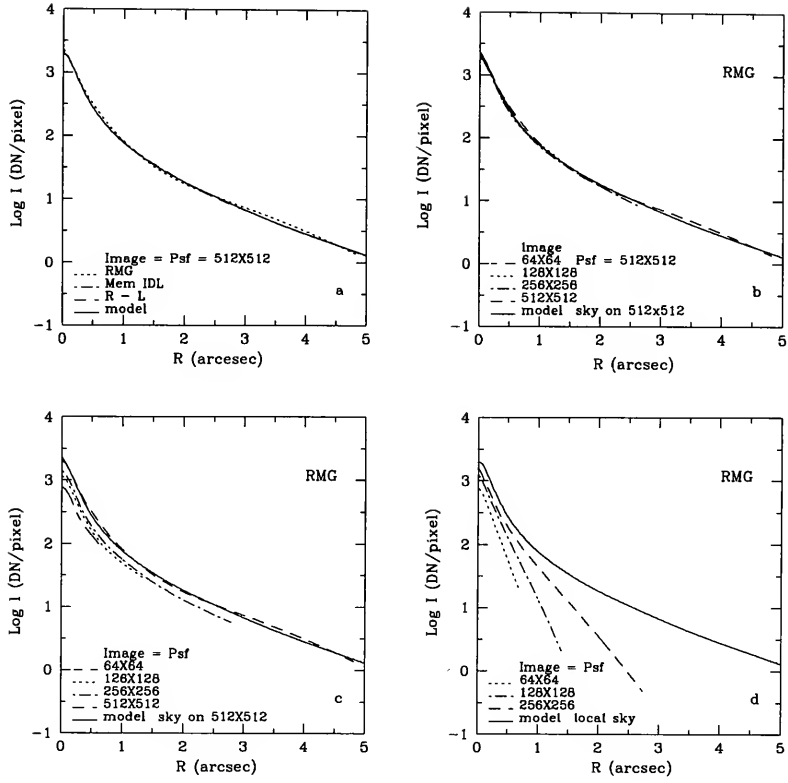


Figure 3. (a) Comparison of the theoretical surface brightness profile with those obtained from the adopted deconvolution procedures; (b,c,d) RMG surface brightness profile: effects of different sizes adopted for the PSF and data matrices.

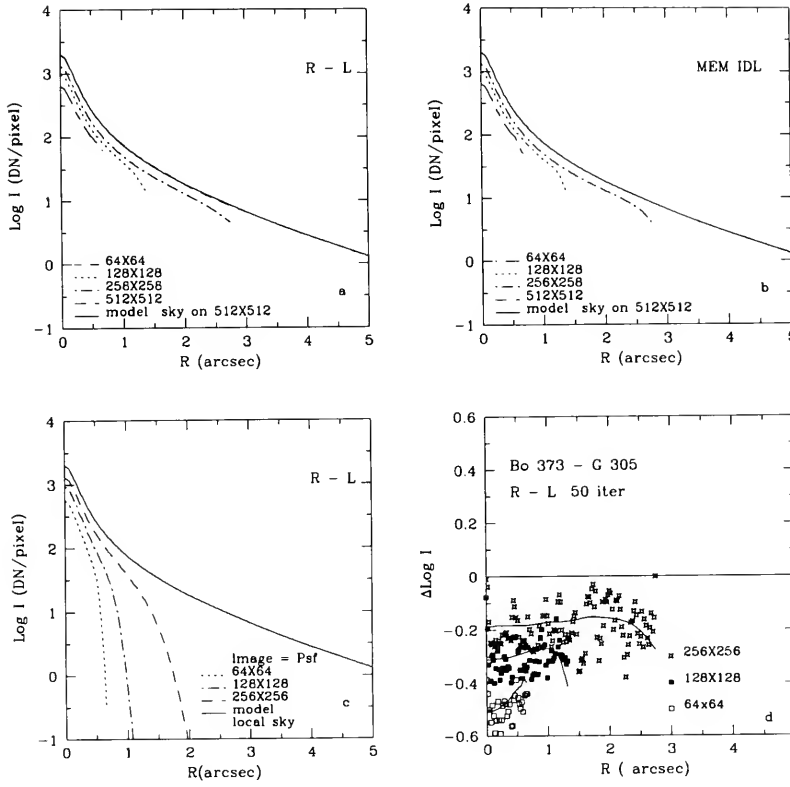


Figure 4. (a,b) Effects of different sizes of the PSF and data matrices on the profiles deconvolved with R-L and MEM; (c) Same as (a) with the assumption of local sky; (d) Residuals predicted by the model simulations with varying the sizes of the PSF and data matrices (solid lines) compared with real HST data of Bo373 = G305.

4. PSF and data matrices smaller and smaller (256×256, 128×128, 64×64), R-L and MEM (Fig. 4a-c): the results are qualitatively similar to those obtained with the RMG (note that the RMG deconvolution operates in the profile domain) but a much stronger edge effect appears.

In Fig. 4d we show the profiles of Fig. 4a differentially with respect to the results of the R-L over 512×512. The models indicate that the use of smaller data matrices produces a shift in the overall brightness profile, which is found also when this test is performed on real data (e.g., the cluster Bo373 = G305).

In summary, it is evident that much care must be taken in order to ensure that all the necessary information is amply contained in the frames to be analyzed. The combination of an overestimated background (such as our "local" sky) coupled with an insufficiently extended PSF may lead to large discrepancies in the derived profiles, which could be partially reduced by the the use of the RMG deconvolution method, as all the profiles are implicitly extrapolated outside the actual matrix size, albeit with a Gaussian falloff.

References

Bendinelli, O. 1991, ApJ, 366, 599

Bendinelli, O., Cacciari, C., Djorgovski, S., Federici, L., Ferraro, F. R., Fusi Pecci, F., Parmeggiani, G., Weir, N., & Zavatti, F. 1993, ApJ, 409, L17

Buonanno, R., Buscema, G., Corsi, C. E., Ferraro, I., & Iannicola, G. 1983, A&A, 126, 278

Buonanno, R., & Iannicola, G. 1988, PASP, 101, 294

Busko, I. 1993, in *Restoration Newsletter*, Space Telescope Science Institute, Baltimore, 1, 60

Fusi Pecci, F., Battistini, P., Bendinelli, O., Bonoli, F., Cacciari, C., Djorgovski, S., Federici, L., Ferraro, F. R., Parmeggiani, P., Weir, N., & Zavatti, F. 1993, A&A, in press

Hanisch, R., 1993, in *Restoration Newsletter*, Space Telescope Science Institute, Baltimore, 1, 76

Hollis, J. M., Dorband, J. E., & Yusef-Zadeh, F. 1992, ApJ, 386, 293

Lucy, L. B. 1974, AJ, 79, 745

Lucy, L. B. 1992, AJ, 104, 1260

Richardson, W. H. 1972, J. Opt. Soc. Am. 62, 55

Weir, N. 1991, in *Proceedings of the 3rd ESO/ST-ECF Data Analysis Workshop*, P. Grosbøl & R. Warmels, eds., European Southern Observatory, Garching, 115

The Restoration of HST Images and Spectra II
Space Telescope Science Institute, 1994
R. J. Hanisch and R. L. White, eds.

Surface Brightness Parameters from Deconvolved PC Images of Elliptical Galaxies

Laura Ferrarese and Holland C. Ford

Johns Hopkins University and Space Telescope Science Institute, Baltimore, MD 21218

Abstract. The spherical aberration affecting the HST primary mirror (Burrows et al. 1992) can seriously alter the core brightness profile of elliptical galaxies. Excellent results in recovering the original HST performances have been obtained by means of deconvolution techniques, improved and updated during the past two years in order to suit HST data (Hanisch 1993). In this paper we test the performance of deconvolution in recovering the original surface brightness profile in the case of elliptical galaxies observed with the HST Planetary Camera (PC). We analyze whether the results of deconvolution depend on the intrinsic morphological properties of the galaxy (e.g., the brightness profile or the ellipticity), or on small variations of the PSF due to jitter, position on the chip, or spectral shape of the point source.

1. Introduction

During 1992 we have obtained HST Planetary Camera (PC) images of a magnitude limited sample of 12 elliptical galaxies in the Virgo cluster (Jaffe et al. 1994, van den Bosch et al. 1994, Ferrarese et al. 1994). The aim of our project is to study the morphological properties of the inner region, derive the core brightness profiles, and look for evidence of nuclear mass concentrations and photometrically decoupled cores. In order to accomplish these goals the full capabilities of HST are needed. In fact, ground based observations are severely impaired by atmospheric seeing; it has been established (Peletier et al. 1990) that seeing effects can extend to radii 5 to 10 times the seeing FWHM, which is unacceptable considering that typical core radii are of the order of only a few arcsec at the distance of the Virgo cluster.

Unfortunately, the spherical aberration affecting the HST primary mirror can seriously alter the morphological and photometric properties of the cores. In the past three years, deconvolution techniques have been extensively improved in order to suit HST data, nevertheless, it is a legitimate doubt to assume that there are cases in which deconvolution may not give satisfactory answers or, even worse, may introduce spurious effects. We have therefore built an extensive set of model elliptical galaxies to test whether deconvolution techniques are able to restore the original surface brightness parameters of the cores, i.e., ellipticity, position angle of the major axis, deviations of the isophotes from pure ellipses, and brightness profile. In this paper, we will discuss the results obtained for the brightness profile, and refer to a separate paper for the results on the other surface brightness parameters (van den Bosch et al. 1994).

2. Models

We have built a set of model elliptical galaxies spanning a wide range of brightness profiles, ellipticity, position angle of the major axis, and isophotal shape. In order to study the effects of the aberrated HST PSF, each model has been convolved with a PSF appropriate for chip 6 of the PC and filter F555W (all of our galaxies have been observed with this instrumental configuration). Poisson noise and read out noise have been added to the convolved, or aberrated, image. The models have been built in order to give a signal-to-noise ratio between 20 and 30 at the galaxy

center after convolution with the HST PSF. Finally, the aberrated image has been deconvolved using the Richardson-Lucy (RL) algorithm. The RL iterations converge to the maximum likelihood solution for Poisson statistics, and are therefore well suited for deconvolving HST PC data. In addition, the RL deconvolved data is forced to be positive and the deconvolved image has good photometric linearity (White 1993). The PSFs used for convolution and deconvolution have been created using the Tiny TIM software (Krist 1992), which is able to model variation in the PSF due to the instrumental configuration (for example, camera and filters), position on the PC chip, spectral shape of the point source, and jitter due to the HST pointing.

In the following sections we will discuss the effects on the results of deconvolution due to:

1. Intrinsic shape of the brightness profile. A brightness profile which flattens up towards the galaxy center is expected to be less affected by convolution than a very steep brightness profile, and, as a consequence, deconvolution is expected to give better results for a flatter profile rather than for a steeper one;

2. The ellipticity of the galaxy. The ellipticity of an elongated galaxy will decrease towards the center as a consequence of the convolution with the aberrated HST PSF. Since the brightness profile is found by integrating along isophotes, larger deviations are expected in the brightness profile of deconvolved galaxies when the ellipticity is higher;

3. Small changes in the PSF, as due to the position of the PSF on the PC chip, jitter, and spectral shape of the point source.

2.1. Dependence on the Brightness Profile Shape

We have considered four different forms for the brightness profile, three with one component only, and one with two separate components, covering a wide range in shape and steepness in the crucial (as far as convolution and deconvolution are concerned) inner 5 arcsec: a de Vaucouleurs law

$$\Sigma(r) = \Sigma_0\, 10^{-3.33[(r/r_0)^{1/4}-1]} \tag{1}$$

where r is along the major axis of the galaxy, $\Sigma_0 = 5$ DN and $r_0 = 13$ arcsec; a Hubble law

$$\Sigma(r) = \frac{\Sigma_0}{(1 + r/r_0)^2} \tag{2}$$

with $\Sigma_0 = 2000$ DN and $r_0 = 4.3$ arcsec; an exponential profile

$$\Sigma(r) = \Sigma_0\, e^{-r/r_0} \tag{3}$$

with $\Sigma_0 = 1500$ DN and $r_0 = 13$ arcsec; an exponential profile plus a central Gaussian component

$$\Sigma(r) = \Sigma_{01}\, e^{-r/r_0} + \Sigma_{02}\, e^{-r^2/2\sigma^2} \tag{4}$$

with $\Sigma_{01} = 1500$ DN, $r_0 = 13$ arcsec, $\Sigma_{02} = 500DN$, $\sigma = 1.6$ arcsec.

The brightness profiles described by Eqs. 1 to 4 are sketched as solid lines in the upper panels of Fig. 1 for an elliptical galaxy with constant position angle and mild ellipticity ($\epsilon = 0.2$), and located in the Virgo cluster. In the same figure, the dashed lines represent the same, but now aberrated, brightness profiles, derived by convolving the model images with an HST PSF and adding noise. As can be seen, substantial differences are present in the inner 2 arcsec, the aberrated profile being lower and flatter than the unaberrated one. As in the case of images blurred by atmospheric seeing, light is removed from the innermost part of a galaxy and pushed to intermediate radii, while the surface brightness at large radii remains undisturbed. In contrast with the seeing PSF, however, the HST PSF is very stable and very accurately known, and this allows the successful use of deconvolution techniques which are inapplicable in the case of seeing convolved images. The

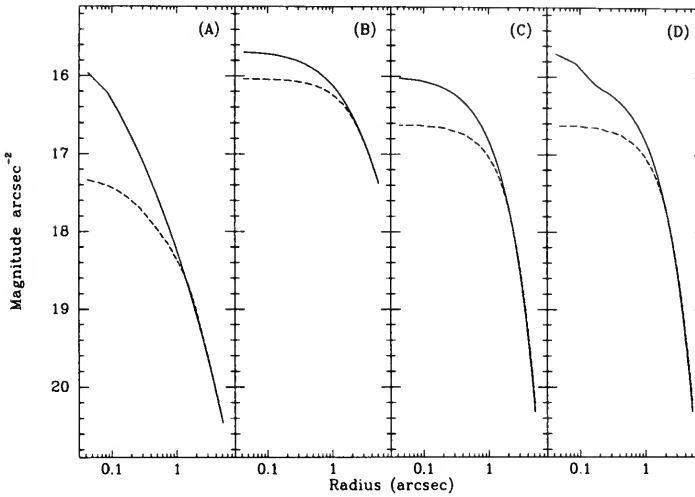


Figure 1. The solid lines in the upper panel show the brightness profiles corresponding to (A) a de Vaucouleurs law (Eq. 1 in the text), (B) a Hubble law (Eq. 2), (C) an exponential law (Eq. 3), (D) an exponential law plus a central component (Eq. 4). The dashed lines show the same brightness profiles as seen by the aberrated HST PC. The lower panels show the residuals between the deconvolved brightness profiles and the expected brightness profiles (solid lines in the upper panels).

process of convolving model galaxies with an HST PSF has been carried out using several HST PSFs, differing one from the other for the position on the chip, the spectral shape, or the amount of jitter. The results show that the aberrated image does not change significantly due to variation of the PSF.

At this stage, the RL algorithm was employed in order to determine if and how differences in steepness and curvature of the brightness profile influence the results of deconvolution. In Fig. 2 we show the difference between the unaberrated brightness profile shown in Fig. 1 as solid line and the deconvolved brightness profile obtained from the aberrated images after 100 RL iterations. The PSF used for the deconvolution was the same as the one used for the initial convolution. The original brightness profile is reproduced by RL deconvolution with accuracy higher than 0.05 magnitudes arcsec^{-2} at radii larger than 0.2 arcsec and accuracy higher than 0.15 magnitudes arcsec^{-2} at smaller radii.

2.2. Dependence on the Ellipticity

We have also determined whether the accuracy of the deconvolution techniques depends on other surface brightness parameters, such as ellipticity, position angle and departure of the isophotes from pure ellipses. Fig. 3 shows the difference between unaberrated brightness profile and deconvolved brightness profiles (after 100 RL iterations) in the case of four elliptical galaxies with a de Vaucouleurs profile (Eq. 1), constant position angle, but different ellipticities, $\epsilon = 0.0$, $\epsilon = 0.2$, $\epsilon = 0.4$, $\epsilon = 0.6$. The same PSF has been used for convolution and deconvolution. From the figure, it is apparent that deconvolution performs slightly better at lower ellipticities, nevertheless, as in the case presented in Fig. 2, we find that deviations in the inner 0.2 arcsec are not higher than 0.15 magnitudes arcsec^{-2}.

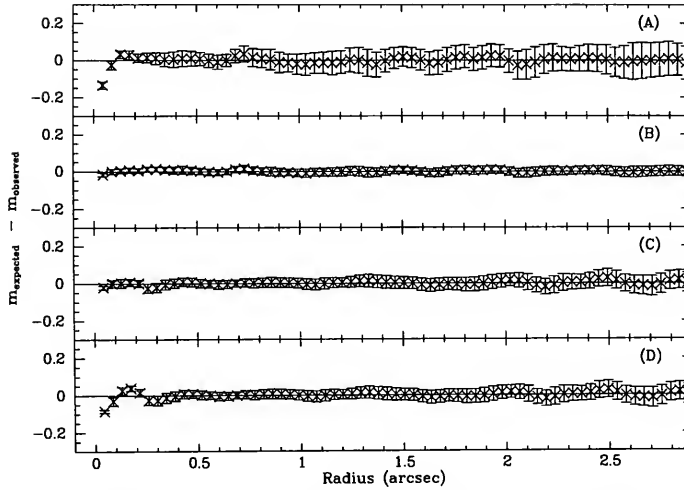


Figure 2. Residuals between the deconvolved and expected brightness profiles. The original brightness profile is given by (A) a de Vaucouleurs law, (B) Hubble law, (C) exponential, (D) exponential plus a central Gaussian component.

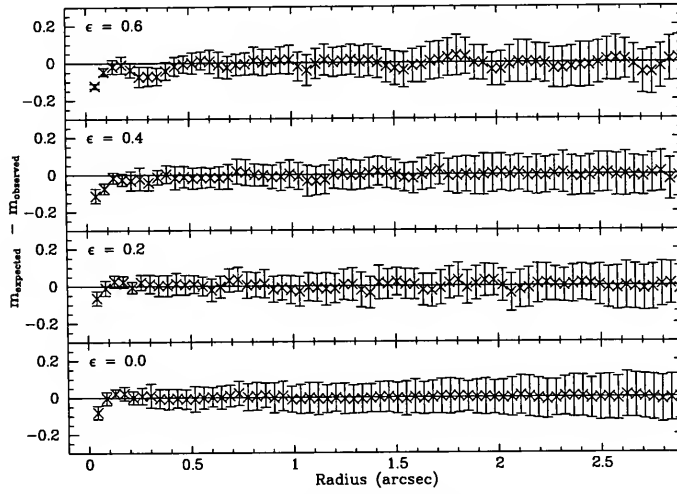


Figure 3. Residuals between the deconvolved and expected brightness profiles for a galaxy with $\epsilon = 0.0$, $\epsilon = 0.2$, $\epsilon = 0.4$, $\epsilon = 0.6$.

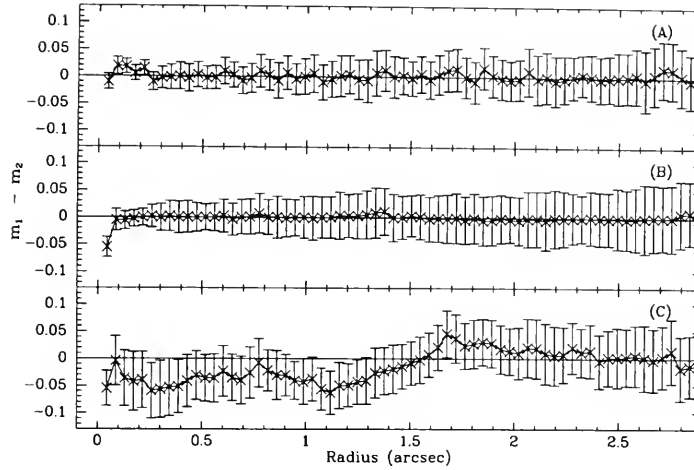


Figure 4. Residuals between the deconvolved brightness profiles for an E2 galaxy obtained with PC PSFs differing in (A) the amount of jitter, (B) the spectral shape of the stellar component), and (C) the position of the PSF on the chip.

In the same way, we found that the results of deconvolution are not affected by variations of the major axis position angle (models were built with position angle varying up to 90 degrees in a 4 arcsec range) or by deviations of the isophotes from pure ellipses (models were built with fourth cosine Fourier coefficient as high as 0.05 in absolute value).

2.3. Dependence on the PSF Shape

Finally, we have studied the effects of small variation in the PSF on the results of deconvolution. The detailed PSF shape depends on the position on the chip, on the spectral shape of the point source, and on jitter due to the HST pointing. The following PSFs have been considered:

1. K star, observed on 1 Jan 1993 at chip position (200,200) with 0 mas jitter
2. K star, observed on Jan 1, 1993 at chip position (200,200) with 30 mas jitter
3. O star, observed on Jan 1, 1993 at chip position (200,200) with 0 mas jitter
4. K star, observed on Jan 1, 1993 at chip position (600,600) with 0 mas jitter

Fig. 4 shows the difference between the deconvolved brightness profiles of a model galaxy (de Vaucouleurs profile, $\epsilon = 0.2$) deconvolved using different PSFs, in particular: Fig. 4A shows the difference between the brightness profiles obtained by deconvolving with PSF 1 and PSF 2, differing in the amount of jitter, 0 mas in one case and 30 mas in the second case (30 mas is likely to be the highest jitter that can occur, at least in fine lock mode); Fig. 4B shows the difference between the brightness profiles obtained by deconvolving with PSF 1 and PSF 3, differing in the spectral shape of the point source (K and O stars, respectively); Fig. 4C shows the difference between the brightness profiles obtained by deconvolving with PSF 1 and PSF 4, differing in the position of the PSF on the chip.

The largest errors are due to variations of the PSF with position in the chip (Fig. 4C), which can cause deviations up to 0.1 magnitudes in the inner 1.5–2 arcsec. The case presented in the figure is quite extreme, since the two PSFs used have been modeled for positions far apart one from the other and close to the edge of the chip, where asymmetries in the PSF shape are larger. Our models shows that if the PSF used for deconvolution is modeled for the position corresponding to the center of the galaxy, the deconvolved brightness profile deviates from the original by less than 0.05 magnitudes arcsec^{-2}, and these deviations are limited in the inner 0.2 arcsec.

3. Conclusions

Our conclusions can be summarized in the following points:

1. The aberrated HST PSF modifies the brightness profile of an elliptical galaxy up to 2 arcsec from the center, in a way which is not sensitive to the exact shape of the brightness profile, ellipticity, major axis position angle, and small PSF variations.

2. The brightness profile derived from a Richardson-Lucy deconvolved image reproduce the original brightness profile with accuracy higher than 0.15 magnitudes arcsec^{-2} at radii smaller than 0.2 arcsec, and higher than 0.05 magnitudes arcsec^{-2} at larger radii. The performance of deconvolution is not noticeably affected by the intrinsic shape of the brightness profile nor by other parameters such as ellipticity, isophotal twist, or deviations of the isophotes from pure ellipses. In addition, if the PSF used for deconvolution is chosen wisely, i.e., close to the center of the galaxy, deconvolution is robust against small PSF variations due to jitter, spectral shape of the point source and position of the PSF on the chip.

3. Studies by van den Bosch et al. (1994) show that more care is needed when deriving the ellipticity, the major axis position angle, and the deviations of the isophotes from pure ellipses, especially in the case in which these parameters vary with distance from the center. In general, it is safe to assume that ellipticity, position angle, and Fourier coefficients are recovered correctly by Richardson-Lucy deconvolution up to 0.5 arcsec from the galaxy center.

References

Burrows, C. et al. 1991, ApJ, 369, L21

Hanisch, R. J., ed., 1993, *Restoration Newsletter*, Space Telescope Science Institute, Baltimore

Jaffe, W., Ford, H. C., O'Connell, R. W., Ferrarese, L., & van den Bosch, F. C. 1994, in preparation

Krist, J. 1992, "The Tiny TIM User's Manual," Version 1.0, Space Telescope Science Institute, Baltimore

Peletier, R. F., Davies, R. L., Illingworth, G. D., Davis, L. E., & Cawson, M. 1990, AJ, 100, 1091

van den Bosch, F. C., Ferrarese, L., Jaffe, W., Ford, H. C., & O'Connell, R. W. 1994, in preparation

White, R. L. 1993, *Restoration Newsletter*, R. J. Hanisch, ed., Space Telescope Science Institute, Baltimore, 1, 11

The Restoration of HST Images and Spectra II
Space Telescope Science Institute, 1994
R. J. Hanisch and R. L. White, eds.

Maximum Likelihood Estimation of Galaxy Morphology: Faint HST Wide Field Camera Images

Kavan U. Ratnatunga, Richard E. Griffiths, and Stefano Casertano

Center for Astrophysical Sciences, Johns Hopkins University, Baltimore, MD 21218

Abstract. A modeling approach based on the maximum likelihood method has been developed to extract quantitative morphological and structural parameter estimates for faint galaxy images obtained with the Hubble Space Telescope (HST) Wide Field Planetary Camera (WFPC). We model both the galaxy image and the instrumental characteristics of the WFPC, including the complex Point Spread Function (PSF), the error in the Analog-to-Digital Converter (ADC), and the positive noise bias due to faint cosmic rays and undetected warm pixels. Because convolved galaxy images are compared directly with the observations, we avoid the need for deconvolution, which is difficult and potentially unstable for faint images.

1. Introduction

The HST Medium Deep Survey (MDS) Key Project (Griffiths et al. 1992, 1994) includes observations of a large number of random WFC fields. They are taken in parallel mode, when another detector is being used on a primary target at a single pointing extending over two or more orbits. Most of the objects detected are faint galaxy images. For statistical analysis of the large scale structure of the universe, the survey is generating a catalog of galaxy magnitudes, colors, half light radii, axis ratios and morphology. In order for this catalog to be well-defined despite the relatively long span over which the observations have been obtained, and to ensure uniformity in the processing, we need reliable error estimates for all evaluated parameters to combine information from different fields, with different integrated exposure times.

In the traditional approach to image analysis, CCD observations are first calibrated to subtract the bias and dark current and eliminate variations in the detector sensitivity. Bright cosmic rays are then removed, by stacking multiple observations if available, otherwise by using one of several algorithms that identify cosmic ray events in the image. Defective pixels may be removed by interpolation. For HST, the brighter images are then deconvolved to restore the images to the high resolution of the core of the PSF.

This approach can be very successful when dealing with bright (high signal-to-noise) images. The result is a cosmetically clean image on which morphological classification and other measurements can be carried out in a non-parametric, model-independent way. However, this procedure does have some drawbacks when applied to a large sample of galaxies. Some quantities, such as magnitudes, are very difficult to measure accurately and without bias on deconvolved images; even for those quantities that can be measured, the error distribution is distorted and probably non-Gaussian. Intercomparison of galaxy properties does require, implicitly or explicitly, the adoption of a model. The primary disadvantage is that deconvolution of faint images is likely to be unstable (Schade and Elson 1993).

In the alternative approach described below we assume a simple parametrized model for the galaxy image and transform it to the observed domain. Parameters used in the model definition are then estimated by maximizing the likelihood function, which is the probability of obtaining each observed image for any set of galaxy parameters. Tests of the procedure on simulated galaxy

images show that derived model parameters and their error estimates are unbiased. The likelihood ratio between fits to different types of models can be used for a coarse morphological classification.

2. Model Fitting

The properties of an ensemble of faint galaxies can be defined by the distribution of the parameters of the models that best fit each individual object. This is only practical for faint images if the number of adjustable parameters to be fitted is kept to a minimum; a possible initial set would include total magnitude, half-light radius, ellipticity, orientation and centroid position. For brighter images, where more signal is available, we can fit additional shape parameters, such as bulge-to-disk ratio. The model also includes instrumental parameters, such as the sky background (estimated locally) and amplitude of the noise. The PSF and the distribution of faint cosmic rays are measured separately. It should be pointed out that the modelling approach for extended objects is relatively insensitive to small-scale errors in the description of the PSF.

For any set of model parameters, a smooth galaxy image is generated and convolved with the expected PSF. The expected sky brightness is added and the image is multiplied by the detector sensitivity (flat field). The probability of the observed value is then computed for each pixel using the adopted error distribution. Since the parameters are estimated as the maximum likelihood model that fit the observations, we can ignore defective pixels without any need to interpolate them. The likelihood function is defined as the sum of the natural logarithm of the probability for individual pixels, including both galaxy pixels and neighboring sky pixels (in fact, pixels do not have to be separated into 'galaxy' and 'sky'). For a Gaussian error distribution, the likelihood function is proportional to the χ^2 of the distribution; however, the method allows for a different error distribution can be used, to take into account both the uneven probability distribution of different data numbers due to the ADC error, and the wide positive tail due to faint cosmic rays and warm pixels.

The choice of models is based on standard galaxy models and include an exponential disk and an $r^{1/4}$ law. Because the method involves a non-linear optimization, the choice of initial guesses for the parameters is quite important; we generally use initial values estimated from moments of the image, as well as the results of the deconvolution if available. Optimization uses the Numerical Algorithm for Maximum Likelihood Estimation (NAMaLiE) developed by Ratnatunga and Casertano (1991). The information on errors of the estimated parameters and their correlation can be obtained from the covariance matrix, which is the inverse of the Hessian (matrix of second order derivatives) at the peak of the likelihood function. Extensive numerical simulations have shown that both parameter estimates and their errors are unbiased.

This approach also requires an error map for the calibrated image, which is not generated by the current HST WFPC pipeline calibration. We use a simple noise model to estimate the rms noise in each pixel. Special consideration must be given to noise in the calibration images (bias, dark and flat field), which, unlike image noise, does not decrease as multiple exposures are stacked together, unless the exposures are shifted with respect to one another. In some cases involving multiple exposures, the noise could be estimated from the internal scatter between individual exposures, plus the contribution of the calibration files.

The contribution of faint cosmic rays to the noise is found to be described adequately by a Weibull distribution with power law index of 0.25; convolved with a Gaussian distribution for shot and read noise, this is a good representation for the total 'noise' distribution over the full range in data numbers. The two additional parameters required to describe the distribution of cosmic rays are derived from a single fit for the whole CCD, and therefore do not increase the number of parameters to be fitted for each image. Empirically, the scatter between different exposures is found to be small.

3. Simulations

The galaxy models that have been extensively tested include an exponential law, meant to represent a disk galaxy, and an $r^{1/4}$ law, representing ellipticals. Both are fully described by six adjustable parameters: total magnitude, half-light radius, axis ratio, position angle, and the coordinates of the center. When both fits are performed on a simulated image of either kind, the value of the likelihood function is very effective in choosing the "right" model down to typical magnitudes of $V = 23$ or $I = 22$ with a single-orbit WFPC exposure, and about one magnitude fainter with the expected parameters for WFPC 2 observations (see Table 1). (Six orbits are required to reach one magnitude fainter with either camera.)

Fits can be successfully performed for fainter galaxies, but at one magnitude fainter than the value in Table 1, the ability to discriminate between models is substantially impaired, and therefore the fitted half-light radius may be biased by the model adopted. Not surprisingly, the bias is smaller than the error in the estimated parameters.

Table 1. Magnitude limits for classification of faint galaxies (single HST orbit).

Detector	Filter	Exp. Time	Sky Mag.	Lim. Mag.
WFC	F555W (V)	2400	23.00	23.00
WFC	F785LP (I)	2400	21.75	22.00
WFPC2	F555W (V)	2400	23.00	24.00
WFPC2	F814W (I)	2400	21.75	23.50

In order to fit more complex models successfully, better signal-to-noise ratio is required. Typically, we estimate that the image must be one magnitude brighter for each additional parameter to be fitted. An additional complication arises for combined models consisting of an exponential (disk) and an $r^{1/4}$ law (bulge) together. We find that the magnitude difference between the disk and the bulge correlates strongly with the ratio of their sizes, even for bright simulated images. This indicates that an image fit has poor leverage to determine the parameters of the two components independently.

We note also that stellar images can be easily fitted about two magnitudes fainter than the limits in the table, thanks to the smaller number of free parameters. However, we have not investigated the effects of errors in the PSF, which may have a greater impact on stellar than on galaxy images. Such issues have been discussed extensively for stellar photometry packages that use the fitting approach, such as DAOPHOT.

4. Discussion

While extensive experience has been garnered in the application of maximum likelihood to stellar photometry with HST (see for example Stetson 1994), comparatively little has been done for galaxy images (see Schade and Elson 1994). What we present here is only a first step, albeit a successful one. The real test of the method comes from its application to actual data.

In the application to real HST observations, we generally do achieve convergence on galaxy parameters to the magnitude limits given in Table 1. However, significant ambiguity remains in the interpretation of the results. Real galaxy images do not follow the simple models used here: spiral arms, bars, double nuclei, dust lanes, and bright knots of star formation all add complications that cannot be properly modeled with the limited information available in typical images. The only recourse in our exploratory runs has been to estimate a best fit using a simple galaxy profile and to inspect the residual distribution for additional visual clarification. More detailed results will be reported in Griffiths et al. (1994) and Ratnatunga, Griffiths, and Casertano (1994).

We should also point out that a key ingredient in achieving our limited success on actual data has been the improved calibration of WFC images, optimized for the quantitative analysis of faint images. The details are reported in Ratnatunga et al. (1994).

Acknowledgments. This work is based on observations taken with the NASA/ESA Hubble Space Telescope, obtained at the Space Telescope Science Institute, which is operated by the Associations of Universities for Research in Astronomy, Inc., under NASA contract NAS5–26555. Coordination and analysis of data for the Medium-Deep Survey is funded by STScI grants GO 2684.0X.87A and GO 3917.0X.91A. We acknowledge the helpful comments from the full HST Medium Deep Survey team.

References

Griffiths, R. E. et al. 1992, in *Science with the Hubble Space Telescope*, P. Benevenuti & E. Schreier, eds., European Southern Observatory, Garching, 13

Griffiths, R. E. et al. 1994, ApJ, submitted

Ratnatunga, K. U.,& Casertano, S. 1991, AJ, 101, 1075

Ratnatunga, K. U., Griffiths, R. E., Casertano, S., Neuschaefer, L. W., & Wyckoff, E. W. 1994, in *HST Calibration Workshop Proceedings*, C. Blades, ed., Space Telescope Science Institute, Baltimore

Ratnatunga, K. U., Griffiths, R. E.,& Casertano, S. 1994, in preparation

Schade, D. J., & Elson, R. A. W. 1993, AJ, 105, 1581

Schade, D. J., & Elson, R. A. W. 1994, in preparation

Stetson, P. B. 1994, this volume

The Restoration of HST Images and Spectra II
Space Telescope Science Institute, 1994
R. J. Hanisch and R. L. White, eds.

High Resolution Measurements from HST Power Spectra

P. Nisenson and E. Falco

Harvard-Smithsonian Center for Astrophysics, 60 Garden Street, Cambridge, MA 02138

R. Gonsalves and S. Ebstein

Electro-Optics Technology Center, Tufts University, 4 Colby Street, Medford, MA 02155

Abstract. Power spectral analysis (PSA) has proven to be an important tool for high spatial resolution measurements from ground based interferometry. It is particularly useful for parameter estimation of stellar diameters, binary separations, and stellar asymmetries. While many images from HST are far too complicated for PSA, there are a class of scientific problems for which it is a very effective technique for handling the problems created by the HST point spread function. We have been applying PSA to images of gravitationally lensed QSOs, and other objects with some surprising results.

1. Introduction

Power spectral analysis (PSA) has long been used for parameter estimation from interferometric data. It is particularly useful when only a few parameters are to be obtained from an image, such as the estimate of stellar diameters, or the measurement of binary separations, intensity ratios and position angles. For example, PSA was used to measure the diameter of the debris from the explosion of SN1987a (Karovska et al. 1991) and estimate the asymmetry in the envelope (Papaliolios et al. 1989) from speckle data recorded over a three year period. Despite the very limited SNR in the data, accurate measurements were obtained from the time when the envelope was only a few milli-arcseconds in diameter, until it was over 40 mas. PSA has also been extremely useful for rapid screening of images to detect duplicity or asymmetries, where the images are complicated by a complex point spread function (as is the case for the atmospheric PSF or the HST PSF). We have applied PSA to a number of different parameter estimation problems from HST WF/PC and FOC image data. In this paper, we will describe our approach to applying PSA and a show an example of its use with WF/PC images of gravitationally lensed QSOs.

2. Power Spectral Analysis

The first step in PSA is to Fourier transform the image $I(x, y)$ and calculate the square modulus of the FT to generate the power spectrum, $p(u, v)$. So,

$$p(u, v) = |\ \mathrm{FT}[I(x, y)]\ |^2 \ . \tag{1}$$

Here FT denotes the Fourier transform. The power spectrum generally has a very wide dynamic range so we rescale it for display with the following operations. First, to reduce the level of the DC component we reset it to equal the value of an adjacent frequency:

$$p(1, 1) = p(1, 2) \, . \tag{2}$$

Next, we rearrange the power spectrum quadrant by quadrant so that the zero frequency is in the center of the array, the 1st quadrant is in the upper right, the second in the upper left, etc. Then we

rescale the array to have values from 1.0 to 10.0 and take the base 10 logarithm, so the final array used for display, $p_d(u,v)$ has the form

$$p_d(u,v) = \log\left(9\,\frac{p(u,v) - \min(p(u,v))}{\max(p(u,v)) - \min(p(u,v))} + 1\right) \tag{3}$$

where min and max are the minimum and maximum of the arrays, respectively.

Display of this rescaled array with a false color display program such as SAOIMAGE or IDL allows observation of a range of sometimes subtle effects which relate to high resolution detail in the image. Working in frequency space has the advantage that the smallest scale features in the image are the largest scale in the power spectrum, so close binaries or image elongation may be immediately obvious in the display. Conversely, the large scale features in the image, the nonuniform background and the extensive wings in the point spread function affect mostly the low frequency region of the power spectrum, so they can be ignored in the inspection process.

An important application of PSA is to detect and characterize binary images. The power spectrum for a binary is just a cosine function whose frequency is proportional to the binary separation, whose modulation is related to the relative intensity of the two components, and whose fringes are perpendicular to the position angle of the binary. Since close binaries give widely spaced fringes, it is easy to fit the fringes to obtain these three parameters. After determining the position angle, one can average the power spectrum in a direction perpendicular to the direction of the fringes, increasing the SNR for the fit and reducing the fit to a one-dimensional operation. This allows accurate measurement even for low SNR data.

For resolved objects, one can fit other simple models to extract diameters. For example, the power spectrum of a disk just has the form $\mid J^2{}_1(r)/r^2 \mid$ and this may be modified to account for a limb-darkening function. If just the diameter of the object is of interest, one can azimuthally average the power spectrum into an array of increasing radius annuli, increasing the SNR for the fit. For asymmetric objects, one can fit an elliptical function to the power spectrum, giving two more extracted parameters (the ratio of minor to major axes of the ellipse, and its position angle).

In all cases, one can restrict the range of frequencies to which the fit is applied. Specifically, one excludes the very lowest frequencies where the large scale nonuniformities in the image have the greatest effect, and one limits the fit to a frequency range where the SNR is sufficiently high. In all cases, calibration of the fits are required by fitting to power spectra of unresolved stars with roughly the same SNR and from a relatively close field position. One can also use PSFs from a program such as Tiny TIM for this calibration, though using both is probably the safest option. Again, because one is doing a global fit, the detailed two-dimensional structure in the PSF, which is still poorly known, is not nearly as important as it is for image deconvolution.

Another operation that proves to be useful for many objects is to apply a spatial filter to the image before generating the power spectrum. One can perform a range of operations including simple extraction of regions of interest, subtraction of background, removal of cosmic rays, etc. These operations need to be performed very carefully to minimize their effect on measurements made in the power spectra.

Another interesting feature of PSA is that one can use it to detect resolved structure on scales small compared to the telescope diffraction limit and the camera sampling limit. The telescope diffraction limit is the frequency at which the telescope aperture transfer function drops to zero. However, objects which are somewhat smaller than the diffraction limit, or binaries with separations less than the diffraction limit, show significant drops in power in the high frequency region of their power spectra as compared to truly unresolved objects. Therefore, while the HST diffraction limit in the visible is only 40 mas (and the PC sampling limit is only 86 mas) stellar diameters and binary separations to a few mas could be estimated from high enough SNR data.

3. Results

We have been applying PSA to a number of scientific problems with some substantial success. We have been using it to analyze PC images of known gravitational lenses and QSOs which are candidates for lensing (Falco 1993). One inspects the power spectra for unusual features such as unexpected asymmetries or unusually rapid falloff at high frequencies. It is very easy to quickly assess whether any of the candidate QSOs have close companions by looking for fringes in the power spectra. Our greatest success to date has been in analyzing PC images taken on a known gravitationally lensed system, Q0142–100, which has a bright image to the west and a fainter image to the east. The lensing galaxy has been detected from ground based images (Surdej et al. 1988) and models for the system were generated by the same group. Examination of the power spectra from the PC images reveals that, in fact, the brighter image is actually double, more or less in a radial direction from the weaker image.

Fig. 1 plots the higher contours of the PC image of 0142 in the upper left, and its power spectrum in the upper right. For comparison, the image of a star is shown in the lower left and its power spectrum is in the lower right. The power spectrum of 0142 shows fine fringes which are about 15° from horizontal, due to the widely separated pair of lensed images. It also has a highly elongated envelope to the fringes which is characteristic of a close binary. In comparison, the stellar power spectrum is much more symmetric, particularly at high frequencies. If we fit the envelope of the power spectrum with a binary model, we find that best fit gives a separation of the two components of 1.3 PC pixels or about 58 mas. Since this separation is too close to actually resolve the two components in the image, one would not know whether this was an elongation of the object or a truly separated binary image. However, gravitational lens optics require it to produce two separated images rather than an elongated image. Surdej et al. (1988) have generated different models of the system which depend on the exact superposition of the lensing galaxy and the QSO. One model generates only two images, another breaks the bright image into separate tangential images, while a third gives radial components for the bright image. Our detection should precisely proscribe the system geometry. The fit also gives a relative intensity of the two components as 5 to 4. Due to the centro-symmetry of power spectra, it is not possible to tell which component is the brighter. However, this can be ascertained from the image, as described below.

Fig. 2 shows cuts through the 0142 and field star power spectra. The upper two curves are x–y cuts through the star power spectra, showing no obvious asymmetry between two directions. The lower two curves are cuts through the 0142 power spectrum. The lowest curve is across the fringe and the higher curve is along the fringe. The strong asymmetry is obvious.

We have also gone back to the image and subtracted the PC PSF generated by Tiny TIM for that position, with alignment between the brightest pixel of 0142 and the PSF. There is a strong residual, somewhat fainter peak to the west of the brighter peak, presumably corresponding to the second component. A surface plot of the residual peak after subtraction is shown in Fig. 3. A similar subtraction from the field star results in much smaller residuals.

Better PSF modeling is needed to get really accurate fitting in the image. However, earlier attempts at detection of close companions using standard point source fitting routines (before it was found from the power spectrum) did not find this peak. Knowing exactly where to look from the power spectra makes the point source fitting much easier. This extremely close pair of images would have only a few days time delay between the optical paths, so if the QSO varies in brightness and regular observations were made, it could be an ideal candidate for making time delay measurements in order to calibrate the extra-galactic distance scale. One would expect that the relative brightness of the two components of the bright source would vary with time and cross-correlating the separate brightness variations would allow extraction of the time delay.

We are currently analyzing the data for a number of other lens systems and lens candidates. We hope to be able to use PSA for measuring the shape and position of lensing galaxies, for the detection of additional components, and for possible detection of structure associated with the QSO.

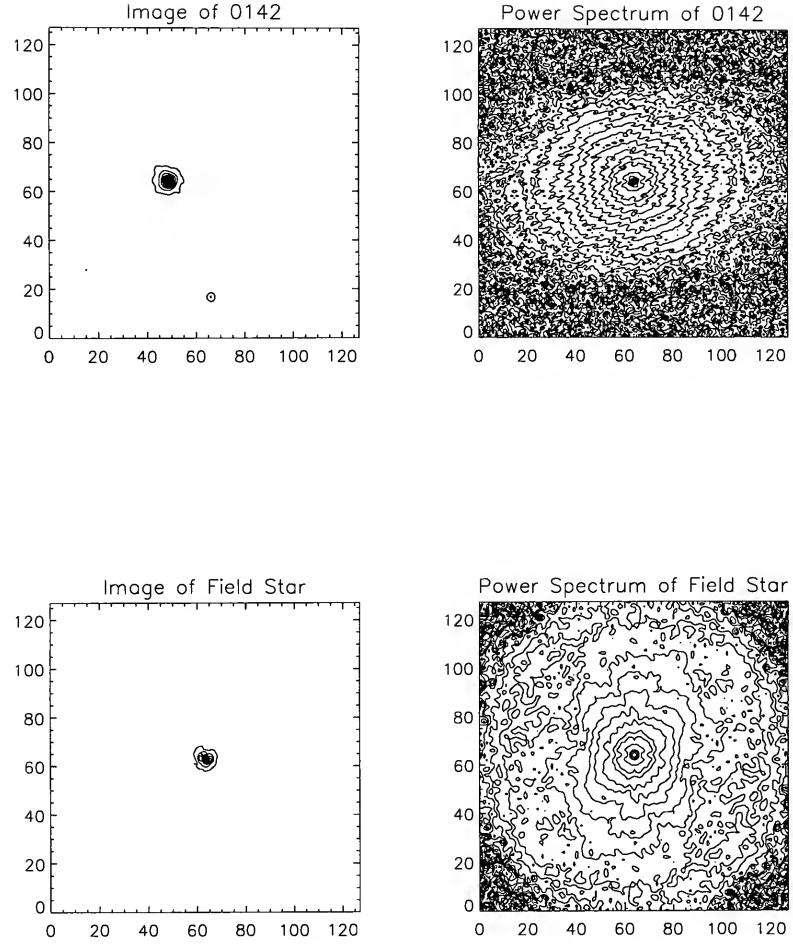


Figure 1. Images and power spectra for Q0142–100 and a field star.

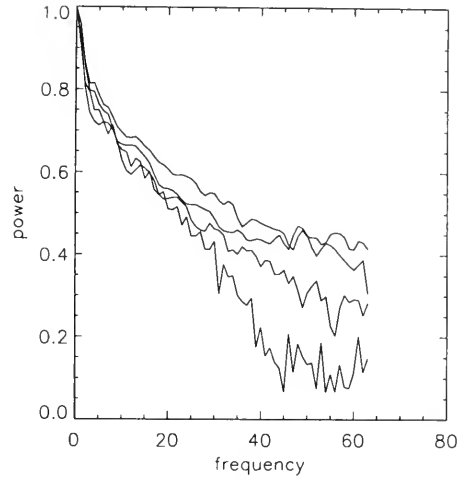


Figure 2. Cuts through Q0142–100 (lower curves) and field star (upper curves.

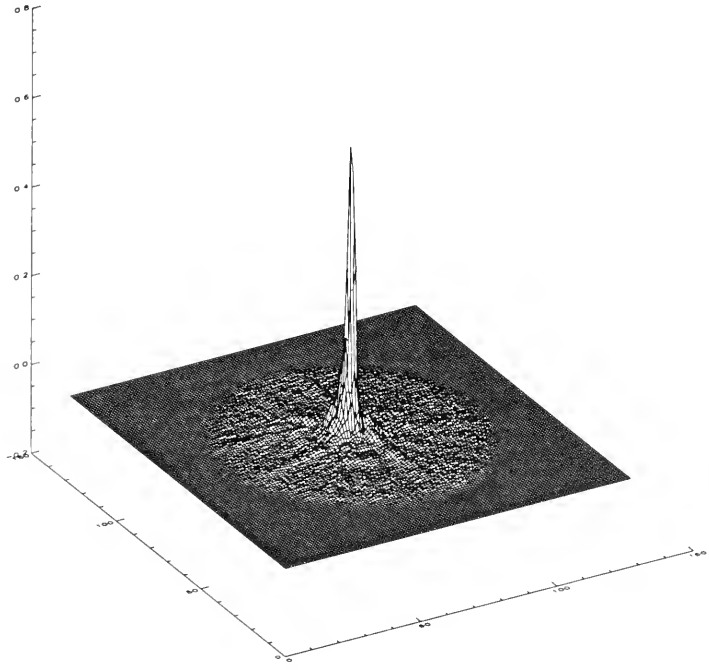


Figure 3. Residual image after subtraction of PSF from Q0142–100 image

4. Summary

Power Spectral Analysis could be an important tool for detecting and measuring a range of high spatial resolution features in HST images. It is particularly useful for the detection of close binaries, stellar asymmetries, and object diameters. Precise fitting to power spectra allows extraction of information and scales small compared to the sampling limit or even the telescope diffraction limit. Extraction of sub-regions of images would allow its use even for complicated objects like globular clusters. We have applied the technique to the analysis of gravitationally lensed QSOs and shown that some important scientific results can be obtained with it. A number of other scientific projects for which it would be useful include tracking the expanding nebula in SN1987a, detection of close binaries in globular clusters, and search for asymmetries in giant and supergiant stars.

Acknowledgments. This work was carried out with partial support from NASA grant NAGW-3577.

References

Falco, E. E. 1993, in *Gravitational Lenses in the Universe*, 31st Liege International Astrophysics Colloquium, J. Surdej, D. Fraipont-Caro, E. Gosset, S. Refsdal, & M. Remy, eds., Université de Liège, Liège, 127

Karovska, M., Koechlin, L., Nisenson, P. Papaliolios, C. & Standley, C. 1989 ApJ, 340, 435

Papaliolios, C., Karovska, M., Koechlin, L., Nisenson, P., Standley, C., & Heathcote, S. 1989, Nature, 338, 565

The Restoration of HST Images and Spectra II
Space Telescope Science Institute, 1994
R. J. Hanisch and R. L. White, eds.

Cramér-Rao Bound — Accuracy of HST Image Restoration

R. A. Gonsalves, T. S. Zaccheo, and S. M. Ebstein

Department of Electrical Engineering, Tufts University, Medford, MA 02155

P. Nisenson

Center for Astrophysics, Cambridge, MA 02128

Abstract. The Cramér-Rao bound is an analytical expression that describes the minimum obtainable mean square error associated with a given estimate of a parameter. This paper presents a compact and simple form of the bound for restoration techniques such as an inverse or a Wiener filter estimate. We present both 1-D and 2-D examples.

1. Introduction

The Cramér-Rao (CR) bound is an analytical expression that can be used to assign a lower limit to the minimum obtainable mean square error (MSE) associated with an estimate, $\hat{\mathbf{a}}$, of an unknown object, $\mathbf{a}$. In our formulation $\mathbf{a}$ is a vector of parameters and usually represents the unknown intensities of a 2-D object. The bound can be used for any restoration technique such as a Wiener filter estimate or an iterative algorithm like the conjugate gradient or Richardson-Lucy, provided certain statistics are known.

To calculate the bound one must know the probability density function (PDF) of the observed data, $\mathbf{d}$, given $\mathbf{a}$, namely, $p(\mathbf{d}|\mathbf{a})$, and the bias of the estimate. Typically, $p(\mathbf{d}|\mathbf{a})$ is Gaussian, Poisson, or some combination of both. In this paper we present results only for the Poisson case. The "given $\mathbf{a}$" assumption and the calculation of the bias are the source of two problems which we address in this paper.

Regarding the bias, we can usually calculate it for an estimate which is linear in the data, i.e., for estimates such as the inverse or Wiener filter. If the estimate is nonlinear in the data, the bias may be difficult to calculate.

Regarding the assumption that $\mathbf{a}$ is known, we assert that, of course, $\mathbf{a}$ is not known. Otherwise we would not want to estimate it. We provide calculations for the two obvious extremes: we calculate the bound as though $\mathbf{a}$ were indeed known, just to check the analyses; then we calculate the bound using the estimate $\hat{\mathbf{a}}$, as though it were the true value of $\mathbf{a}$. The latter approach provides a recipe for calculation of the bound in real estimation problems: one makes the estimate, then puts that estimate into an algorithm which calculates the bound.

2. The Cramér-Rao Bound

Analytic expressions for the Cramér-Rao bound are well known. The developments appear in texts such as Van Trees (1968), Whalen (1971), Kendall and Stuart (1973) and Snyder (1975). For the one-parameter problem where $\mathbf{a}$ is to be estimated by an unbiased estimate, $\hat{\mathbf{a}}$, the simplest possible case, the CR bound is

$$E\{(\hat{\mathbf{a}} - \mathbf{a})^2\} \geq 1/E\{(\partial L/\partial \mathbf{a})^2\}, \tag{1}$$

where

$$L = ln(p(\mathbf{d}|\mathbf{a})) \tag{2}$$

and E{ } denotes a statistical average over what is random in **d**. The left side of (1) is the mean square error for the estimate and the right side is the bound. No estimate for **a** can have a mean square error smaller than the bound.

From (1) we observe the following regarding changes in **a**: if the PDF change is large the error will be small; if, however, the PDF change is small the error is large. The CR bound estimates the sensitivity of the data to changes in **a**.

The multi-parameter, biased estimate requires calculation of the Fisher information matrix, **F**, a bias vector, **b**, and a correlation matrix **C**. They are defined by the elements

$$F_{ij} = E\left\{\frac{\partial L}{\partial a_i}\frac{\partial L}{\partial a_j}\right\}, \quad b_i = E\{\hat{a}_i\} - a_i \quad and \quad C_{ij} = \frac{\partial E\{\hat{a}_i\}}{\partial a_j}. \tag{3}$$

Then the CR bound on $\hat{a}_j$ is the j^{th} diagonal of the matrix **R**:

$$\mathbf{R} = \mathbf{C}\mathbf{F}^{-1}\mathbf{C}^T + \mathbf{b}\mathbf{b}^T. \tag{4}$$

The right hand matrix $\mathbf{CF^{-1}C^T}$ is commonly referred to as the "variance-covariance matrix" (Bury 1976).

3. CR Bound for Poisson Statistics

We model the data, **d**, as a set of independent, Poisson random variables with rate vector **g** given by

$$\mathbf{g} = \mathbf{P}\mathbf{a}, \tag{5}$$

where **P** is a distortion matrix. The PDF of **d**, given **a**, is

$$p(\mathbf{d}|\mathbf{a}) = \prod_{k=0}^{N} \frac{e^{-g_k} g_k^{d_k}}{d_k!} \tag{6}$$

and the **F** matrix has elements

$$F_{ij} = \sum_{k=0}^{N} \frac{p_{ki}p_{kj}}{g_k}. \tag{7}$$

If we define a diagonal matrix **Y** with diagonal elements $1/g_k$, **F** is

$$\mathbf{F} = \mathbf{P}^T\mathbf{Y}\mathbf{P}, \tag{8}$$

with inverse

$$\mathbf{F}^{-1} = \mathbf{Q}^T\mathbf{Y}^{-1}\mathbf{Q}, \tag{9}$$

where $\mathbf{Y}^{-1}$ is a diagonal matrix with elements g_k and **Q** is the inverse of **P**.

To evaluate **Q**, and $\mathbf{F}^{-1}$, we assume that **P** is a block circulant matrix, as would be the case if **P** represents a convolution of the object with a point response function and if we neglect the end effects. In this case **Q** is also a block circulant matrix and can be evaluated with Fourier techniques.

We are working towards the evaluation of Eq. (4). In the case where $\hat{\mathbf{a}}$ is an unbiased estimate, **b** is zero and the matrix **C** is an identity. Then $\mathbf{R} = \mathbf{F}^{-1}$ and the CR bound on $\hat{\mathbf{a}}$ is a vector with elements

$$e_i^2 = \sum_k (Q_{ki})^2 g_k. \tag{10}$$

This is also a convolution-like operation, but with the sum over index k, it is actually a correlation operation.

This equation provides a simple prescription for obtaining the bound associated with an unbiased estimator such as the inverse filter. Construct the inverse filter, square its elements in the spatial domain, and convolve the transpose of the resulting function with **g**. The only problem with this approach is that **g** is unknown. However, reasonable approximations can be obtained by using an estimate based on $\hat{\mathbf{a}}$.

If the estimate $\hat{\mathbf{a}}$ is biased and is given by

$$\hat{\mathbf{a}} = \mathbf{W}\mathbf{d}, \tag{11}$$

where **W** is any block circulant matrix, then we can show that

$$\epsilon_i^2 = \sum_k (W_{ki})^2 g_k + b_i^2. \tag{12}$$

The recipe for ϵ_i^2 is modified by the addition of the squared bias term and by the substitution of the filter W_{ki} for the inverse filter Q_{ki}.

In our examples we use a Wiener filter described in the spatial frequency domain by

$$W(\omega) = \frac{P^*(\omega)(1+z)}{|P(\omega)|^2 + z}, \tag{13}$$

where $P(\omega)$ is the Fourier transform of the point spread function, normalized to unit area, and z a noise control parameter.

4. 1-D Example

Fig. 1 (a) shows an input object, **a**, containing two pulses, each shaped by (350,700,350) counts in a background of 20 counts. This is convolved with the point spread function (PSF) **p** shown in (b). The observable data, **d**, is Poisson, driven by the rate vector **g**, the convolution of **a** with **p**. A typical realization is shown in (c).

Fig. 1 (d), (e) and (f) are estimates of **a** for an inverse filter, a Wiener filter, and the Richardson-Lucy (RL) algorithm. The noise control parameter in the Wiener filter was chosen to yield approximately the same bias (in the peaks) as the RL algorithm.

Fig. 2 shows the sample mean and RMSE (square root of the MSE) for 100 trials of the three estimation procedures. It also shows the CR bound calculations for the inverse and Wiener estimates. On the left are the sample means. We see that the inverse estimate is unbiased and that both the Wiener and RL estimates underestimate the peaks by about 20%. The latter two are biased estimates.

On the right are the sample RMSE's and two of the bounds. The bounds are smooth. We see that the bounds for the inverse and for the Wiener estimates closely match the sample RMSE's. This provides verification that the CR bounds are useful.

We present no bound for the RL algorithm because we do not know an analytic form for the bias. We note the general superiority of the RL estimate, which is known to be optimal for Poisson statistics. The RMSE is considerably smaller and the bias is about the same as the Wiener filter estimate.

Finally we address the (inappropriate) use of **a** in calculating the bound. Of course, **a** is unknown. We are compelled to use the estimate, $\hat{\mathbf{a}}$, of **a** in the calculation and this leads to errors. In Fig. 3 we show the effect. On the left is the bound for **a** known; on the right is the bound calculated for a single estimate, $\hat{\mathbf{a}}$. The latter has the general shape and amplitude of the former. We have not yet analyzed the differences. However, the bound on the right, which is actually calculable, seems to be useful.

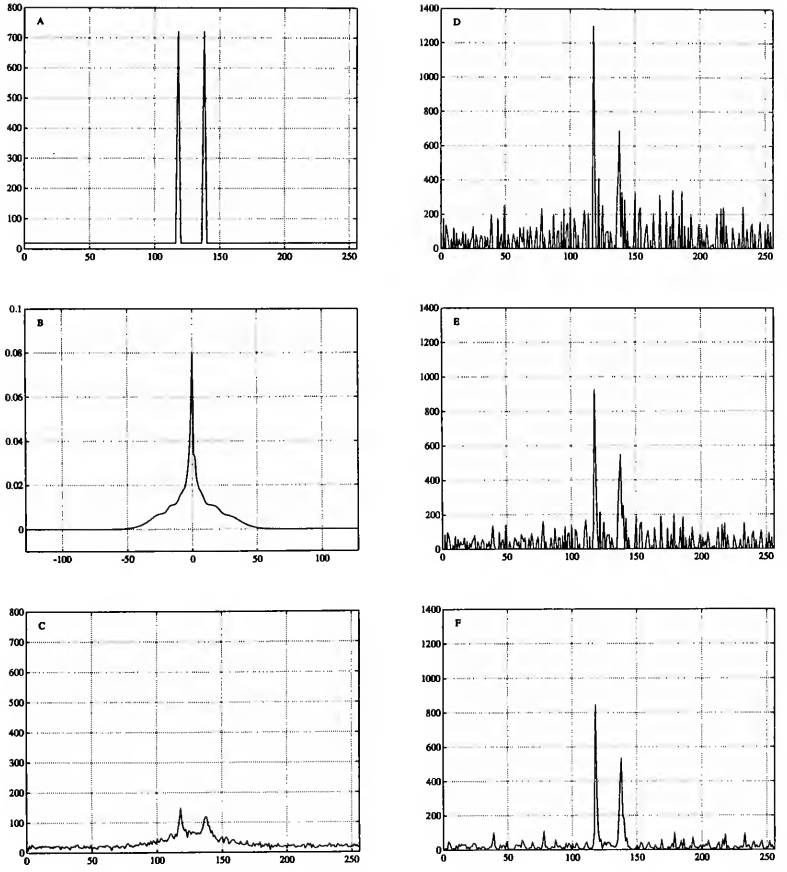


Figure 1. One dimensional data and typical estimates. (A) Original object containing two triangles each with a total count of 1400 on a background of 20 counts. (B) One dimensional blurring function obtained by integrating a Tiny TIM PSF. (C) Typical observed data corrupted by Poisson noise. (D) Typical inverse filter estimate. (E) Typical Wiener filter estimate. (F) Typical Richardson-Lucy estimate. The estimated data was clipped for presentation purposes.

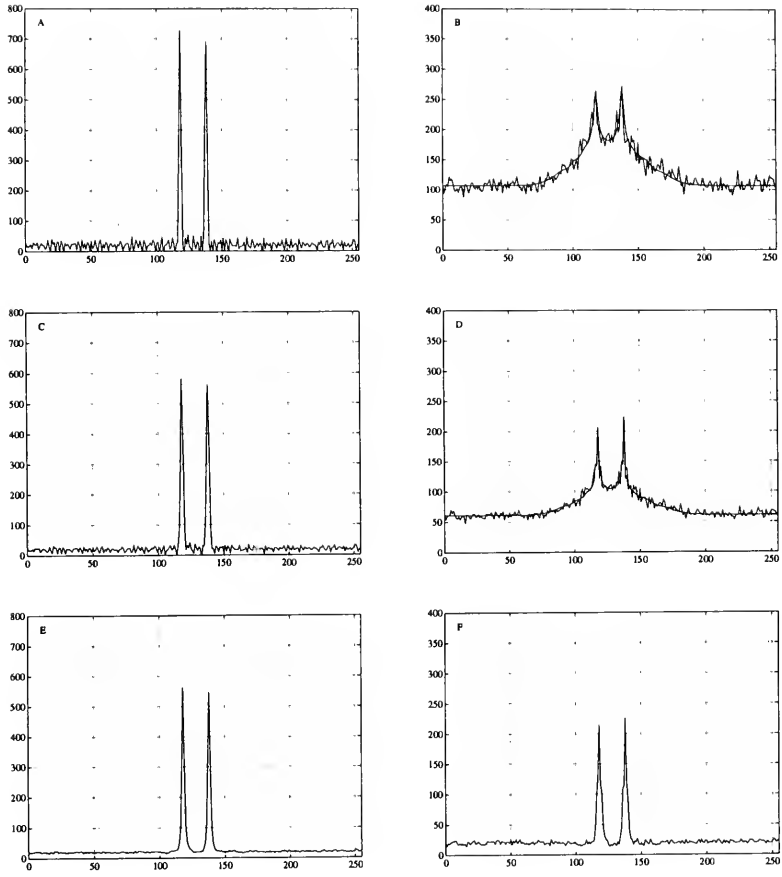


Figure 2. Sample statistics for 100 trials and CR bound calculations. (A) Sample mean for the inverse filter estimate. (B) Square root of the bound and sample RMSE for the inverse filter estimate. (C) Sample mean for the Wiener filter estimate. (D) Square root of the bound and sample RMSE for the Wiener filter estimate. (E) Sample mean for the Richardson-Lucy estimate. (F) Sample RMSE for the Richardson-Lucy estimate.

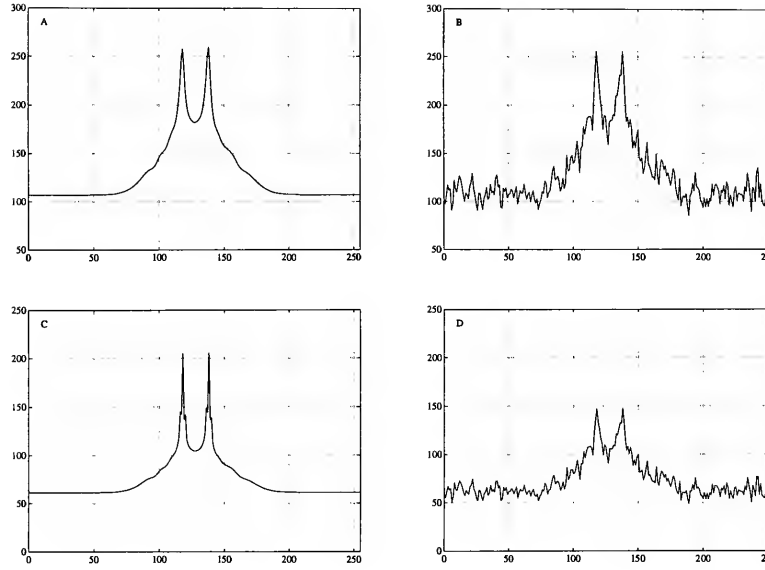


Figure 3. Theoretical CR bounds and estimates from typical data. (A) Theoretical RMSE bound for the inverse filter. (B) Approximate RMSE for the inverse filter, based on the estimate. (C) Theoretical RMSE bound for the Wiener filter. (D) Approximate RMSE for the Wiener filter, based on the estimate.

5. 2-D Example

Fig. 4 (a) is a ground-truth object, **a**, and (b) is a data image, **d**. It is **a**, convolved with a ST ScI-supplied PSF, **p**, and corrupted with Poisson noise. The max and min of **d** are 14,847 and 23. Fig. 4(c) shows a Wiener filter estimate, $\hat{\mathbf{a}}$, with z = 0.005. Fig. 4(d) shows the CR bound on the estimate. It was calculated based on the estimate, $\hat{\mathbf{a}}$, not on the ground-truth image, **a**.

Fig. 5 shows two 1-D cuts through the images and bounds. On the left are the sample RMSE's for 50 trials of the estimate. On the right is the CR bound for the first trial, based on the first estimate. We see that the bound does indeed provide a reasonable estimate for the MSE associated with the estimate.

6. Summary

We have presented a recipe for approximating the MSE associated with an estimate of an object based on blurred, Poisson data. We give an explicit formula , Eq. (12), for the bound when the bias in the estimate is known, as is the case with a Wiener filter estimate.

When the bias for an estimate, **b**, is unknown we can only suggest an approach. The suggestion is to find a Wiener filter estimate which yields the same residual noise power in data space as **b**, averaged over a local region. The composite Wiener-based bound may be a reasonable approximation to the exact calculation which requires knowledge of the bias. We intend to pursue this suggestion in future work; and to pursue a direct calculation of the analytic form of the bias for the RL algorithm, based on the residual error in data space.

References

Bury, K. V. 1976, *Statistical Models in Applied Science*, John Wiley & Son, New York

Kendall, M. G., & Stuart A. 1973, *The Advance Theory of Statistics (Vol. 2)*, Hafner, New York

Snyder, D. L. 1975, *Random Point Processing*, John Wiley & Son, New York

Van Trees, H. L. 1968, *Detection, Estimation, and Modulation Theory (Part 1)*, Wiley, New York

Whalen, A. D. 1971, *Detection of Signals in Noise*, Academic Press, San Diego

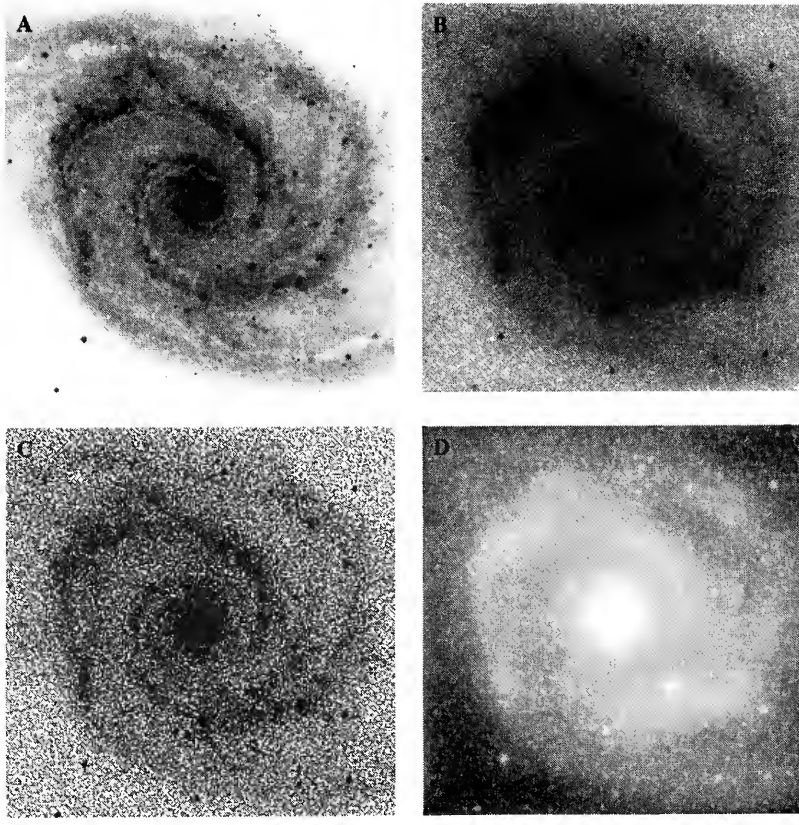


Figure 4. Two dimensional example. (A) Image of M51 obtained from ST ScI. (B) Blurred image corrupted by Poisson noise. (C) Typical Wiener estimate. (D) Cramér-Rao bound estimated from a single restoration.

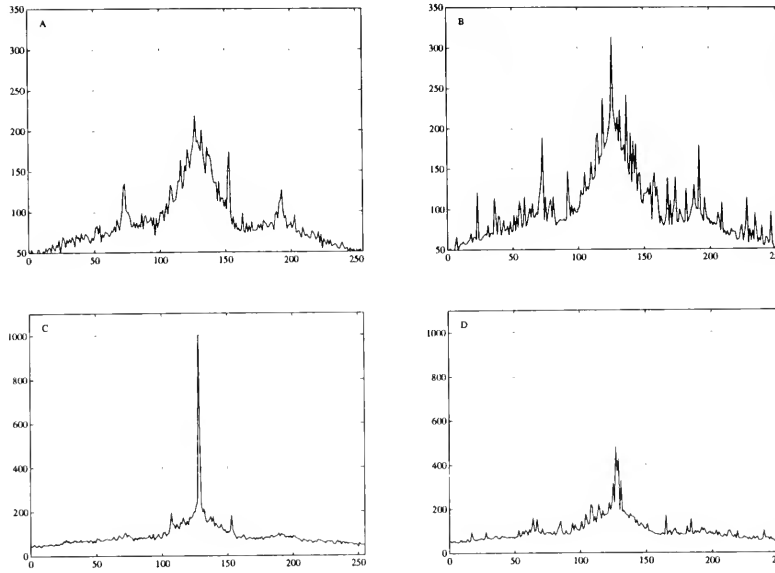


Figure 5. Sample RMSE for 50 trials and CR bound for a Wiener filter estimate; for two slices through the estimate. (A and C) Sample RMSE. (B and D) Corresponding bounds estimated from one trial.

The Restoration of HST Images and Spectra II
Space Telescope Science Institute, 1994
R. J. Hanisch and R. L. White, eds.

Imaging Performance of HST Compared to Future Ground-Based Adaptive Optics Systems

Adeline Caulet[1] and Wolfram Freudling

ST-ECF, Karl-Schwarzschild-Str. 2, Garching bei München, D-85748 Germany

Abstract. Simulations of a distant cluster of galaxies at $z \geq 2$ are presented for the HST cameras COSTAR-corrected FOC, WFPC II, the Advanced Camera, and for an adaptive optics system that may equip a future 8-meter class telescope. Except for the FOC and PC II, the faint objects are detected in 200 minutes, with identifiable morphological types. The Advanced Camera offers the biggest advantages due to its large field, great sensitivity, and high angular resolution.

1. New Generation Instruments

The first generation of imaging cameras FOC, WFPC I, and spectrographs FOS and GHRS suffer presently from OTA spherical aberration at the primary mirror. After the Servicing Repair Mission of December 1993 and the installation of the corrector system COSTAR, HST should become diffraction limited at wavelengths throughout the visible spectrum.

The second generation wide-field CCD camera WFPC II has essentially the same field of view (FOV) and pixel size as WFPC I with an overall improved efficiency due to the lower read noise, improved charge transfer efficiency, and intrinsically uniform and stable flat field. The other second generation instruments, to be installed in the mid-90's, are the infrared imaging camera NICMOS and the imaging spectrograph STIS.

Finally, a third generation imaging instrument, the Advanced Camera (AC), is being proposed for operation around 1999 to replace the existing cameras, and to offer the benefits of new technologies — megasize CCD arrays of high sensitivity from UV to near IR wavelengths. Towards the end of this decade, very large aperture (8 to 10 meter) ground-based telescopes will become available, working with active and/or adaptive optics systems that correct some of the image blur caused by atmospheric distortion. In this paper, we examine how these advanced instruments compare for the imaging of large extended faint objects.

2. Simulation of a Distant Cluster of Normal Galaxies

We have simulated the observations of a very distant cluster of galaxies at high redshift. Distant clusters of galaxies probe the evolution of galaxies and the large scale structure of the Universe at early epochs. The studies of clusters of galaxies at large redshifts will be important with the new generation instruments since detailed studies of the morphological composition of clusters and the evolution with look-back time of low surface brightness distant objects will be feasible.

The simulations are made for the HST instruments FOC, WFPC II, and the Advanced Camera, and for an 8-m telescope on the ground equipped with adaptive optics. The characteristics of the cameras and detectors are summarized in Table 1. A comparison of the fields of view of post-COSTAR HST imaging instruments with the Advanced Camera (AC) and the ground-based adaptive optics camera (AO) is shown in Fig. 1 superimposed on the noise-free cluster simulation.

---

[1] Affiliated to the Astrophysics Division, Space Science Department of the European Space Agency, F-75738 Paris, France

Table 1. Instrumental parameters used in the simulations.

Camera	F-ratio	array	pix size (μm)	pix size (milliarcsec)	FOV (arcsec)	readnoise (e^-)	darknoise (e^-/pix/hr)	sky (mag/arcsec^2)	qe% (at 5500Å)
WF I	12.9	800	15	100	80	13	108	23	50
WF II	12.9	800	15	100	80	7	36	23	35
PC II	28.3	800	15	46	36	7	36	23	35
FOC	151	512	24	14	7	0	2.3	23	10
AC	24	4096	15	54	220	3	22	23	90
GB-AO	–	2000	–	5	10	3	20	22	90

The Advanced Camera is ideally suited to mapping large scale structures due to its 200 arcsec FOV, i.e., 1.5 Mpc on the sky at $z \geq 2$. HST has the advantages of a space-based large telescope: a dark sky background and high angular resolution. The Airy disk (i.e., the resolution power) of the 2.4-m HST has a radius of 0.″058 at λ 5500 Å. From Table 1, it is seen that the detectors on the wide-field and planetary cameras undersample the PSF by a factor > 3. The pixel size for the proposed wide-field mode of the Advanced Camera is half critically sampled. The FOC detector fully samples the PSF. For the 8-m, the resolution power is 0.″017, and we have assumed a detector pixel of 0.″005.

2.1. Modelling the Cluster

The steps involved in the simulations are the following: (1) definition of the cluster parameters, (2) sampling the image to match the detector pixel size, (3) calculating the count rates (sky background, dark noise, galaxies), (4) PSF modelling and convolution, and (5) adding Poisson noise and read-out noise.

Cluster, Foreground Stars Using the package ARTDATA in IRAF, we have simulated a rich cluster of 1000 galaxies, adopting a Hubble law for the spatial density function, and a Schecter luminosity function. A uniform surface density of foreground faint stars was added. A scaling factor was applied to match the angular size of normal galaxies between z=2 and 3. A spiral disk of diameter 40 kpc at such a redshift extends over 6 arcsec (H_0=100 km s^{-1} Mpc^{-1}, q_0=0). The cluster extends over the whole AC field, and has a core radius of $\geq$ 100 kpc, typical of present-day regular clusters (Bahcall 1975). No evolution other than cosmological has been considered in the simulations.

Since we are simulating the same cluster viewed by cameras of different resolution power, the positions of the same objects were calculated for each detector pixel grid, and the task MKOBJECTS was run for each case.

Count Rates The adopted surface brightness averaged over the central 0.″1 × 0.″1 is 24.4 mag arcsec^{-2} for a normal bright spiral in this distant cluster, and 23.3 mag arcsec^{-2} for a normal elliptical. Such values are predicted for normal bright galaxies at z=2–3. The Lyα galaxy near the damped Lyα system toward the QSO PHL957 has V $\sim$ 26 mag arcsec^{-2} averaged over the whole galaxy, or V=23.6 mag integrated over 3x3 arcsec^2 (Lowenthal et al., 1991). The count rates per pixel for the sky and objects were derived using the relations given in the FOC (V4.0, p. 72, eqns. 4 and 5) and WFPC II (V1.0, p. 47) handbooks and the relation between the specific intensity of an extended source and the V surface brightness:

$$I^B(\lambda 5560) = 4.3 \times 10^{13} \times 10^{-0.4\times(V+AB_V)} \times \Omega \text{ (in phot cm}^{-2}\text{ s}^{-2}\text{ sr}^{-1}\text{)}\,, \qquad (1)$$

where $\Omega = 4.6 \times 10^{-15}$ sr is the solid angle of one FOC pixel.

Filter F480LP was used for FOC, and F555W for WFPC II. AB_V is $\sim$ 0 mag for sky and galaxies from the table of extinction coefficients as a function of spectral type and wavelengths (p. 49, V1.0 WFPC II handbook). The dark counts were added to the sky counts. The count rates for the AC and AO cameras were obtained using the WFPC II count rate formula, adjusted for the

DQE, angular resolution, and telescope aperture. Finally these simulated observations represent 5 coadded exposures of 40 minutes each.

PSF For the HST instruments, the PSFs were modelled using TINYTIM. The PSF for the center of the CCD was used in all cases. The PSF of the AC was assumed to be identical to the WFPC II PSF scaled by the pixel size ratio of the cameras (AC/WFPC II). For the adaptive optics camera, we have assumed an 8-m diffraction-limited core containing 20% of the flux, and a seeing-limited halo containing the rest. This AO PSF is modelled by the sum of two Gaussians: one narrow, high peak for the core, and one wide, low peak for the halo. A (very) good seeing of 0.″3 was assumed. This PSF is probably overly optimistic in the wavelength region considered here. The realization of such a diffraction-limited PSF with adaptive optics is currently anticipated only for the infrared domain.

Convolution with the original images was done using DCON, a FFT based fast image convolution task under IRAF provided by Richard Hook (1993).

Noise Poisson noise was added to the images using the task MKNOISE. The same task also added read noise appropriate for 5 co-added exposures (Table 1).

2.2. Comparison of the Simulations

The simulated observations are presented in the figures. Fig. 1 shows the whole field before PSF convolution. Fig. 2 (panels a–d) are the simulations obtained for WFPC II and for the AC, while Fig. 2 (panels e–f) show the simulations for the AO field. Simulations for FOC (not shown) show only noise for this distant cluster. For the large spiral seen in the field, the signal-to-noise ratio, averaged over 0.″1 × 0.″1 around the peak is $< S/N >= 4$ for the AC image (2×2 pixels), 0.6 for the AO image (20×20 pixels), 2.2 for the WFC II image (1 pixel) and 1.2 for the PC II (2×2 pixels). However, because of the high resolving power of the AO camera in the PSF core, S/N is effectively much higher near the object center, and an extended low surface brigthness spiral can be recognized in the adaptive optics image, whereas it looks like noise in the PC frame. This work shows the advantage of the AC camera over the other instruments. WFC I would also have detected this high redshift cluster with a lower S/N, but the aberrated PSF makes the morphological identification questionable (Freudling and Caulet 1993).

3. Conclusions

The comparisons of the simulated WFPC II and AC observations with the 8-m telescope observations show that, even with the assumption of very good conditions and an advanced state of adaptive optics technology on an 8-m class telescope, HST is still very competive for the observations of faint, extended objects. We did not use any image restoration for the simulated adaptive optics observations because the simulated cluster is too faint to benefit from such attempts. Because of the intrinsic FOV limitations of adaptive optics from the ground which are difficult to overcome by technological advances, the AC will have a unique capability of large FOV, high resolution deep imaging. Observations of high-redshift clusters are just one example of science programs which will be possible only with the HST, independent on the rate of progress in adaptive optics technology.

References

Bahcall, N.A. 1975, ApJ, 198, 249

Lowenthal, J. D., Hogan, C. J., Green, R. F., Caulet, A., Woodgate, B. E., Brown, L., & Foltz, C. B. 1991, ApJ, 377, L73

Hook, R. 1993, private communication

Freudling, W., & Caulet, A. 1993, in 5th ESO/ST-ECF Data Analysis Workshop, P. J. Grosbøl & R. C. E. de Ruijsscher, eds., ESO Conf. Proc. No. 47, p. 63

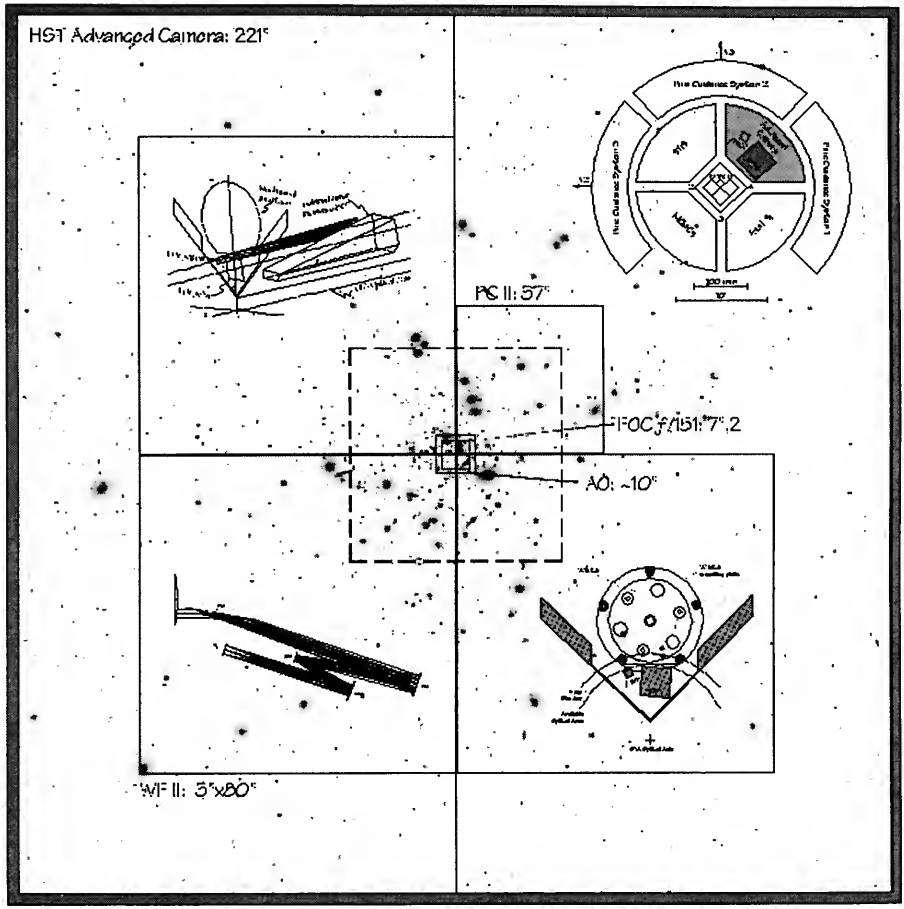


Figure 1. Simulation of a distant cluster of galaxies. The fields of view of the HST cameras are represented before PSF convolution.

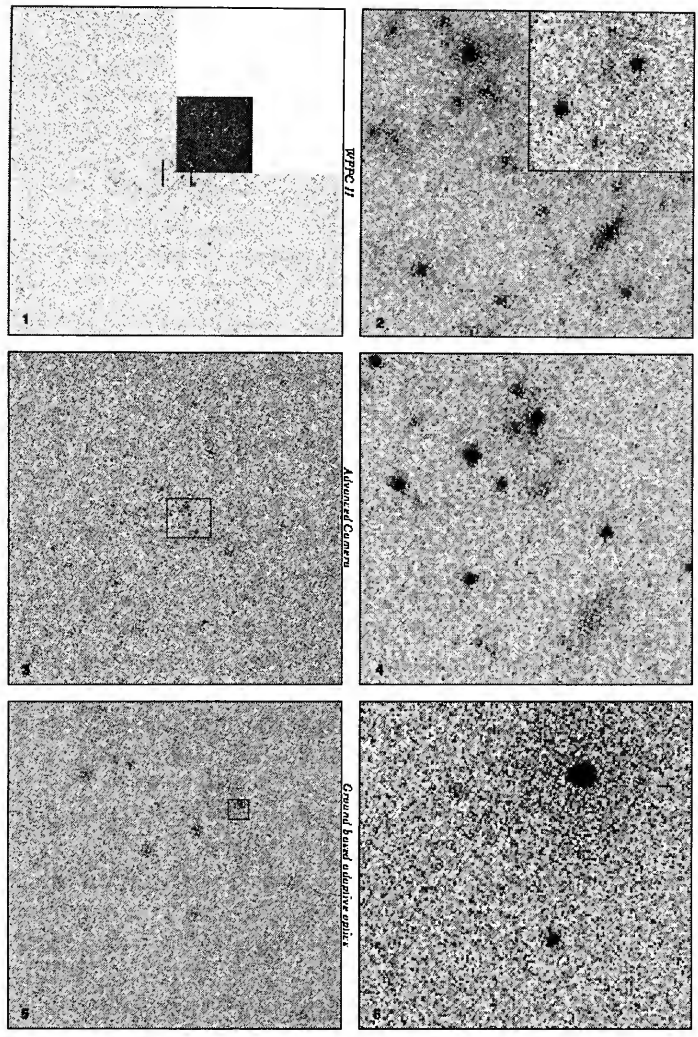


Figure 2. Simulations obtained for WFC II, PC II, the Advanced Camera, and the ground-based adaptive optics system. Only 1000 × 1000 pixels are shown here for the AC (e.g., the field within the dashed outline of Fig. 1). The PC II insert in frame a) has its intensity scaled by the square of the WFC/PC pixel size ratio (4.7). The enlargments of the image sections (boxes, left-hand column) in the right column have an image scale inversely proportional to the actual pixel size in milliarcsec, with the PC insert being magnified by a factor of ~ 2. The spiral mentioned in the text is seen towards the bottom right of frames b), d) and e).

Summing Up

The Restoration of HST Images and Spectra II
Space Telescope Science Institute, 1994
R. J. Hanisch and R. L. White, eds.

Scientific Results from Deconvolved Images

William C. Keel
Department of Physics and Astronomy, University of Alabama, Box 870324, Tuscaloosa, AL 35487

Abstract. I review various regimes in which deconvolution is and is not the technique of choice for analysis of HST images. The difficulty of getting adequately deconvolved images depends largely on required angular field, attainable signal-to-noise, and the dynamic range of the target; this last factor limits many interesting investigations in the presence of spherical aberration, exacerbating noise amplification and uncertain knowledge of the point-spread function. Photometric validation issues are also important; real data are used to show how well completely different approaches agree on the intensity profiles of faint galaxies.

Some examples of scientific results in the realm of galaxy structure and evolution are given, which have required deconvolved data. Some highlight are disk structures near galactic nuclei, fine structure in synchrotron jets, morphological evolution of medium- and high-redshift galaxies, and significant galaxy merging at moderate redshifts. An additional set of results has been greatly aided by deconvolution, including study of concentrated cores in galaxies and structural parameters of faint galaxies.

1. Introduction

Many of the presentations at this workshop have, quite properly, centered on problems in image restoration — mathematical underpinnings of algorithms, proper image bases for reconstruction, biases in measurement from deconvolved images. These might leave a depressing view of the prospects for getting scientific results from deconvolution, but the opposite is more nearly the case. I will review here some experience as to when image restoration is the right tool for the job, and describe some of the significant science that has come from analysis of deconvolved HST imagery.

Deconvolution is clearly a powerful analysis tool. The astronomical community's experience with HST and IRAS data has started to spread expertise in use of deconvolution techniques beyond its traditional base in aperture synthesis, and is leading to a more sophisticated view of data and measurement than was needed to interpret images blurred by the nearly Gaussian PSF of long-exposure images from the ground. In retrospect, we made it so long without them because eye, brain, and mathematics can deal with this Gaussian-like PSF very naturally; in many cases, we didn't even realize what visual processing was going on! Deconvolution algorithms are especially potent for astronomers as part of a software toolkit including image modelling, convolution, noise tracking, and measurement routines.

2. Practical Validation of Results

A major question in the minds of most astronomical users was discussed by Ivan King at this workshop — the issue of validation of the output from deconvolution. Astronomers are deeply concerned not only about morphology but with photometric integrity. It is of limited use to accurately distinguish thousands of stars in a restored cluster image if the positions and intensities are subject to unknown biases produced by the reconstruction procedures. Several presentation here have dealt with photometric issues for stellar images. For regimes useful in analyzing galaxy structure, existing techniques already yield sufficient accuracy to learn interesting and otherwise

inaccessible parameters. Fig. 1 shows an example of cross-validation, using data taken from Keel & Windhorst (1993). The image is of the faint radio galaxy 53W044 at redshift $z = 0.311$. The surface-brightness profile along the local major axis has been evaluated in three quite different ways. Direct measurements were made on deconvolved images, one made using the STSDAS implementation of the Lucy-Richardson algorithm run to $\chi^2 = 1$, and the other using a hybrid CLEAN with noise model. The CLEAN result was constrained to have a resolution equal to the Nyquist limit for sampling in the WFC, and the Lucy-Richardson result is demonstrably very close to the same effective resolution. For galaxies with smooth, symmetric structure, modelling allows an independent comparison. In this case, a family of models with various bulge:disk intensity ratio, and scale lengths for bulge and disk components, was generated and convolved with the empirical PSF (from a foreground star projected 5 arcseconds away). The model from this family that fit best in a χ^2 sense was numerically realized and measured in the same way as the deconvolved images. All three surface-brightness profiles track to within a few per cent at all radii over a dynamic range exceeding 100:1 (5 magnitudes). The three techniques could scarcely have more different basic principles and sensitivity to details of application, so this treatment is an empirical demonstration that HST, in its aberrated state, can deliver reliable surface-brightness profiles for galaxies at substantial redshifts; this means that we can do reliable quantitative studies of galaxy morphology.

The two-dimensional galaxy modelling used in this validation exercise introduces a crucial issue. When is deconvolution the analysis tool of choice, and just as important, when is it the wrong tool? Frequently, the science we seek is not directly in the image, however crisp. In imaging a star cluster, the desired result may be a Hertzsprung-Russell diagram; in observing the inner regions of an elliptical galaxy, the goal may be a high-resolution azimuthally averaged intensity profile. Problems of this kind, where the astrophysical background gives us strong *a priori* knowledge of the size or shape of the target objects, lend themselves to modelling and direct comparison with data in the observed domain, thus avoiding any possible biases from deconvolution. Specific tools for these applications exist, both for stars (Stetson 1994) and galaxies (Keel & Windhorst 1993, Windhorst et al. 1993b, Ratnatunga 1994). However, there are many problems for which the universe is not so cooperative: the targets may have unknown or complex structure. In these cases, deconvolution is the only way to retrieve a faithful representation of the object's properties. It may be the only way to match measurements made with various instruments. As a concrete example, if we observe some galaxy with a small aperture and the FOS, the only way to find what fraction of the galaxy's total light was included if an HST image is available involved deconvolving the image and reconvolving with a PSF appropriate to the wavelength range observed spectroscopically. Finally, producing faithful images for public release is not a trivial need, for use in education and in letting the taxpayers know they're still getting something for their money.

In seeing how different investigators use or avoid deconvolution algorithms, and in the choice for various problems, one can see some philosophical differences in how the results are approached, and what deconvolution is supposed to do. Some users see it as an operation on the data, approximating what would have been seen with a more favorable PSF, including noise and artifacts so that the quality of the processed data is immediately apparent. Other users want to go straight to "truth", and want to see a best-estimate model for the object with noise suppressed as irrelevant. Algorithms are certainly available to do both, and cases in between as well.

Before discussing specific science results, it is useful consider various regimes of deconvolution, where its practice may be affected by angular extent (in choice of instrument, mode, sampling, and extent of PSF changes) and signal-to-noise ratio (controlling the available dynamic range). For example, a planetary image with a good PSF needs only Fourier reconstruction (Cunningham & Anthony 1993; note that their reconstruction tests used only a space-invariant PSF and are thus applicable only to small objects), while this is completely unacceptable in noise properties for faint galaxies. As size increases and S/N drops, the degree of difficulty (measured in computational expense, investigator's time, trouble with subtle instrumental effects, and perhaps number of false starts) grows. Contrast issues are also important in determining which classes of problems are amenable to deconvolution. Fig. 2 shows where some well-known observations fall in these terms.

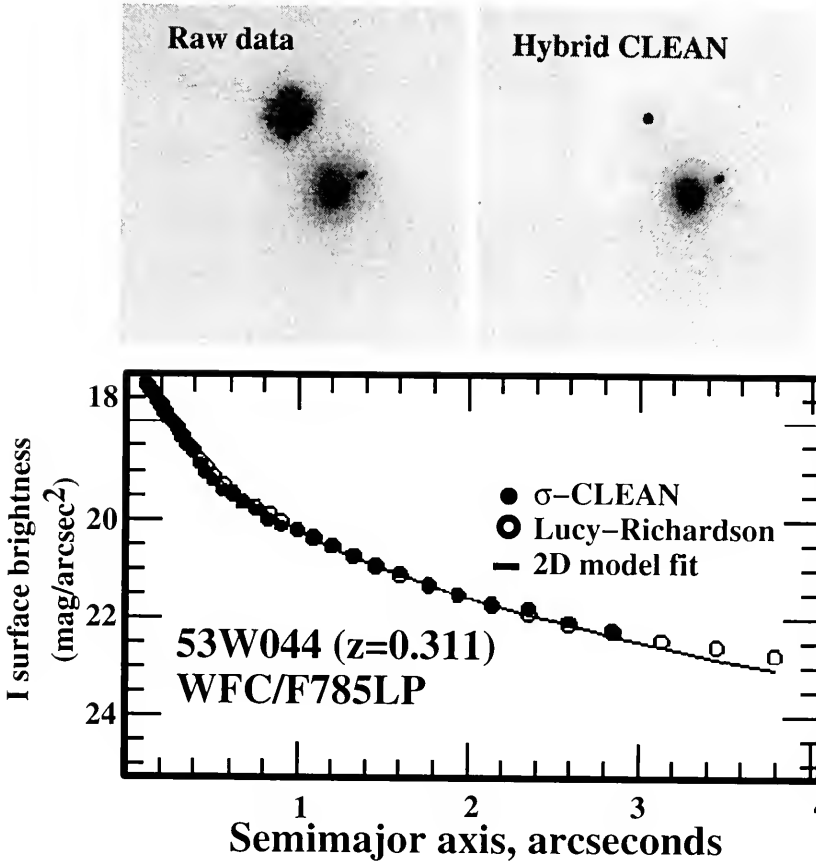


Figure 1. Comparison of surface-brightness profiles obtained through modelling and direct comparison with the aberrated data, and using both Lucy-Ruchardson and σ-CLEAN deconvolution.

The faint-galaxy results I will stress fall across the bottom in this diagram, which makes them relatively forgiving subjects for restoration; the dynamic range is limited by signal-to-noise ratio rather than PSF errors or sampling.

3. HST Science: Structure and Evolution of Galaxies

This could easily be a shopping list of HST's greatest hits. I will exercise the reviewer's arbitrary prerogative to limit the discussion to a field of particular interest where deconvolved images have had a profound impact, the structure and evolution of galaxies. This has been effective ground for deconvolution, because many targets have less contrast than, say, QSO host galaxies, and we are limited by S/N before PSF uncertainties or algorithmic problems. Many of these results require restoration only to the Nyquist limit of WFC sampling, sidestepping issues of undersampling.

There are many results that have required deconvolution, things we simply wouldn't know without this capability. The discussion moves roughly in order of typical distance.

3.1. Galactic Nuclei

These regions have proven fertile ground for imaging and deconvolution, being strongly condensed and bright. Numerous surprises have appeared in such images, especially of active nuclei.

It has long been suspected that active galactic nuclei are driven by central massive objects (massive black holes) surrounded by accretion disks, sometimes with jets arising along the rotation axis of the accretion disk. In the radio galaxies NGC 4261 and 3C 449 (Jaffe & Ford 1993), HST images have revealed dark disklike structures around the nuclei, oriented perpendicular to twin jets seen at centimeter wavelengths. These structures have characteristic dimensions of parsecs, much too large to be the accretion disks responsible for the emission properties of these nuclei and mass loss into the putative black holes, but may represent material which has already settled into orbits aligned with the yet-unseen inner disk.

Most active galactic nuclei fall into two types, classified on the basis of their emission-line spectra. Broad-line objects have emission from gas at a wide range of velocities, covering in the most extreme cases a range $0.1c$, while narrow-line objects show similar ionization and energy-input requirements but show only emission from lower-density gas with characteristic velocity of a few hundred km s^{-1} (which is also often present in broad-line objects). An important breakthrough was the recognition, driven by polarimetry, that some narrow-line active nuclei possess a dense inner broad-line region that is blocked from our line of sight, but seen by (polarized) scattered light (as reviewed by Antonucci 1993). This can fit nicely with an accretion-disk picture if the scattering region is a thick disk aligned with (perhaps outside) the accretion disk proper. Early PC observations of the nearby Seyfert galaxy NGC 1068 (Lynds et al. 1991) in fact resolved the continuum emission of the nucleus into a region of typical size 10 parsecs, perhaps a direct detection of this scattering region.

Some active nuclei exhibit jets that are detectable in the optical and ultraviolet. Most are dominated by synchrotron radiation, so that observations in this range are sensitive to highly relativistic electrons with relativistic Lorentz factors $\gamma \sim 10^6$ as well as to the details of the magnetic field structure in which they radiate. The FOC has proven very effective for these objects, since they have high surface brightness, small angular size, and greatest contrast against the background starlight in the blue wavelengths where the FOC is most sensitive. The well-known jet in M87 was a prime HST target, reported by Boksenberg et al. (1992) with the FOC and also observed with the PC in the deep red (e.g. Lauer et al. 1992a). Restoration shows many of the same features observed at 2 cm by the VLA (Biretta et al. 1983, not without its own level of deconvolution). Subtle differences between the radio and UV structures are particularly intriguing, offering the possibility of tracing the fine structures in which electrons are accelerated and respond to the magnetic field geometry. The subtlety of any such differences helps focus attention on the detailed behavior of different deconvolution algorithms, taxing fidelity of both intensity and geometry at low count rates. Observations of additional jets have revealed contrasts between rather smooth structure (PKS

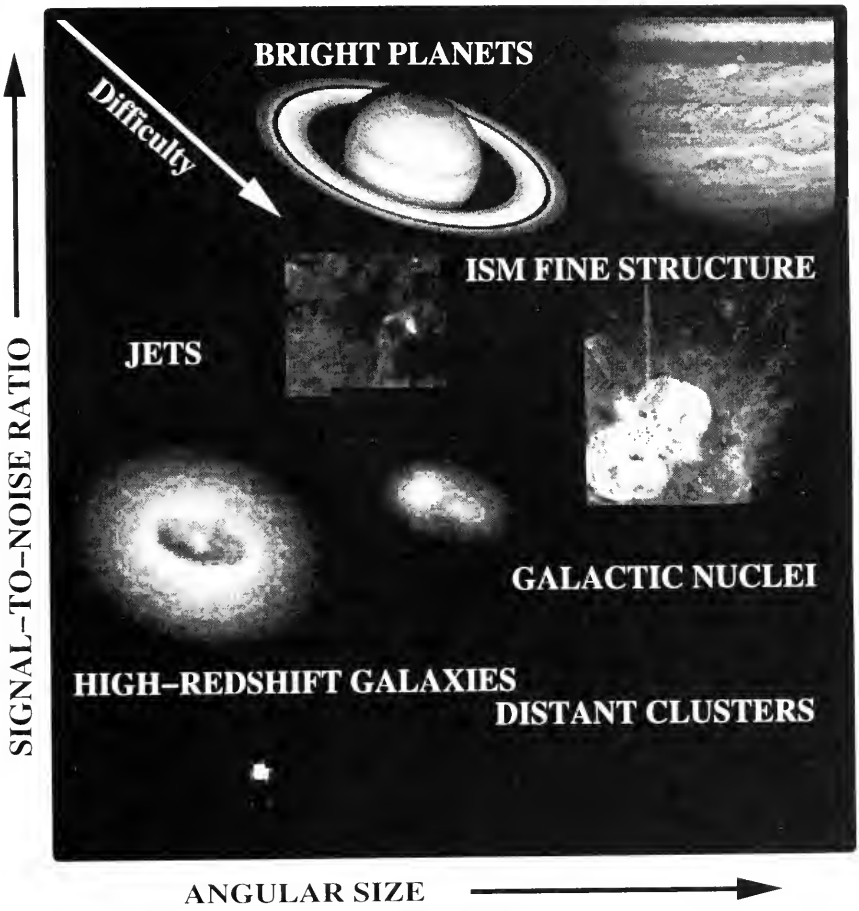


Figure 2. Schematic display of factors affecting the difficulty of image restoration of HST data. Angular size enters primarily through changes in the form of the PSF. Some well-known images are included at the appropriate locations to graphically illustrate the various regimes relevant to different scientific aims.

0521-36, Macchetto et al. 1991a) and hints of the kinds of filamentary features seen in parts of the M87 jet (3C 66B, Macchetto et al. 1991b). A further synchrotron jet was discovered serendipitously in NGC 3862 (Crane et al. 1993).

3.2. Galaxy Mergers and Star Formation

Ground-based data obtained at a variety of wavelengths have shown that bursts of star formation can be induced by strong tidal interactions and mergers. The high resolution allowed by deconvolved HST images has added a new richness to this picture, by showing very luminous blue star clusters. These have been seen now in the cases of NGC 1275 (Holtzmann et al. 1992), Arp 220 (Dowling & Shaya 1992), NGC 7252 (Whitmore et al. 1993), and a "quiet" merger in the compact group NGC 6027 (Seyfert's Sextet; Sulentic et al. 1994). The properties of these clusters open new windows on the history of mergers, with timescales and cluster ages deduced from their colors and in some cases spectra. These may be the precursors of globular clusters, an important issue in understanding the merger history of galaxies.

3.3. Morphology of Distant Galaxies

High angular resolution opens the possibility of seeing transformations in the forms of galaxies over cosmic time. For any but elliptical and S0 galaxies, morphological classification requires deconvolution rather than modelling with a manageable number of components. The galaxy content of rich clusters is a particularly rich field for inquiry; the color distributions of cluster members have long been known to show excess blue galaxies at redshifts $z > 0.3$ (the Butcher-Oemler effect; see Lavery, Pierce, & McClure 1992 and references therein). One mechanism for the color change is shown in a spectacular way by WFC imaging of the cluster 0939+4713 at $z = 0.4$ (lookback time 5.2×10^9 years for $H_0 = 50$ km s^{-1} Mpc^{-1} and $q_0 = 1/2$) presented by Dressler et al. (1994). While nearby rich clusters are uniformly dominated by E and S0 galaxies, all Hubble types are present in 0939+4713; many of the blue galaxies are structurally late-type spirals, along with a few apparent mergers. WFPC2 should allow similar observations to redshifts $z \sim 0.8$. The surprisingly rapid evolution of the galaxy population in clusters with cosmic time may be due to a combination of gas stripping from the hot intracluster medium and merging with subsequent starburst-driven galactic winds.

The ability to discriminate arbitrary structures of small angular scale has allowed detection of gravitationally lensed galaxies, as in the cluster AC 114 (Couch et al. 1992). This presents the possibility of measuring structure in extremely distant galaxies, using the "gravitational telescope" of deep cluster potentials. Distinguishing between even and odd-parity structure in such a lensed image will in principle allow distinction between features in the background galaxy and foreground lens, and may eventually allow us to probe structures on angular scales beyond the direct resolution of even corrected HST optics.

The observation of merging galaxies locally has led to the suggestion that mergers might be an important part of galaxy evolution, perhaps driving important bursts of star formation and nuclear activity, changing disk galaxies into ellipticals, and certainly changing the comoving space density of galaxies. The ability of deconvolved HST images to resolve fine structure at large redshifts allows a new test for this. Since mergers seen today are dominantly between galaxies that are gravitationally bound to one another, rather than from random encounters of unrelated galaxies, one signature of a strong role for mergers will be increasing numbers of galaxy pairs and tight groups wih increasing redshift and lookback time. Such an effect has indeed been reported, by Burkey et al. (1994) for pairs in a set of parallel WFC images, and by Casertano et al. (1994) for pairs and groups from the Medium-Deep Survey. Statistically well-defined criteria for such associations give membership rates of about 7% for nearby galaxies and 34–40% for faint galaxies (I magnitudes 21–24, redshifts $z = 0.3$–0.7), suggesting that the merger rate changes with redshift approximately as $(1 + z)^3$. This change is important in interpreting counts of faint galaxies, and is interestingly close to the rate at which QSOs and radio galaxies evolve with cosmic time.

Working to yet higher redshift has been more difficult, but feasible, with spherical aberration. Galaxies have been found with "normal" morphology and size out to $z = 2.4$, as in the case of the radio galaxy 53W002 (Windhorst et al. 1992). This galaxy has an elliptical-like intensity profile and scale length comparable to those of nearby ellipticals with radio sources. Seen at a lookback time only $1\text{–}2\times10^9$ years later than the peculiar galaxies at the highest redshift, such objects provide clues to the timescales of galaxy formation. So far, galaxies with strong central condensation have been reported out to $z = 2.4$. Further imaging studies will be crucial in understanding whether this epoch is when most galaxies approach their present dynamically relaxed structures. We may already be seeing galaxy formation, if we but had a clue what the process looks like.

The highest-redshift galaxies known are the very powerful radio sources found from the 3C and 4C surveys. At redshifts as high as $z = 3.8$, they show spectacular, if poorly-understood, morphologies. Their optical (that is, emitted UV) images are frequently elongated and clumpy. The elongation is correlated with the direction of the radio structure in the so-called alignment effect, which is in evidence only for $z > 0.5$. HST images show striking correlations in detail between emitted ultraviolet and radio structures, in the cases of 4C 41.17 ($z = 3.8$, Miley et al. 1992) and 4C 28.58 ($z = 2.9$, Miley 1993). Connecting these early views of galaxy evolution with what we see at lower redshifts may be the outstanding problem here. Are these systems rare examples of galaxian pathology, irrelevant to the evolution of galaxies in general, or are they symptomatic of many galaxies during a turbulent initial epoch? Once again, only the ability to image larger samples to fainter levels will clarify these matters. It is intriguing that we see symmetric, centrally condensed galaxies at redshifts less than 2.5, and that the peculiar elongated and clumpy systems occur mostly at yet higher redshifts; one might be tempted to call this an evolutionary pathway, invoking astronomers' well-known ability to generalize from a sample of two or more.

4. Science Aided by Deconvolution

There is another category of scientific results where deconvolution plays a significant role, even when the final measurements use another technique. This operates on the principle that if you don't know it's there, you can't measure it. In many cases, deconvolved images are important in knowing what analysis technique to use for the problem at hand, and whether in fact another technique might be more appropriate. Several examples come to mind from the realm of galaxy structure and evolution.

4.1. Central Peaks in Galaxy Profiles

The presence of strong central concentrations in the starlight of galaxies is important in understanding their dynamics, and in searching for the massive compact objects thought to exist in (at least) active galactic nuclei. They have been probed using combinations of modelling and deconvolution in several galaxies. From an early science verification image, Lauer et al. 1991 found a strong central peak in the previously undistinguished galaxy NGC 7457. Further PC observations confirmed ground-based measurements of a central "spike" in M87 (Lauer et al. 1992a), and showed that the central concentration is more pronounced that could have been demonstrated from ground-based data alone. A similar concentration also appears in the Local Group elliptical M32 (Lauer et al. 1992b). Pending spectroscopic observations post-COSTAR, galaxies with these features are the best candidates for hosting massive central black holes. It is noteworthy that this signature is not uniquely linked to nuclear activity, suggesting that the prerequisites for such activity are more common than its actual occurrence.

The core of the Andromeda galaxy, M31, has also been suspected to harbor a massive central object from ground-based spectroscopic data (Dressler & Richstone, 1988, Kormendy 1988). The HST images analyzed by Lauer et al. (1993) show an unexpectedly complex situation — two cores of different scale size and brightness, separated by 0.49 arcseconds. This discovery accounts for a long-standing puzzle, the offset of the brightness peak of M31 from the isophotal center (Nieto et al. 1986). This nucleus was observed using the *Stratoscope II* balloon-borne telescope (Light

et al. 1974), which showed this central asymmetry; had their image been somewhat deeper the character of the nucleus might have become apparent. The double core of M31 poses outstanding theoretical questions. The most likely explanation is that the more compact object is the remnant core of a galaxy that has merged with M31; to have such a configuration last long enough to have a reasonable chance of our observing it probably requires that both nuclei harbor massive black holes. These conclusions could have been reached from modeling and fitting alone, but deconvolution was a strong aid in knowing what to fit.

4.2. Modelling Galaxy Images

For more distant galaxies, structural parameters are often better estimated by model-fitting (Schade & Elson 1993) — once deconvolution has given assurance that there is not too much nonaxisymmetric structure. Even in fitting multicomponent models to the aberrated data in a χ^2 sense, a deconvolved image can give initial values for the parameters that vastly speeds convergence to the best-fit set (Windhorst et al. 1993a, 1993b). Results from such procedures include an angular-size – redshift relation for field galaxies (Griffiths et al. 1994) and identification of weak disks in early-type galaxies (Keel & Windhorst 1993).

5. Current Challenges

Many of the successes listed above are relatively forgiving targets for image reconstruction. In faint-galaxy work, we are limited more often by signal-to-noise ratio than by knowledge of the PSF or residual effects from the restoration. For galactic nuclei, the data quality is usually high enough that some noise amplification can be tolerated. Some of the most challenging restoration tasks are in the regime of high dynamic range, exemplified by QSO host galaxies, gravitational lenses, stellar ejecta, and even comets. In all these cases, reconstruction is vulnerable to small errors in the PSF, and the inevitable photon noise in low-signal regions is dominated by the Poisson noise from aberrated photons properly belonging to bright parts of the image. This problem is particularly severe for the kinds of problems just listed — where the primary interest is in faint extended structure close to very bright (usually unresolved) sources. Here, noise amplification can destroy the interesting signal, since the noise is set by statistics unrelated to the true local brightness. Nisenson (1994) has described in these proceedings some indirect analysis methods for gravitational lenses; early image-restoration work on the multiply imaged QSO PG 1115+080 (Groth et al. 1991; see also Kristian et al. 1993) amply illustrated how sensitive such results can be to both noise and PSF accuracy.

The issue of PSF accuracy has been addressed in several contributions. Analysis of existing imagery has already run into the limits of present calculations (with TinyTIM) and empirical libraries in several instances. Crowded-field photometry has reached a precision limited by the our ability to track changes in the PSF across the WF/PC field. Empirical libraries, though guaranteed to have the right optical mapping, are usually too sparsely sampled near the edges of the WF/PC chips, and break down for objects with extreme spectral shapes (including strong emission lines). This makes both wide field and high dynamic range restorations some of the most stringent tests for image restoration techniques, and just as important of our ability to properly compute the appropriate PSF.

6. Where Are We Going Now?

What will optical astronomers be doing with deconvolution in the coming years, especially with HST? Algorithms and PSF characterization will only improve with experience, so that we will be able to do a better job on data already in the archive. The amount of data already taken insures a rich harvest for the taking as our understanding of the instruments and optics improves. Further, as we have heard here, the post-COSTAR images will still display residual aberrations that can be corrected by the techniques that were utterly necessary for handling the aberrations in the first

place. Crudely put, there will be little excuse for seeing even diffraction rings in fully processed post-COSTAR images!

What have optical astronomers learned from handling HST imagery? Any experience that forces more of the community to understand the data-taking and data analysis processes more completely is to that extent a positive step, albeit in this case not one that any of us would have chosen. Many optical astronomers have now followed their colleagues in radio interferometry into regular and informed use of deconvolution algorithms. The HST optics have driven a significant amount of work in algorithm development and in understanding their theoretical underpinnings; comparison of the presentations here with the first image-restoration workshop shows how much these applications have matured. More astronomical results than ever before are benefiting from restoration, and we can look forward to seeing more complete analyses of hard-won data as a result. Finally, prospects for adaptive optics on the ground open new avenues for science through image restoration. Systems in the optical are likely to produce PSFs with a sharp core and significant diffuse halo; if the PSF can be determined locally, this reduces to something that is, in the mathematicians' terms, "a problem previously solved."

References

Antonucci, R. R. J. 1993, ARAA 31, 473

Biretta, J. A., Owen, F. N., & Hardee, P. E. 1983, ApJ, 274, L27

Boksenberg, A. et al. 1992, A&A 261, 393

Burkey, J. M., Keel, W. C., Windhorst, R. A., & Franklin, B. E. 1994, ApJ, submitted

Casertano, S. et al. 1994, BAAS, 25(4), 1399

Couch, W., Ellis, R., & Sharples, R. 1992, ST-ECF Newsletter 17, 13

Crane, P., et al. 1993, ApJ, 402, L37

Cunningham, C. C. & Anthony, D. 1993, Icarus 102, 307

Dowling, D. & Shaya, E. 1992, BAAS, 24, 728

Dressler, A. & Richstone, D. O. 1988, ApJ, 324, 701

Dressler, A. et al. 1994, ApJ, submitted

Griffiths, R. E. et al. 1994, ApJ, submitted

Groth, E. J. et al. 1991, in *The First Year of HST Observations*, A. L. Kinney & J. C. Blades, eds., Space Telescope Science Institute, Baltimore, 192

Holtzmann, J. et al. 1992, AJ, 103, 691

Jaffe, W. & Ford, H. 1993, reported in Maran, S. P., & Kinney, A. L. 1993, PASP, 105, 447

Keel, W. C. & Windhorst, R. A. 1993, AJ, 106, 455

Kormendy, J. 1988, ApJ, 325, 128

Kristian, J. et al. 1993, AJ, 106, 1330

Lauer, T. R. et al. 1991, ApJ, 369, L41

Lauer, T. R. et al. 1992a, AJ, 103, 703

Lauer, T. R. et al. 1992b, AJ, 104, 552

Lauer, T. R. et al. 1993, AJ, 106, 1436

Lavery, R. J., Pierce, M. J., & McClure, R. D. 1992, AJ, 104, 2067

Light, E. S., Danielson, R. E., & Schwarzschild, M. 1974, ApJ, 194, 257

Lynds, C. R. et al. 1991, ApJ, 369, L31

Macchetto, F. et al. 1991a, ApJ, 369, L55

Macchetto, F. et al. 1991b, ApJ, 373, L55

Miley, G. K., Chambers, K. C., van Breugel, W. J. M., & Macchetto, F. 1992, ApJ, 401, L69

Miley, G. K. 1993, in *Science with the Hubble Space Telescope*, ESO Conference and Workshop Proceedings, 44, P. Benvenuti & E. Schreier, eds., European Southern Observatory, Munich, 1

Nisenson, P., et al. 1994, this volume

Ratnatunga, K., et al. 1994, this volume

Schade, D. J. & Elson, R. A. W. 1993, AJ, 105, 1581

Stetson, P. 1994, this volume

Sulentic, J. W., Arp. H., & Rabaça, C. 1994, ApJ, submitted

Whitmore, B. C., Schweizer, F., Leitherer, C., Borne, K., & Robert, C. 1993, AJ, 106, 1354

Windhorst, R. A., Mathis, D. F., & Keel, W. C. 1992, ApJ, 400, L1

Windhorst, R. A., Gordon, J. M., Mathis, D. F., Neuschaefer, L. W., Pasquerelle, S. M., Schmidtke, P. C., Ferro, A. J., Keel, W. C., & Burkey, J. M. 1993a, ApJ, submitted

Windhorst, R. A. et al. 1993b, ApJ, submitted

The Restoration of HST Images and Spectra II
Space Telescope Science Institute, 1994
R. J. Hanisch and R. L. White, eds.

Where Have We Been, Where Are We Now, Where Are We Going?

T. J. Cornwell

National Radio Astronomy Observatory, Socorro, NM, 87801

Abstract. Three years ago, STScI hosted the predecessor of this meeting. I attempt to contrast our understanding of HST deconvolution now with then. The papers presented here can be split into two categories: deconvolution algorithms in general, and tricky HST-dependent details. I discuss how we have progressed in these two areas and where we might expect to be at a putative next meeting in the series, three years from now.

1. Introduction

It is a pleasure to be here at the Institute for the second meeting on Image Restoration of HST Images and Spectra. The last time I was in this building was at the first such meeting about three years ago, some months after the discovery of the problems with the HST optics. In the months between those two events, I participated in a panel to advise NASA on the potential of image processing for correcting images for the effects of the spherical aberration. My first reaction on hearing of the problems was that, yes, deconvolution as we do it in radio astronomy could be a substantial help, given enough photons and knowledge of the point spread function. Well, it turned out that the HST deconvolution problem was very tricky in a number of different ways. At the first workshop, I think there was a tendency to minimize the difficulties and to concentrate upon debates as to which was the best algorithm to use. At this workshop, I see less interest in disputes over the merits of Richardson-Lucy versus MEM and more interest in tackling the difficult aspects of the data. In this summary, I want to discuss our deepening appreciation of deconvolution algorithms, then how we did with the tricky stuff, and then conclude by addressing the questions in the title of this talk. Before any of this, I want to record some noteworthy (and telling) comments from attendees.

2. Some Comments from Attendees

I heard the following statements about Image Restoration. Some of these were made by one person more or less in the words I put down here, and some were recurrent themes in the meeting. My commentary follows:

- It's often not necessary. (*King*)
  - *This is true if one has a well-posed scientific question that can be posed directly to the data (i.e., to the aberrated image). If not, then one often has to look at an image for qualitative information (e.g., what does the object look like?)*
- It's too mysterious. (*King*)
  - *I sympathize with this view. Many forms of deconvolution algorithm are too mysterious. For example, just what does the Richardson-Lucy algorithm assume about the sky? More on this below.*
- It's impossible. (*Anon.*)

- *What's impossible? Evidently, to find a solution to the convolution equation $g = Hf$ when the PSF H is singular. Well, I think we all know that by now. The real point is what extra information do we need to get a unique solution, and what are the properties of that solution.*

- It's too expensive. (*many people*)
  - *Most processing described at this meeting can be done on a Sparc IPX. To run HST costs about 1 Sparc IPX per minute.*

- It's too hard. (*many people*)
  - *The technical obstacles are formidable: a spatially variant and poorly known PSF, strange noise, cosmic rays, undersampling, field rotation, etc. A better statement would be that it's very hard but not too hard.*

3. Deconvolution

As the risk of offending by over-simplifying, the deconvolution algorithms presented here can be classified in the following way:

Classic inverse filter. The inverse filter doesn't work, of course, but one can sneak up on it by using various iterative schemes that eventually converge (or, more appropriately, diverge) to the inverse filter. Sneaking up is not quite the right phrase; perhaps knowing when to slam on the brakes expresses the danger of this approach more accurately. One subtle point is that if the cue to hit the brakes comes from the data, then the method is no longer linear and hence loses most of its attraction (*Coggins*).

Richardson-Lucy algorithm. RL is known to be a Maximum Likelihood estimator for photon noise. What else do we know about it? Well, it also needs the brakes slammed on at the right moment (*White*). Also, we don't actually understand the nature of the extra information used by RL to generate a unique solution. This is very similar to the situation with the CLEAN algorithm used in radio interferometry. CLEAN is quite mysterious. It seems to work well, but no one can describe the properties of the solution generated by CLEAN. The same is true of RL. Both RL and CLEAN solve Maximum-Likelihood equations via an iterative scheme that must impose some bias towards a particular type of solution. The problem is that we cannot quantify that bias or what the preferred type of solution is.

Regularized deconvolution. It's been my experience that regularization is loved by mathematicians and Bayesian methods are loved by physicists. This split is somewhat meaningless since the equations to be solved are often much the same and therefore the results are often similar.

Bayesian methods. There were two interesting developments in Bayesian methods presented at this meeting. First, we saw the marriage of Bayesian priors with photon noise to yield an improved RL algorithm (thus addressing my objection to RL above) (*Molina*). Second, we heard about the *Pixon* (*Puetter*). The idea behind the Pixon is quite simple and beautiful: it says that since there is nothing magical about the basis function used for expressing an image — one should allow it to be dictated by the data. This has the self-evident quality of many smart ideas. However, I think we need to know more about how exactly a Pixon basis is to be chosen. This seems to be another place where heuristics can creep in, which is ironic if one considers a goal of Bayesian methods to be the elimination of *ad hoc* tricks.

Blind deconvolution. I think that, for me, one of the most interesting and satisfying topics at this conference was blind deconvolution: recovery of the PSF function as well as an image of the sky (*Christou, Schulz*). This is personally satisfying because I have worked on more or less the

same problem in radio interferometry. There we know some constraints on the PSF (mainly the known spatial frequency content and also that the closure relations have to be obeyed) and we can add sufficient prior information about the sky (e.g., positivity and emptiness) to allow recovery of both PSF and sky brightness. In the application to the HST, blind deconvolution allows recovery of the PSF in cases where the calculated PSF is insufficiently accurate. There remain many issues to be settled, though:

- The starting point for the algorithms presented seems to be too important.
- What is the role of prior information in constraining the solution? More information should resolve the starting point problem.
- One always gains by adding more information in the form of known imaging physics. For example, one can ask whether the closure relations are enforced? Alternatively, is the PSF forced to be the Fourier transform of a cross-correlation of a pupil plane with phase errors? If not, then it should be.

4. Tricky stuff

Turning now to the tricky instrumental stuff, what has been solved since the last meeting? First, I'll discuss instrument-related problems, then I'll turn to algorithms.

PSF poorly known: Solved. Either find PSF by blind deconvolution or by using one of the new PSF prediction codes (*Redding*).

Under-sampling of focal-plane: Not much said here.

Signal-to-noise ratio: Solved. Strangely, this was a big worry three years ago. I think that this represents a set of then-unresolved concerns over the effect of the PSF halo on imaging weak sources. We now know that for weak, point-like sources the answer is to simply increase the observing time proportionately to the loss in Strehl ratio. For complex sources, we know that the answer is much more complicated.

Cosmic rays: Solved. The signature of cosmic rays is that no PSF halo is seen. Working algorithms to exploit this fact exist (*Adorf*). Ironically, it seems that this signature will disappear after the servicing mission (*Biretta*).

Readout noise: Solved. Several algorithms now account for readout noise in the formal description of the measurement equation (*Núñez and Llacer; Snyder*).

Field rotation: Solved. Joint deconvolution of several epochs does the trick. The coupling with undersampling must still be a problem, though it was not discussed as such.

Object color variations: Not solved. Since the PSF changes with wavelength, objects of different colors observed with a wide-band filter will have different PSFs. This seems to have vanished from consideration.

In the area of algorithms, we have made considerable progress:

Photometry bias: Solved. Some of the deconvolution algorithms can be forced to have less bias (*Lucy*) or one can use a photometry package (e.g., DAOPHOT) directly (*Stetson*).

Spatially variant PSF: Not solved. Some progress in understanding but no well-used working algorithms. (*Hunt and McNown; Adorf*). From my experience tackling an SVPSF in radio interferometry, I would advise investigators to look at a number of different approaches carefully before implementing any one method. The computational issues tend to dominate and one has to think carefully about how to best solve the relevant equations. In radio

interferometry, although the equations become more tractable by embedding the imaging in a three-dimensional space, the optimum computational solution comes from using local approximations to the PSF.

Null-space stuff, e.g., mottling in RL: Solved. The RL algorithm produces mottled images if allowed to iterate too long. There seem to be a number of *ad hoc* ways around this problem (*White*). CLEAN also produces strange stuff in the null-space (i.e., unsampled regions of Fourier space) but the main effect seems to be easily-recognizable stripes in the image plane.

Performance limits: Solved. The Cramér-Rao bound work is very interesting (*Gonsalves*). Also, the incorporation of the error analysis into deconvolution looks promising (*Bouyoucef, Roques, and Fraix-Burnet*).

Multi-channeling: Solved. Both MEM and RL can handle multiple constraint images.

5. Where Have We Been? Where Are We Now? Where Are We Going?

A conference summary necessarily glosses over or outright ignores a lot of interesting work. For that I apologize. However, I'm going to compound any such error by attempting to summarize our past, present and future even more briefly:

3 years ago

- Only one algorithm was required and it was either RL or MEM, depending on who was talking.
- Everyone thought that the PSF was knowable a priori.
- As a community we had little experience with the deconvolution of HST data.

Now

- We now know that from zero to many algorithms might be applicable.
- We have seen that it pays to view the PSF an unknown to be determined.
- As a community, we now have lots of experience of HST deconvolution.

3 years from now?

- Our model of the telescope will be more complete: including SVPSF and a better noise model.
- As our understanding of blind deconvolution improves, we will realize that it pays to view the aberrations (rather than the PSF) as unknowns to be determined.
- We will have produced lots of documentation on deconvolution.
- Post-servicing, the HST will have a better PSF. However, as a community, we'll still have a lot to do. The PSF for WFPCII is not perfect and varies across the field.

Finally, I should say that ST ScI has done a great job in pushing the cause of image restoration in the community via software, algorithms, user support, newsletters and via these meetings. Personally, I have enjoyed both these meetings tremendously, and learned a lot. I look forward to the next meeting in the series.