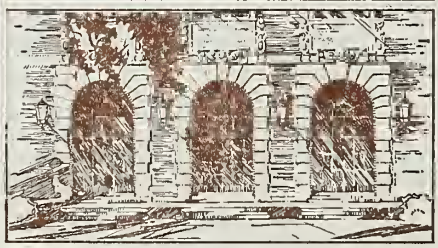


Center for Advanced Computation

UNIVERSITY OF ILLINOIS AT URBANA-CHAMPAIGN
URBANA, ILLINOIS 61801

CAC Document No. 139

ON METHODS FOR DIRECT QUANTIFICATION OF PATTERN ASSOCIATIONS

by

Robert M. Ray III

November 1974

CAC Document No. 139

ON METHODS FOR DIRECT QUANTIFICATION OF PATTERN ASSOCIATIONS

by

Robert M. Ray III

Center for Advanced Computation
University of Illinois at Urbana-Champaign
Urbana, Illinois 61801

November 1974

This work was supported in part by the National Aeronautics and Space Administration and the EROS Program of the Department of Interior through NASA Grant NGR 14-005-202, and in part by the Advanced Research Projects Agency of the Department of Defense through Contract DAHC04-72-C-0001.

ABSTRACT

Rationale is presented for the development of more effective measures of pattern association that may be determined by direct evaluation of pattern similarities. A general notation is suggested for mathematical representation of patterns as multidimensional probability distributions. With respect to this notation, measures of pattern distance, pattern dissimilarity, and pattern correlation are developed that are expressible directly in terms of initial pattern quantizations. The measure of pattern correlation given may be computed invariant with respect to individual pattern sizes, positions, and proximate orientations. The concepts employed would seem well suited for both geometric and network models of pattern information processing.

1. Introduction

The deficiencies of established correlation techniques for effective quantification of pattern similarities have discouraged greatly to date the development of methodologies of pattern recognition based on methods of direct comparison [1,2,3]. Clearly, where patterns of the same class may differ in size, position, orientation, and degree and nature of distortion, conventional template-matching procedures are inappropriate. Thus, with exceptions (see, for example, Widrow [4,5]), an apparent majority of researchers have chosen to pursue analytic methodologies of pattern recognition, i.e., methodologies in which pattern classification depends either upon analysis of transformation and deformation invariant pattern properties, attributes, or features (e.g., statistical methods) or upon analysis of invariant structural relationships between pattern components (e.g., syntactic methods).

There remains, however, in philosophical opposition to all analytic methodologies of pattern recognition the basic hypothesis of gestalt--that there exist, as the most elementary units of perception, holistic organizations of phenomena, unitary perceptual entities, or wholes whose phenomenological characters defy analytic description and are only apprehensible directly. Under this assumption, patterns themselves are necessarily their only valid characterizations. To the extent then that in a particular context meaningful categories of patterns derive directly from basic similarities of gestalt, we must consider all analytic methodologies of pattern recognition inappropriate to the task at hand.

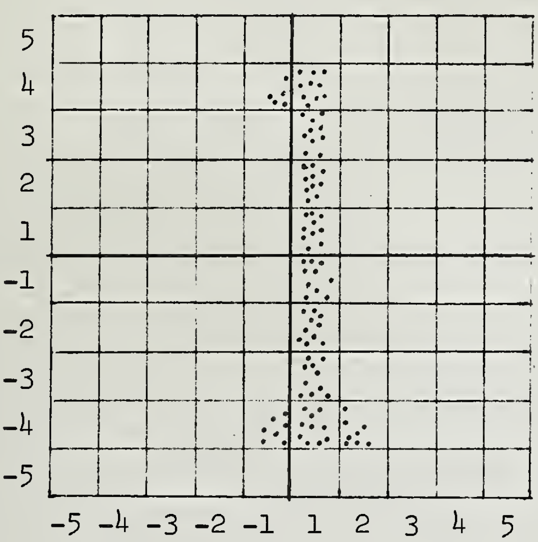
Adopting philosophically the premise that patterns are their own most valid characterizations, while acknowledging the inadequacies of conventional correlation methods of pattern similarity measurement, we consider a fundamental problem of pattern information processing research to be the development of more general and more effective measures of pattern association that may be expressed and computed directly in terms of initial pattern quantizations.

Relying greatly on mathematical concepts long employed by social scientists for modeling economic and social interaction patterns within urban and regional environments, below we suggest a general representation of quantized patterns as probabilistic spatial distributions of information and, with respect to a particular mathematical notation, develop numerical indices of pattern distance, pattern dissimilarity, and pattern correlation that are expressible directly for any two patterns so defined. These measures of pattern association are first developed geometrically for planar pictorial patterns as coefficients of spatial congruence between pairs of two-dimensional probability distributions. The indices presented, however, appear applicable as congruence measures for multidimensional probability distributions in general. In particular, the index of pattern correlation presented is invariant with respect to individual pattern sizes, positions, and proximate orientations and continuous with respect to individual pattern deformations. The concepts employed point toward general network models of pattern information processing that permit conceptualization of both patterns and associations between patterns as probabilistic network distributions of pattern-specific information quanta.

2. Patterns and Pattern Distance

While generalities surrounding the concept pattern make difficult any single definition, to assist the present mathematical discussion we offer the following: a pattern is a unitary organized set of quantized information whose probabilistic spatial (and/or temporal) distribution over some set of sampling elements characterizes some more complex phenomenon source.

If we adopt at least provisionally the above definition, we may represent mathematically any particular pattern f as a partitioned array $(W|X)_f$, as tall as there are sampling elements of f, where W is a matrix of coordinates indicating the relative spatial positions of all elements of f, and X is a vector of positive reals indicating the proportional distribution of quantized units of information across all pattern elements. For lack of any existing term, we will refer generally to these quantum units of pattern information--whose distribution characterizes a particular pattern source--simply as pattern quits (quantization units). Also, for mathematical convenience we will assume normalization of pattern intensities, i.e., equalization of recorded quit totals (or in the case of pictorial patterns normalization of overall levels of brightness or darkness), so that $\sum_i x_i = 1$. Thus the representation of a particular pattern f given by $(W|X)_f$ may be considered a discrete probability distribution X of pattern quits over a spatially arrayed set of sampling elements with centroids W. (See Figure 1.)



f

$$(W|X)_f = \left[\begin{array}{rr|l} -1 & 4 & .05 \\ 1 & 4 & .10 \\ 1 & 3 & .10 \\ 1 & 2 & .10 \\ 1 & 1 & .10 \\ 1 & -1 & .10 \\ 1 & -2 & .10 \\ 1 & -3 & .10 \\ -1 & -4 & .07 \\ 1 & -4 & .10 \\ 2 & -4 & .08 \end{array}\right]$$

Figure 1. Example pattern quantization and mathematical representation.

With such mathematical notation we consider the following pattern association measurement problem: given a set of patterns F, determine a symmetric non-negative scalar index of pattern distance, $D_{f,g} = d[(W|X)_f, (Y|Z)_g]$, pairwise computable for all $f \in F$ and $g \in F$, such that $D_{f,g}$ approaches zero as the spatial congruence of the probability distributions of f and g increases.

For any two patterns f and g quantized in terms of the notation given above we may establish such an index of pattern distance in the following manner. We may determine a weighted correspondence of elements between f and g such that there exists maximal proximity between elements corresponding between f and g. We may then take as our measure of pattern distance the weighted sum of squared distances between all pairs of elements corresponding between f and g where the weights of the sum reflect the degree of correspondence between each element pair.

Let the two patterns f and g consist of m and n elements respectively and let their quantized representations be denoted $(W|X)_f$ and $(Y|Z)_g$. We represent a particular weighted correspondence of elements between f and g as a matrix Q (m x n) satisfying

$$(1) \quad \sum_{i}^{m} q_{i,j} = z_j \qquad j = 1, \ldots, n$$

$$(2) \quad \sum_{j}^{n} q_{i,j} = x_i \qquad i = 1, \ldots, m$$

$$(3) \quad q_{i,j} \geq 0 \qquad i = 1, \ldots, m$$

$$j = 1, \ldots, n.$$

Let $\pi_{f,g}$ denote the set of all Q matrices satisfying (1), (2), and (3) for given X of f and Z of g. Now by normalization of pattern intensities $\sum_i x_i = \sum_j z_j = 1$, hence $\sum_i \sum_j q_{k,j} = 1$. Since also $q_{i,j} \geq 0$ for all i and j, we may consider any $Q \in \pi_{f,g}$ to be a discrete joint probability distribution of "quit correspondences" between the elements of f and the elements of g. Alternatively, any $Q \in \pi_{f,g}$ represents a probabilistic matching or connection of the quits of f with the quits of g.

Now assuming fixed geometries for f and g (for example, a font recognition problem where all quantized patterns may be assumed standardized with respect to positions, sizes, and orientations), let S be the matrix of squared distances between all elements of f and all elements of g given directly by

$$(4) \quad s_{i,j} = \sum_k (w_{i,k} - y_{j,k})^2 \qquad \begin{array}{l} i = 1, \ldots, m \\ j = 1, \ldots, n. \end{array}$$

Our formulation of pattern distance between f and g is then

$$(5) \quad D_{f,g} = \min_{Q \in \pi_{f,g}} \sum_i^m \sum_j^n q_{i,j} \, s_{i,j} = \min_{Q \in \pi_{f,g}} \operatorname{tr}(Q'S)$$

where again $\pi_{f,g}$ is the set of all Q matrices satisfying (1), (2), and (3). Note that such a measure of pattern distance may be interpreted as a minimal mean squared distance of spatial separation between corresponding quits of f and g.

Now the optimization problem given by (1), (2), (3), and (5), where $\sum_i x_i = \sum_j z_j = \gamma$ but γ not necessarily unity, may be recognized as the <u>Hitchcock</u> or <u>transportation problem</u> of linear programming [6,7,8]. Typically, the problem requires determination of a matching between a spatially distributed set of economic supplies and a spatially distributed set of demands such that the total cost of all material movements from suppliers to buyers is minimal. For such problems, computational algorithms are well known and solution properties well documented [9,10]. Thus, a variety of computational procedures exist that can be employed to determine simultaneously an optimal set of weighted correspondences between pattern elements (an extremal joint probability distribution of quit correspondences) and the minimal value of pattern distance yielded by these correspondences.

We may note also at this point that the measure of pattern distance presented should be useful not only for pattern recognition applications per se but also for numerous other applications where there is needed some composite scalar measure of the spatial congruence of pairs of probability distributions. For example, the measure presented would seem well suited as a measure of ecological association between spatially distributed populations of social and biotic communities within ecosystem analyses [11,12].

3. Pattern Dissimilarity

The index of pattern distance presented above provides a measure of the spatial congruence of patterns under the assumption that individual pattern positions, sizes, and orientations may be regarded as standardized or, for other reasons, must be taken as fixed. For most pattern recognition applications, however, no such conditions will prevail. Hence, we remain faced with the problem: given a set of patterns F whose overall characters may be considered independent of individual pattern positions, sizes, and orientations, determine a symmetric non-negative index of pattern dissimilarity, $\Delta_{f,g} = \delta[(W|X)_f, (Y|Z)_g]$, pairwise computable for all $f \in F$ and $g \in F$, such that $\Delta_{f,g}$ approaches zero as the similarity of f and g increases.

As an extension of the method presented above for measurement of in situ pattern congruence, we establish an index of pattern dissimilarity in the following manner. We determine not only a weighted correspondence of elements between f and g but also a spatial registration of f with respect to g such that there results maximal spatial congruence of elements corresponding between f and g. We then take as our criterion of pattern dissimilarity the weighted sum of resulting squared distances between all pairs of elements corresponding between f and g where the weights of the sum again reflect the extent of correspondence determined for each pair of pattern elements.

Let a particular spatial registration of f with respect to g be denoted $\sigma WR + JT'$ where J is the vector $(1, \ldots, 1_m)'$, T is a translation vector, σ is a scale factor, and R is any additional legitimate linear transformation, e.g. a proper rotation. For a given registration of f with respect to g, let S (m x n) be the resulting matrix of squared distances between the elements of f and the elements of g.

$$(6) \quad s_{i,j} = \sum_k [\sigma(\sum_l w_{i,l} r_{l,k} - t_k) - y_{j,k}]^2 \qquad i = 1, \ldots, m \quad j = 1, \ldots, n$$

Let $\Sigma_{f,g}$ denote the set of S matrices obtainable for f and g by (6) over all positive scalars σ, all translations T, and all legitimate transformations R.

Our criterion of pattern dissimilarity may then be formulated:

$$(7) \quad \Delta_{f,g} = \min_{Q \epsilon \pi_{f,g},\ S \epsilon \Sigma_{f,g}} \sum_i^m \sum_j^n q_{i,j} s_{i,j}$$

$$= \min_{Q \epsilon \pi_{f,g},\ S \epsilon \Sigma_{f,g}} \operatorname{tr}(Q'S).$$

Note that such an index of pattern dissimilarity may be interpreted theoretically as a minimal mean squared error of registration between all corresponding quits of f and g.

In the following sections we present a number of alternative mathematical techniques which, in specific combinations, provide computational solutions to (7). The methods developed all yield numerical estimates for $\Delta_{f,g}$ via iterative solution of the two interdependent subproblems implied by (7) -- the <u>correspondences problem</u> requiring minimization of $\Delta_{f,g}$ over all $Q \epsilon \pi_{f,g}$ for fixed S, and the <u>transformation problem</u> which requires determination of that spatial registration of f with respect to g that minimizes $\Delta_{f,g}$ over all $S \epsilon \Sigma_{f,g}$ for fixed Q. Since, as pointed out above, we already have at hand established linear optimization techniques for computational solution of the correspondences problem, let us now turn to analysis of the specific transformations required for optimal spatial registration of patterns within pairwise comparisons.

4. The Transformation Problem and Normalization Procedures

Let f and g be two patterns, again with quantizations $(W|X)_f$ and $(Y|Z)_g$, taken from a set of patterns F not assumed to be standardized with respect to individual pattern sizes, positions, and orientations. On the contrary, in this section let us assume that pattern positions and sizes are arbitrary and that also individual pattern orientations may include considerable rotational displacements from prototypical axial alignments. Our problem is to determine that set of translational, scale, and rotational transformations that will bring about that particular spatial registration of f and g, and hence that particular matrix of squared distances S via (6), such that our previous measure of pattern distance, $D_{f,g}$ via (5), might be determined as a minimum over all $S \in \Sigma_{f,g}$ as well as over all $Q \in \pi_{f,g}$. This is precisely the meaning of our measure of pattern dissimilarity $\Delta_{f,g}$ as given via (7).

Now regarding Q as given, consider all possible translations of f with respect to g and with reference to (5) write

$$(8) \quad D = \sum_i^m \sum_j^n q_{i,j} \sum_k [(w_{i,k} - t_k) - y_{j,k}]^2 .$$

To determine the particular T that minimizes D over all translations of f with respect to g, differentiate (8) with respect to t_k to obtain

$$(9) \quad \frac{dD}{dt_k} = \sum_i^m \sum_j^n q_{i,j} \sum_k (2t_k - 2w_{i,k} + 2y_{j,k})$$

from which it may be determined that $t_k = 0$ where $\sum_i \sum_j q_{i,j} w_{i,k} = \sum_i \sum_j q_{i,j} y_{j,k}$, or where $\sum_i x_i w_{i,k} = \sum_j z_j y_{j,k}$. This condition implies that an optimal registration between f and g requires a coincidence of pattern centroids. Let us therefore normalize the positions of both f and g so that centroids are coincident at a common origin, i.e., so that $\sum_i x_i w_{i,k} = \sum_j z_j y_{j,k} = 0$ for each spatial dimension k. Since these new centroids remain invariant over any additional scale and rotational transformations, we may conclude that no further consideration of pattern translations is necessary in minimizing $\Delta_{f,g}$.

Now consider all possible positive scale factors σ applied to f so that

$$(10) \quad D = \sum_{i}^{m} \sum_{j}^{n} q_{i,j} \sum_{k} (\sigma\, w_{i,k} - y_{j,k})^2 .$$

Differentiating D with respect to σ we find that the particular σ minimizing (10) is given by

$$(11) \quad \sigma = [\sum_{i}^{m} \sum_{j}^{n} q_{i,j} \sum_{k} w_{i,k}\, y_{j,k}] \,/\, [\sum_{i}^{m} \sum_{j}^{n} q_{i,j} \sum_{k} w_{i,k}^2].$$

If σ scales optimally W with respect to Y, then by symmetry, σ^{-1} scales optimally Y with respect to W. By an identical analysis we may determine for σ^{-1} the expression

$$(12) \quad \sigma^{-1} = [\sum_{i}^{m} \sum_{j}^{n} q_{i,j} \sum_{k} y_{j,k}^2] \,/\, [\sum_{i}^{m} \sum_{j}^{n} q_{i,j} \sum_{k} w_{i,k}\, y_{j,k}].$$

Hence $\sigma = \sigma^{-1} = 1$ where $\sum_i \sum_j q_{i,j} \sum_k w_{i,k}^2 = \sum_i \sum_j q_{i,j} \sum_k y_{j,k}^2$, or where $\sum_i x_i \sum_k w_{i,k}^2 = \sum_j z_j \sum_k y_{j,k}^2$. This condition implies that optimal registration of f and g requires equality of pattern second moments about the origin. Let us therefore normalize the sizes of both f and g so that second moments equal unity, i.e., so that $\sum_i x_i \sum_k w_{i,k}^2 = \sum_j z_j \sum_k y_{j,k}^2 = 1$. Since these second moments remain unchanged over any rotational transformations that may be required, in minimizing $\Delta_{f,g}$ we may now also exclude all further consideration of pattern sizes.

The above analysis demonstrates that transformed pattern positions and sizes, optimal with respect to the minimal $\Delta_{f,g}$, may be determined directly by normalization procedures independent of whatever correspondences Q are defined between the elements of f and g and independent of whatever rotation R may be chosen to effect maximum spatial congruence of corresponding elements. The specific normalization procedures given might be profitably included as a pre-processing step with the generalized template-matching technique given above in Section 2. providing normalized measures of pattern distance for all pairwise comparisons. In the present context, it remains to determine that particular R and that particular Q (which as we shall see are interdependent) that together yield a minimum value of pattern dissimilarity $\Delta_{f,g}$.

5. Rotation to Maximum Pattern Correlation

To determine simultaneously the rotation R effecting maximal spatial congruence of normalized patterns f and g and the matrix of correspondences Q that together yield a minimal value of pattern dissimilarity $\Delta_{f,g}$, we adopt a hill-climbing computational procedure. With respect to initial orientations of f and g, we compute a first estimate of S via (4) and then a first estimate of Q via (5). Then, with these initial correspondences fixed, we may determine a first estimate of the optimal rotation R in the following manner.

Note that, for f and g normalized and Q fixed, our stepwise optimal value of pattern dissimilarity may be formulated

$$(13)\quad \Delta = \min_{S \epsilon \Sigma_{f,g}} \operatorname{tr}(Q'S)$$

where our problem is again to determine a rotation $R \in \Theta$ (Θ the set of proper planar rotations) yielding a new estimate of S via (6) stepwise optimal with respect to Q.

Define $\underline{Q}$ and $\underline{S}$ as mn x mn diagonal matrices such that

$$(14)\quad (\underline{q}_{1,1}, \underline{q}_{2,2} \ldots, \underline{q}_{mn,mn}) = (q_{1,1}, q_{1,2} \ldots, q_{m,n})$$

and

$$(15)\quad (\underline{s}_{1,1}, \underline{s}_{2,2} \ldots, \underline{s}_{mn,mn}) = (s_{1,1}, s_{1,2} \ldots, s_{m,n}).$$

Then we may express (13) alternatively as

$$(16)\quad \Delta = \min_{R\epsilon\Theta} \operatorname{tr}(\underline{Q}'\underline{S}) = \min_{R\epsilon\Theta} \operatorname{tr}(\underline{QS})$$

where the elements of $\underline{S}$ remain to be determined as a function of the unknown R.

Also define $\underline{W}$ to be mn x k where the first row of W is repeated n times as the first n rows of $\underline{W}$, the second row of W repeated as the next n rows of $\underline{W}$ and so forth. Define $\underline{Y}$ to be mn x k where the entire matrix Y is simply repeated vertically m times.

Now note that ordered diagonal elements of the matrix $[(\underline{W}R-\underline{Y})(\underline{W}R-\underline{Y})']$ are identically equal to the elements of $\underline{S}$, hence we may write

$$(17) \quad \text{tr}\ [(\underline{W}R-\underline{Y})\ (\underline{W}R-\underline{Y})'] = \text{tr}\ (\underline{S})$$

and since $\underline{Q}$ is also diagonal we may restate (16) as

$$(18) \quad \Delta = \min_{R\epsilon\Theta} \text{tr}\ [\underline{Q}(\underline{W}R-\underline{Y})\ (\underline{W}R-\underline{Y})'].$$

Defining $\underline{Q}^{\frac{1}{2}}$, such that $\underline{Q}^{\frac{1}{2}}\underline{Q}^{\frac{1}{2}} = \underline{Q}$, we may write

$$(19) \quad \Delta = \min_{R\epsilon\Theta} \text{tr}\ [\underline{Q}^{\frac{1}{2}}(\underline{W}R-\underline{Y})\ (\underline{W}R-\underline{Y})'\underline{Q}^{\frac{1}{2}}]$$

and after manipulation,

$$(20) \quad \Delta = \min_{R\epsilon\Theta} \text{tr}\ [(\underline{Q}^{\frac{1}{2}}\underline{W}R-\underline{Q}^{\frac{1}{2}}\underline{Y})'(\underline{Q}^{\frac{1}{2}}\underline{W}R-\underline{Q}^{\frac{1}{2}}\underline{Y})].$$

Now substitute $\underline{\tilde{W}} = \underline{Q}^{\frac{1}{2}}\underline{W}$ and $\underline{\tilde{Y}} = \underline{Q}^{\frac{1}{2}}\underline{Y}$ into (20) to obtain

$$(21) \quad \Delta = \min_{R\epsilon\Theta} \text{tr}\ [(\underline{\tilde{W}}R-\underline{\tilde{Y}})'(\underline{\tilde{W}}R-\underline{\tilde{Y}})]$$

which may be written

$$(22) \quad \Delta = \min_{R\epsilon\Theta} \text{tr}\ (R'\underline{\tilde{W}}'\underline{\tilde{W}}R - 2R'\underline{\tilde{W}}'\underline{\tilde{Y}} + \underline{\tilde{Y}}'\underline{\tilde{Y}})$$

or, since the trace of a sum equals the sum of the traces,

$$(23)\quad \triangle = \min_{R\epsilon\Theta} [tr(R'\underset{\sim}{\tilde{W}}'\underset{\sim}{\tilde{W}}R) - 2tr(R'\underset{\sim}{\tilde{W}}'\underset{\sim}{\tilde{Y}}) + tr(\underset{\sim}{\tilde{Y}}'\underset{\sim}{\tilde{Y}})].$$

Now consider the first and last terms of (23). Clearly both are independent of R. Furthermore, by normalization and our definitions of $\underline{W}$, $\underline{Y}$, $\underline{\tilde{W}}$ and $\underline{\tilde{Y}}$ we may write

$$(24)\quad tr(R'\underline{\tilde{W}}'\underline{\tilde{W}}R) = tr(\underline{\tilde{W}}'\underline{\tilde{W}}) = tr[(\underline{Q}^{\frac{1}{2}}\underline{W})'(\underline{Q}^{\frac{1}{2}}\underline{W})]$$

$$= \sum_i \sum_j q_{i,j} \sum_k w_{i,k}^2 = \sum_i x_i \sum_k w_{i,k}^2 = 1$$

and

$$(25)\quad tr(\underline{\tilde{Y}}'\underline{\tilde{Y}}) = tr[(\underline{Q}^{\frac{1}{2}}\underline{Y})'(\underline{Q}^{\frac{1}{2}}\underline{Y})] = \sum_i \sum_j q_{i,j} \sum_k y_{j,k}^2$$

$$= \sum_j z_j \sum_k y_{j,k}^2 = 1.$$

Since also the middle term of (23) may be written $-2tr[R'(\underline{Q}^{\frac{1}{2}}\underline{W})'(\underline{Q}^{\frac{1}{2}}\underline{Y})] = -2tr(R'\underline{W}'\underline{QY})$, we have shown that our rotation problem may be formulated equivalently as

$$(26)\quad \triangle = 2 - 2 \max_{R\epsilon\Theta} tr(R'\underline{W}'\underline{QY})$$

or,

$$(27)\quad \triangle = 2 - 2 \max_{R\epsilon\Theta} tr(R'W'QY).$$

With reference to (27) we notice that solution of (7) is equivalent to solution of

$$(28) \quad P_{f,g} = \max_{R\in\Theta,\ Q\in\pi_{f,g}} \operatorname{tr}(R'W'QY)$$

where there exist the inverse monotonic mappings $\Delta_{f,g} = 2 - 2P_{f,g}$ and $P_{f,g} = 1 - \frac{1}{2}\Delta_{f,g}$. Since $\Delta_{f,g}$ is formulated as a mean sum of squared distances, its lower bound is zero, hence the upper bound for $P_{f,g}$ is unity. We may thus refer to $P_{f,g}$ as the pattern correlation of f and g and solve (28) as an alternative to (7).

Now if we allow R to be any orthogonal transformation, i.e., either a proper or improper rotation, the optimization problem given by any of the above formulations of our pattern transformation problem may be recognized as the Procrustes problem of psychometrics [13,14]. The problem arises in factor analysis and multidimensional scaling applications where it is desirable to compare two sets of factor coordinates independently determined for the same set of variables by rotation of one set to maximum spatial congruence with the other to maximize between-set factor correlations. Mathematically, the problem is closely related to the canonical correlation problem of multivariate analysis.

It has been shown then that for our present problem where W and Y and hence $\underline{W}$, $\underline{Y}$, $\underline{\tilde{W}}$ and $\underline{\tilde{Y}}$ are of full-column rank k, an optimal orthogonal transformation R is given by

$$(29) \quad R = (HL^{-\frac{1}{2}}H')\,\underline{\tilde{W}}'\underline{\tilde{Y}}$$

where H (k x k) and L (k x k diagonal) represent respectively the eigenvectors and the eigenvalues of the matrix $(\underline{\tilde{W}}'\underline{\tilde{Y}}\underline{\tilde{Y}}'\underline{\tilde{W}})$ [15]. Since both $\underline{\tilde{W}}$ and $\underline{\tilde{Y}}$ are of rank k, all roots of $(\underline{\tilde{W}}'\underline{\tilde{Y}}\underline{\tilde{Y}}'\underline{\tilde{W}})$ are positive and we may take as the elements of $L^{-\frac{1}{2}}$ the reciprocals of the positive square roots of the elements of L [13].

While it is true that computation of R via (29) optimizes Δ over all orthogonal transformations, definition of element correspondences Q via (4) (or via (6) where R occurs as a small proper rotation) makes it extremely improbable that the maximum of tr(R'W'QY) will now occur for an improper rotation--that is, a reflection of f about some axis as well as a proper rotation of f with respect to g. Exceptions to this rule occur when comparing patterns whose coordinate matrices, W or Y, are only weakly of full-column rank, i.e., patterns of nominal full-column rank k whose spatial geometries can be accommodated with only slight distortion in a subspace of dimensionality $k' < k$. For example, where two patterns being compared represent quantized left and right parentheses, "(" and ")", and where Q has been given by (4), we would expect the R given by (29) to contain a horizontal reflection. On the other hand, if the two patterns being compared are "M" and "W", both patterns strongly two-dimensional, we would not expect an R computed via the same method to contain a vertical reflection. In any case, where pattern reflections are significant, the determinant of the matrix R (det R) may be computed to detect improper rotations and further action may be undertaken appropriate to the specific application.

In the last section, we presented a method for determining Q optimal with respect to an assumed S. In this section, we have shown how an optimal transformation R, and hence an optimal S, may be determined with respect to a given Q. Since both subproblems are formulated to optimize the same criterion $\Delta_{f,g}$ (or $P_{f,g}$), iterative solution of both yields a value of $\Delta_{f,g}$ optimal at least locally over all $S \in \Sigma_{f,g}$ and $Q \in \pi_{f,g}$. Thus given a set of quantized patterns F for which rotational displacements can be assumed small, a numerically expressible procedure exists for determining $\Delta_{f,g}$ for all $f \in F$ and all $g \in F$.

6. The Network Entropy Formulation of the Correspondences Problem

In Section 2. above, we noted that the problem of determining an optimal set of weighted correspondences between the elements of two patterns (an extremal probabilistic matching of pattern quits) can be solved via well known linear programming techniques, specifically via Hitchcock or transportation algorithms. While any of these computational procedures may be used within a variety of classification methods based on $D_{f,g}$ and $\Delta_{f,g}$, the computational characteristics of pattern information processing in general compel us to look further for a numerical expression of our correspondences problem of a structure more appropriate to parallel computation.

Here our motivation stems from two sources. For technical and economic reasons, we wish to explore the applicability of the pattern recognition methodology presented for special-purpose hardware implementations. For purely scientific and philosophical reasons, we wish not to overlook any meaningful analogies between the numerical methodology itself and naturally-occurring pattern information processes.

Now a problem closely related to the Hitchcock problem (and well known to urban and regional transportation planners) is called the <u>entropy network distribution problem</u> [16,17]. The problem arises where it is desirable to simulate traffic flows within a metropolitan region given data describing distributions of populations and economic activities over some set of analysis zones subdividing the region, zone-to-zone travel times, and estimates of mean travel times for specific types of trips within the region. Borrowing the notation of our pattern correspondences problem, let Δ represent the mean travel time for all home-work commuting trips, let X be the probability distribution of workers over m residential zones, let Z be the distribution of jobs over n employment zones, and let S be a matrix of network travel times between any residential zone and any employment zone. The problem requires determination of a most probable, mean, or maximum entropy joint probability distribution Q with marginals X and Z such that each element $q_{i,j}$ represents the forecasted proportion of all trips occurring between the i-th residential zone and the j-th employment zone. Mathematically, the problem is formulated

$$(30) \quad \max H = - \sum_{i}^{m} \sum_{j}^{n} q_{i,j} \log q_{i,j}$$

subject to (1), (2), and (3) (as given in Section 2. above) and the additional mean-travel-time constraint

$$(31)\quad \sum_{i}^{m} \sum_{j}^{n} q_{i,j}\, s_{i,j} = \Delta.$$

Note that constraint (31) may be considered simply as an a priori specification of overall network distribution efficiency or total energy expenditure.

The solution to the problem is given by

$$(32)\quad q_{i,j} = x_i\, u_i\, z_j\, v_j \exp(-\beta\, s_{i,j}) \qquad i = 1, \ldots, m \quad j = 1, \ldots, n$$

where β represents the Lagrange multiplier associated with constraint (31) and the u_i and v_j are functions of the Lagrange multipliers associated with constraint sets (1) and (2). It has been shown [18] that corresponding to any real β there exists a unique Q maximizing (30) and satisfying (1), (2), and (3) given by (32) where the parameters u_i and v_j may be determined by iterative solution of the equations

$$(33)\quad u_i = [\sum_{j}^{n} z_j\, v_j \exp(-\beta\, s_{i,j})]^{-1} \qquad i = 1, \ldots, m$$

$$(34)\quad v_j = [\sum_{i}^{m} x_i\, u_i \exp(-\beta\, s_{i,j})]^{-1} \qquad j = 1, \ldots, n.$$

Additionally, it has been shown that there exists a monotonic mapping between all β and all feasible Δ such that as β approaches $-\infty$, Δ approaches Δ_{max}, and as β approaches $+\infty$, Δ approaches Δ_{min}, where Δ_{max} and Δ_{min}, respectively, denote the maximum and minimum values of Δ feasible for given S, X, and Z [19,20]. Together these results yield a theoretical basis for iterative determination of the unique Q maximizing (30) and satisfying a particular feasible efficiency constraint (31) as well as constraints (1), (2), and (3). Since Δ_{min} of the entropy network distribution problem is analogous to the pattern association criteria of our present pattern information processing models, these results also imply that we may at least theoretically formulate a maximum entropy correspondence matrix Q*, unique and optimal with respect to our pattern association criteria via equations (32), (33), and (34) with the parameter β set to $+\infty$.

Using theorems developed elsewhere [21] and well known properties of the Hitchcock model, Evans [20] demonstrates several features of the matrix Q* and suggests a strategy by which it may be computed. Let E denote a binary matrix such that $e_{i,j} = 1$ for all subscript pairs (i,j) where $q^* > 0$ and $e_{i,j} = 0$ elsewhere. (Properties of the Hitchcock model imply that E will be sparse.) The desired Q* and the matrix E then interrelate arithmetically in the form

$$(35)\quad q^*_{i,j} = x_i\, u^*_i\, z_j\, v^*_j\, e_{i,j} \qquad \begin{array}{l} i = 1, \ldots, m \\ j = 1, \ldots, n \end{array}$$

where the vector elements u^*_i and v^*_j satisfy the relations

$$(36)\quad u^*_i = [\sum_j^n z_j\, v^*_j\, e_{i,j}]^{-1} \qquad i = 1, \ldots, m$$

$$(37)\quad v^*_j = [\sum_i^m x_i\, u^*_i\, e_{i,j}]^{-1} \qquad j = 1, \ldots, n$$

Despite these simple properties of extremal solutions to the entropy network distribution problem, Evans' method for exposing E, and hence Q*, requires initial solution of the associated Hitchcock problem presumably by traditional techniques.

7. A Heuristic Procedure for Determination of Planar Pattern Dissimilarities

Since for any pairwise pattern comparison computation of pattern dissimilarity via (7) or pattern correlation via (28) is necessarily an iterative hill-climbing procedure, the particular set of quit correspondences Q determined at any one iteration can only be stepwise optimal with respect to the particular transformation R determined previously at that iteration. Thus given a convenient procedure for determining good unbiased approximations of Q, we might choose to hill-climb using at each step only estimates of extremal quit correspondences.

One possible computational strategy for approximating extremal quit correspondences proceeds as follows.

Establish the matrix $\hat{E}$ such that $\hat{e}_{i,j} = \exp(-\beta\ \hat{s}_{i,j})$ where $\hat{S}$ is S scaled linearly to have elements within a specified interval (say between 0 and 1) and β is chosen as large as computational considerations permit (say $\beta = 150$). Initialize $\hat{V}$ as $(1, \ldots 1_n)'$ and determine estimates of the vectors U* and V* via iterative solution of

$$(38)\quad \hat{u}_i = [\sum_j^n z_j\ \hat{v}_j\ \hat{e}_{i,j}]^{-1} \qquad i = 1, \ldots, m$$

$$(39)\quad \hat{v}_j = [\sum_i^m x_i\ \hat{u}_i\ \hat{e}_{i,j}]^{-1} \qquad j = 1, \ldots, n$$

and then estimate Q* via

$$(40)\quad \hat{q}_{i,j} = x_i\ \hat{u}_i\ z_j\ \hat{v}_j\ \hat{e}_{i,j} \qquad i = 1, \ldots, m \quad j = 1, \ldots, n.$$

Now assuming that β is sufficiently large such that $\hat{Q}$ is close to Q*, then we may expect small elements of $\hat{Q}$ to correspond to zero elements of Q*. Hence we may select for each $\hat{q}_{i,j}$ some threshhold value, say $\bar{q}_{i,j} = x_i * z_j$, and approximate Evans' binary matrix E of Section 6. above by re-defining $\hat{e}_{i,j} = 0$ wherever $\hat{q}_{i,j} < \bar{q}_{i,j}$ and re-defining $\hat{e}_{i,j} = 1$ wherever $\hat{q}_{i,j} \geq \bar{q}_{i,j}$. Then, with this new definition of $\hat{E}$ (hopefully Evans' E above), return to iteration of equations (38) and (39) obtaining new estimates of U* and V* and compute a final approximation of Q* via (40).

Now the Procrustes formulation of our pattern transformation problem provided a general solution applicable for comparison of patterns of any dimensionality. In the case of planar pictorial patterns, however, the problem may be resolved in a more direct fashion.

Restricting R to be a proper rotation, let $C = (W'QY)$ for a given Q and write max $\operatorname{tr}(R'W'QY)$ in (28) as

$$(41)\quad f(\alpha) = \max_{\alpha} \operatorname{tr} \begin{bmatrix} \cos\alpha & -\sin\alpha \\ \sin\alpha & \cos\alpha \end{bmatrix} \begin{bmatrix} c_{1,1} & c_{1,2} \\ c_{2,1} & c_{2,2} \end{bmatrix}$$

or equivalently,

$$(42)\quad f(\alpha) = \max_{\alpha} \; (c_{1,2} - c_{2,1}) \sin\alpha + (c_{1,1} + c_{2,2}) \cos\alpha \, .$$

Then let $A = (c_{1,2} - c_{2,1})$ and $B = (c_{1,1} + c_{2,2})$ and write (42) as

$$(43)\quad f(\alpha) = \max_{\alpha} \; (A \sin\alpha + B\cos\alpha).$$

Also, let $K = (A^2 + B^2)^{\frac{1}{2}}$ so that $A = K \sin\phi$ and $B = K\cos\phi$ and

$$(44)\quad f(\alpha) = \max_{\alpha} \; (K \sin\phi \sin\alpha + K\cos\phi \cos\alpha) = \max_{\alpha} \cos(\alpha - \phi) \, .$$

The maximum occurs (at $\alpha = \phi$) as $K = [(c_{1,2} - c_{2,1})^2 + (c_{1,1} + c_{2,2})^2]^{\frac{1}{2}}$. The proper rotation maximizing (28) is then determined by the relations $\sin\alpha = A/K = (c_{1,2} - c_{2,1})/K$ and $\cos\alpha = B/K = (c_{1,1} + c_{2,2})/K$. Therefore, in comparing any two planar patterns f and g, a proper rotation R, stepwise optimal with respect to a given correspondence matrix Q, can be determined directly as a function of the four elements of the matrix $C = (W'QY)$.

The two procedures above for convenient approximation of quit correspondences and direct determination of stepwise optimal rotations, in combination, yield a simple heuristic approach to measurement of pattern dissimilarities for planar patterns. Such a procedure may be programmed as follows:

1. Input, normalize (see Section 2.), and store patterns f and g in terms of quantizations $(W|X)_f$ and $(Y|Z)_g$.

2. Set $\sigma = 1$, $T = (0,0)'$, $R = I$, and $\Delta^o_{f,g} = M$ (M some large value).

3. Compute S via (6).

4. Approximate Q using the heuristic procedure given above in this section, and obtain a new estimate of $\Delta_{f,g}$ via (5).

5. If $|\Delta^o_{f,g} - \Delta_{f,g}| < 0.01$, stop. Otherwise, let $\Delta^o_{f,g} = \Delta_{f,g}$.

6. Compute a new R via the short method given in this section and return to Step 3.

8. Computational Results

To evaluate the effectiveness of such a pairwise pattern comparison procedure, the following experiment was conducted. A test set of ten prototype patterns corresponding to the numerals 1 through 9 and 0 was designed. For convenience, the elements of each prototype were chosen spatially coincident with the cells of a 4 x 8 integer grid and all elements of all prototypes were assigned equal quit densities. Then, sixteen noisy copies of each prototype were generated using the equations

$$(45)\quad W = \sigma_e\,(Y\,R_e) + JT_e'$$

$$(46)\quad X = Z + \chi_e$$

where J denotes the vector $(1, \ldots, 1_n)'$, σ_e, T_e and R_e are, respectively, randomly selected scale, translational, and rotational transformations of prototype spatial coordinates Y, and χ_e represents a random perturbation of prototype quit distributions Z. The ten prototype patterns selected and the sixteen noisy versions generated for each are reproduced in Figure 2. where the sizes of individual pattern elements have been plotted proportional to quit densities.

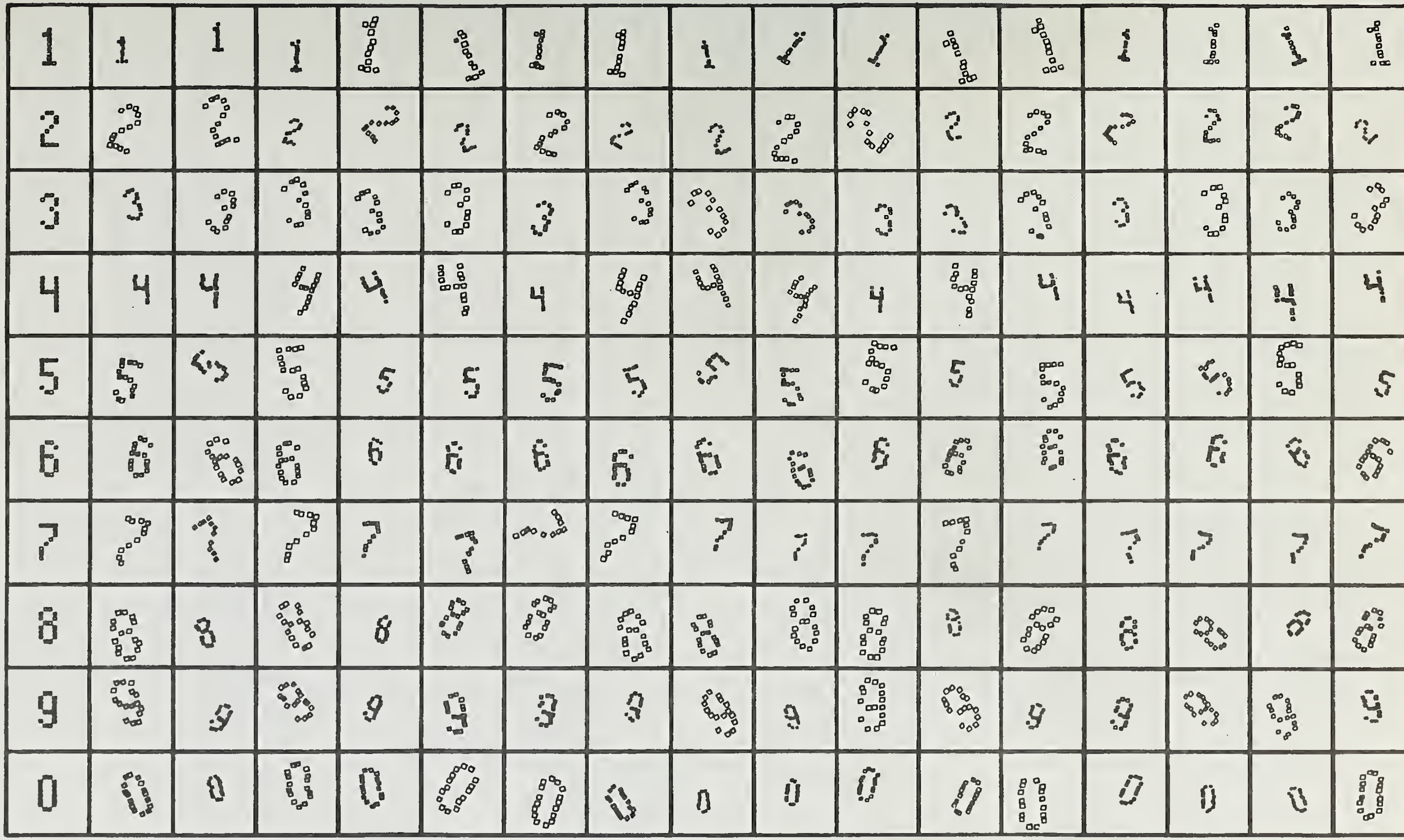


Figure 2. Sixteen noisy copies of each of ten prototype patterns (extreme left). Despite considerable rotational displacement, all 160 noisy patterns were correctly classified using a Fortran implementation of the above techniques.

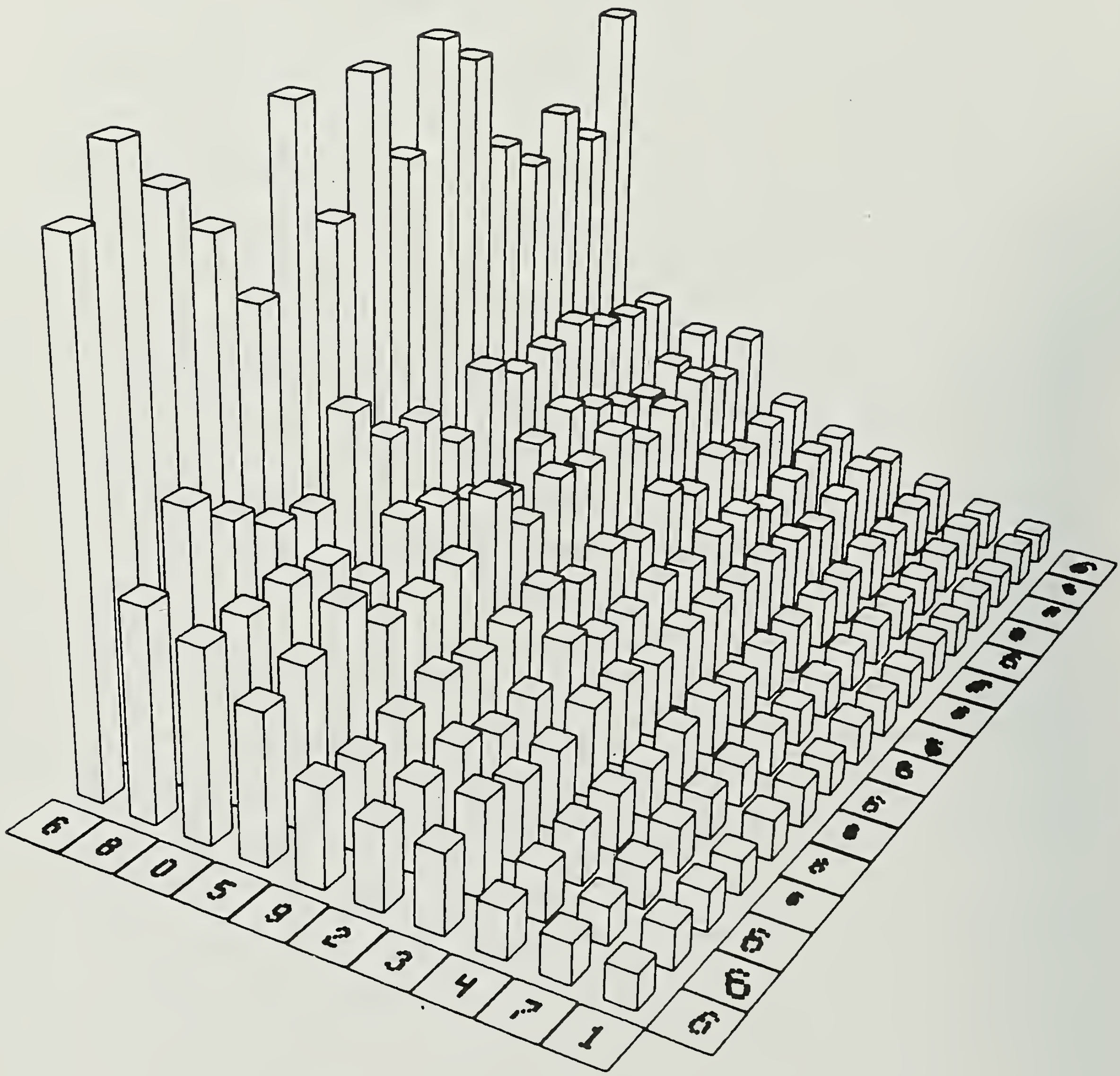


Figure 3. A computer graphic showing the rank order of prototype-to-pattern dissimilarity measures computed for the sixteen noisy "6's" of Figure 2. Individual blocks have been plotted proportional to $1/\Delta_{f,g}$. Also, prototypes have been ordered from left to right in accordance with mean prototype similarity with all noisy patterns depicted.

A Fortran implementation of the above outlined algorithm was executed on the IBM 360/75 of the University of Illinois to compute the 1600 pattern-to-prototype dissimilarity measures. In every case the minimum pattern-to-prototype dissimilarity measure occurred when a pattern was matched with the correct prototype. The total IBM 360/75 CPU time required for computation of all 1600 dissimilarity measures was 1040 seconds, or approximately .65 seconds per comparison. These computation times may be reducible by more efficient programming.

As typical of the results obtained, a graphical presentation of all comparisons for the noisy "6's" is given in Figure 3. There, individual block heights have been plotted proportional to $1/\Delta_{f,g}$ and scaled vertically with respect to the maximum value of $1/\Delta_{f,g}$ occurring within the 160 comparisons. Thus, while exaggerating proportional differences, the display makes plainly visible the rank order of all similarities computed by the pairwise comparison procedure.

9. Conclusions

Adopting the premise that patterns are their own most valid characterizations and relying greatly on mathematical concepts long employed within urban systems modeling, we have posited a new methodology for direct quantification of pattern associations that should serve well as an alternative to conventional template-matching methods.

The mathematical bases of the methods proposed are quite general. Wherever it is reasonable to represent patterns as spatial probability distributions of information, the numerical procedures presented can be employed to obtain specific measures of association between patterns. The methodology is general with respect to the spatial dimensionality of patterns processed. Unlike traditional correlation methods, moreover, it does not depend on any fixed format or order for pattern information sampling and quantization and, in fact, seems relatively insensitive to such considerations.

It is often argued that we stand to gain from research of abstract models of pattern information processing, not only more general methodological bases for technological advancement, but also additional insights into the possible nature of our own mechanisms of perception and information processing. Thus, the feature extraction procedures of our analytic models of pattern recognition have their counterparts within scientific theories of animal vision. In this context, we are hopeful that the abstract models of pattern information processing posited above may lend additional support to existing mathematical and logical bases for holistic mechanisms within perception such as those cooperative processes within human vision hypothesized and extensively investigated by Julesz [22]. To the extent that such reinforcement may be derivable from the above abstractions, we consider it not without significance that our models of pattern communications have strong relationship to, and indeed in this case stem directly from, our models of community.

Acknowledgements

The author wishes to express his appreciation for the interdisciplinary support provided by numerous colleagues at the Center for Advanced Computation in the preparation of this paper. Special acknowledgements are due Professor Daniel L. Slotnick for many discussions concerning the paper's contents. Professor Ahmed H. Sameh lent much assistance with mathematical formulations and notation. The special solution to the planar orthogonal rotation problem was pointed out to the author by Dr. George B. Purdy. A note of appreciation is also given Dr. Hans Mark of NASA's Ames Research Center for numerous comments on an earlier draft.

References

1. Selfridge, O. G., and Neisser, U., "Pattern Recognition by Machine," Scientific American, Vol. 203, No. 18, August 1960, pp. 60-79; reprinted in Feigenbaum, E., and Feldman, J. (eds.), Computers and Thought, McGraw-Hill Book Co., New York, 1963.

2. Minsky, M., "Steps Toward Artificial Intelligence," Proceedings of the Institute of Radio Engineers, Vol. 49, January 1961, pp. 8-30; reprinted in Feigenbaum, E., and Feldman, J. (eds.), Computers and Thought, McGraw-Hill Book Co., New York, 1963.

3. Nagy, G., "State of the Art in Pattern Recognition," Proceedings of the IEEE, Vol. 56, No. 5, May 1968, pp. 836-862.

4. Widrow, B., "The Rubber-Mask Technique -- Pattern Measurement and Analysis," Journal of the Pattern Recognition Society, Vol. 5, No. 3, September 1973, pp. 175-197.

5. Widrow, B., "The Rubber-Mask Technique -- Pattern Storage and Recognition," Journal of the Pattern Recognition Society, Vol. 5, No. 3, September 1973, pp. 199-211.

6. Hitchcock, F. L., "The Distribution of a Product from Several Sources to Numerous Localities," Journal of Mathematics and Physics, Vol. 20, 1941, pp. 224-230.

7. Dorfman, R., Samuelson, P. A., and Solow, R. M., Linear Programming and Economic Analysis, McGraw-Hill Book Co., New York, 1958.

8. Dantzig, G. B., Linear Programming and Extensions, Princeton University Press, Princeton, 1963.

9. Vajda, S., Mathematical Programming, Addison-Wesley, 1961.

10. Hadley, G., Linear Programming, Addison-Wesley, 1962.

11. Duncan, O. D., Cuzzort, R. P., and Duncan, B., Statistical Geography: Problems in Analyzing Areal Data, The Free Press, Glencoe, Illinois, 1961.

12. Pielou, E. C., An Introduction to Mathematical Ecology, John Wiley and Sons, New York, 1969.

13. Schönemann, P. H., "A Generalized Solution of the Orthogonal Procrustes Problem," Psychometrika, Vol. 31, No. 1, March 1966, pp. 1-10.

14. Cliff, N., "Orthogonal Rotation to Congruence," Psychometrika, Vol. 31, No. 1, March 1966, pp. 33-42.

15. Green, B. F., "The Orthogonal Approximation of an Oblique Structure in Factor Analysis," Psychometrika, Vol. 17, No. 4, December 1952, pp. 429-440.

16. Wilson, A. G., Entropy in Urban and Regional Modelling, Pion Limited, London, 1970.

17. Potts, R. B., and Oliver, R. M., Flows in Transportation Networks, Academic Press, New York, 1972.

18. Evans, A. W., "Some Properties of Trip Distribution Methods," Transportation Research, Vol. 4, 1970, pp. 19-36.

19. Evans, A. W., "The Calibration of Trip Distribution Models with Exponential or Similar Cost Functions," Transportation Research, Vol. 5, 1971, pp. 15-38.

20. Evans, S. P., "A Relationship between the Gravity Model for Trip Distribution and the Transportation Problem in Linear Programming," Transportation Research, Vol. 7, 1973, pp. 39-61.

21. Bacharach, M., Biproportional Matrices and Input-Output Change, Cambridge University Press, Cambridge, 1970.

22. Julesz, B., "Cooperative Phenomena in Binocular Depth Perception," American Scientist, Vol. 62, No. 1, January-February 1974, pp. 32-43.