Normalization Techniques for Handprinted Numerals

G. NAGY
IBM Thomas J. Watson Research Center
Yorktown Heights, New York

AND

N. TUONG
Université de Montreal, Montreal, Quebec, Canada*

A family of pattern standardization techniques based on geometrical projection is applied to a file of digitized handprinted numerals obtained from sales clerks. The principle involves transforming a quadrilateral specified in terms of the convex hull of each pattern into a square. The amount of overlap within each class of characters versus the amount between classes is used to evaluate the degree of normalization achieved with respect to other published methods including size and shear normalization through moments.

KEY WORDS AND PHRASES: pattern recognition, character recognition, normalization, projective transformation, central projection, hand printed characters, handwriting, linear transformation, size normalization, mapping, pattern preprocessor
CR CATEGORIES: 3.63, 5.14

Introduction

The subject of this study is a family of pattern normalization techniques based on geometric projection. These techniques are applied to a file of handprinted numerals, some samples of which, both before and after normalization, are shown in Figure 1.

In principle, geometric projection is best suited to the recognition or classification of objects in the three-dimensional world, where it is clear that projective transformations are introduced by changes in the point of view. In terms of plane figures this kind of distortion is equivalent to oblique observation, from a point outside the plane of the figure, of a perfectly formed prototype.

In applying the method to handprinted characters, we must show by experiment that a significant degree of standardization can be attained. The success of the endeavor depends on the appropriateness of our hypothesis regarding the way people print.

* Departement d'Informatique. This paper is abstracted from a technical report of the Departement d'Informatique of the Université de Montreal [6]. The work was done under the financial aegis of the Canadian National Research Council.

Standard preprocessing methods in character recognition include simple scaling, height and width normalization, and size and shear correction through linear transformation [1, 2]. It is of interest to compare the results achieved by geometric normalization with those obtained by the better-known methods.

The experimental work is based on a file of 13,000 numerals obtained from department store clerks in the course of routine sales operations. The sales slips were scanned at the experimental character recognition facility of the IBM Thomas J. Watson Research Center, and the digitized characters were recorded on magnetic tape. Coarse height and line width normalization was included in the scanning routine. A detailed description of the data and further information about the scanning hardware can be obtained from [3].

Method

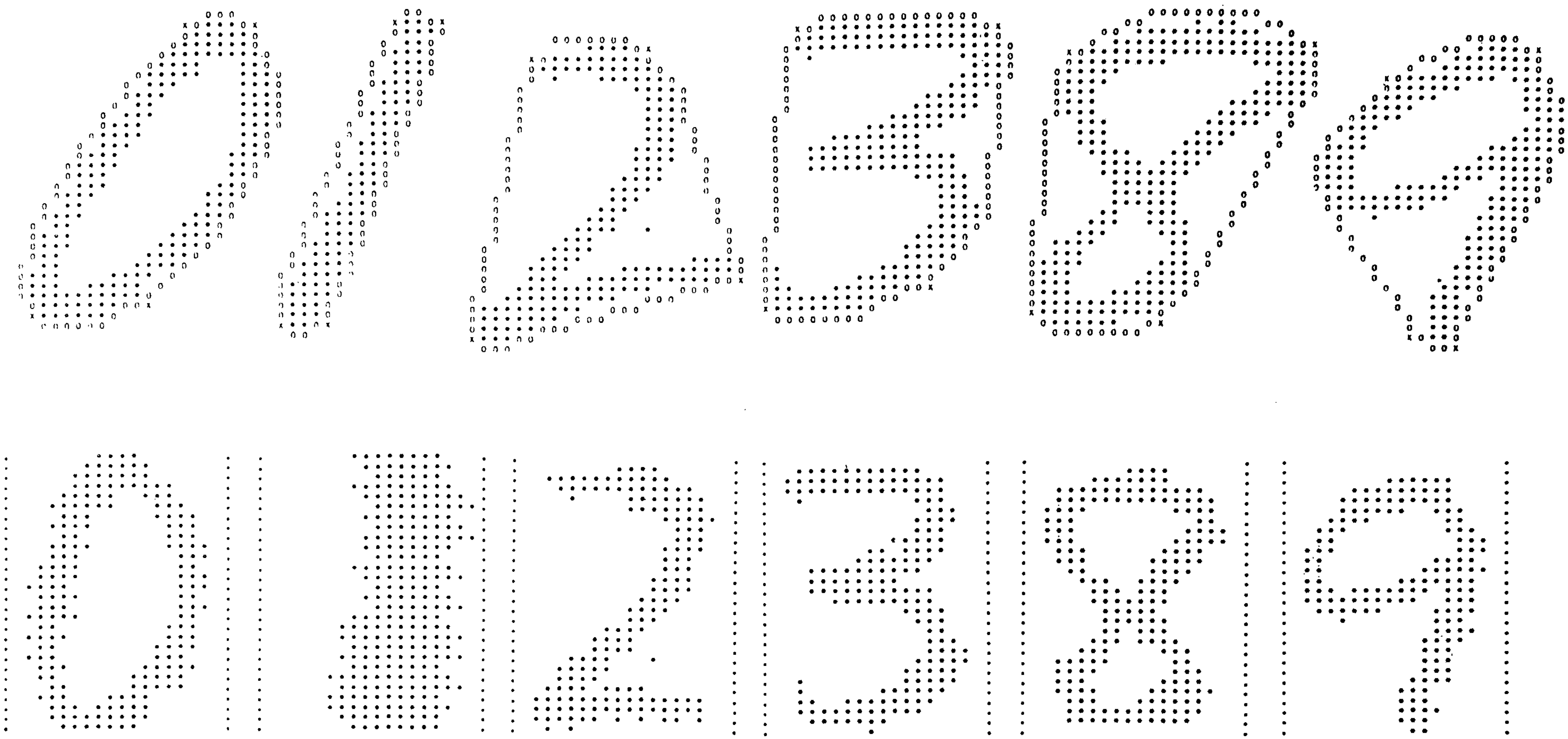
Our goal is to reduce each incoming pattern to a *canonical form* which would, we hope, enhance the similarity between patterns in the same class without obviating the distinguishing characteristics of the different classes. For example, under projective transformation every quadrilateral can be reduced to a square of arbitrary dimension, and every ellipse to a circle (see Figure 2), thus permitting the differentiation of quadrilaterals from ellipses by simple template matching.

Since we regard our canonical transformation merely as the first step in the recognition process, we do not wish to be encumbered by the details of the pattern. The transformation is based solely on the vertices of the *convex hull* [4] of the pattern, which are easily determined for the digitized numerals, as follows.

We find all the black points which are x or y extrema (the "envelope") and select from these a black point which is necessarily a vertex of the convex hull (such as, for example, the rightmost point of the bottom row). Successive vertices are then determined by selecting at each step the point in the envelope with the smallest positive angular displacement with respect to the last point selected. Only ratios of y-increments to x-increments are computed, and the pattern is rounded only once.

Four points are required to specify uniquely a projective transformation [5]. These are derived from the slope of the convex hull; the 45° points are simply spread to allow complete framing of the pattern, as shown on Figure 3. These four points (*fiducial marks*) become the corners of the standard sized rectangular array (24×18 in most of this work) which contains the transformed pattern.

We have arbitrarily chosen, after much discussion and some experimentation [6], to use the vertices of the quadrilateral circumscribing the pattern with sides parallel to the sides of the quadrilateral formed by the four vertices



Fɪɢ. 1. Raw and normalized numerals. The top row represents typical numerals written by busy sales clerks in a department store. The bottom row shows the same characters after a projective transformation.

where the slope of the sides of the convex hull passes through the intercardinal directions (45°, 135°, 225°, 315°).

This method of projective normalization is size and registration invariant within the limits imposed by mesh noise. It is also rotation invariant to the extent that the 45° points are far apart, i.e. the convex hull is not "pointed."

We must be careful, however, not to apply the transformation blindly; a quadrilateral with a very large ratio in the lengths of opposite sides might be better represented by a triangle than a square, while if we allow full rotational invariance the distinction between "6" and "9" is irrevocably lost.

The program includes safeguards against mutilation of numerals such as "7" and certain types of "4" which do not conveniently fit into a quadrilateral frame. If one side of the quadrilateral is abnormally small (in the extreme case it may be a single point) and the opposite side is large, then the form is considered "triangular," and the transformation is performed on one of the four possible parallelograms obtained by completing the triangle. No provision has been included against the fattening of the numeral "1" since this presents only an esthetic drawback.

Bilinear Approximation

For implementation on a digital computer it is convenient to approximate the geometrical projection by means of the transformation shown in Figure 4. While analytically this transformation is less tractable than the projection, it can be computed more rapidly on a set of grid points.

The maximum deviation between the two transformations is calculated in [6] for trapezoid envelopes. An example of the difference on a sample of handprinted numerals is shown in Figure 5. Although the deviation may occasionally exceed the quantization error resulting from the imposition of a coarse grid on the transformed characters, it has the advantage of reducing the unsightly fattening of the lines close to the focus of the geometric projection.

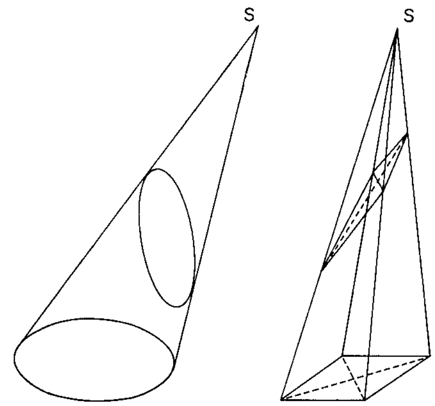


FIG. 2. Geometric projection. Any quadrilateral can be projected into a square by suitable choice of the focus and plane of projection. Arbitrary ellipses are transformed into unit circles, as with linear transformations.

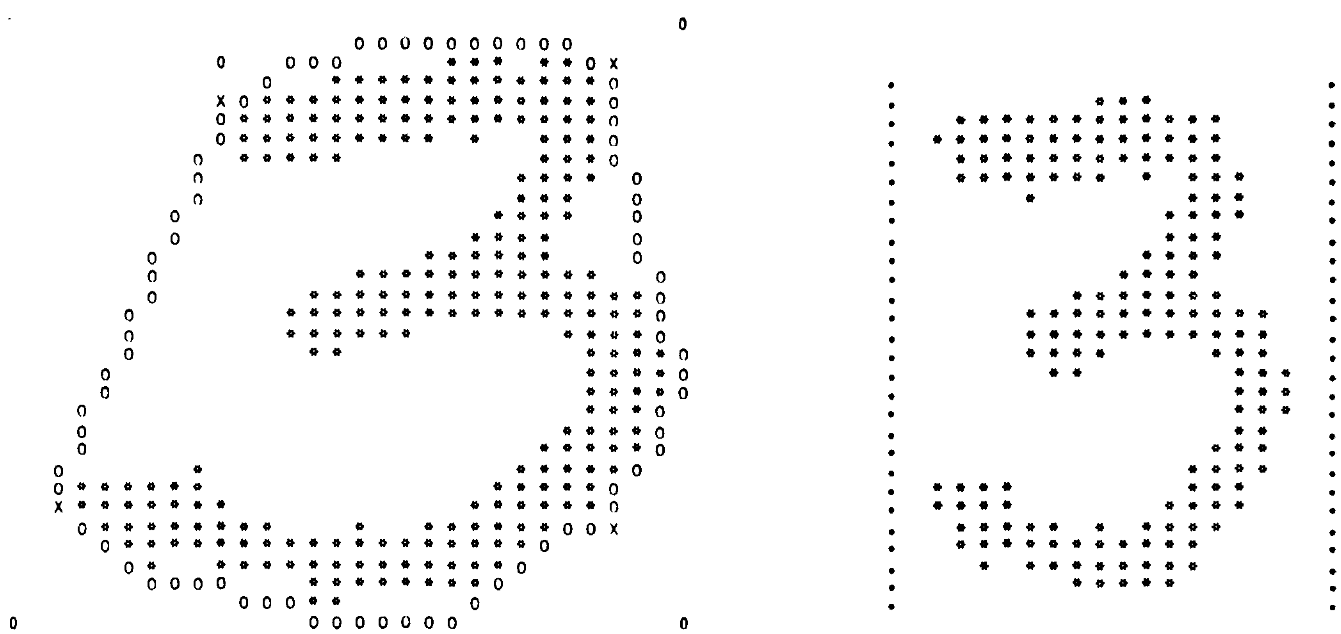


FIG. 3. A heuristic method of defining a transformation. The convex hull of the pattern is shown as a chain of O's. X marks the spots where the slope of the convex hull goes through 45°. The quadrilateral formed by the X's is expanded until it includes the whole character; its vertices are now shown by O's. This quadrilateral is now mapped into the standard 24 × 18 rectangle on the right-hand side.

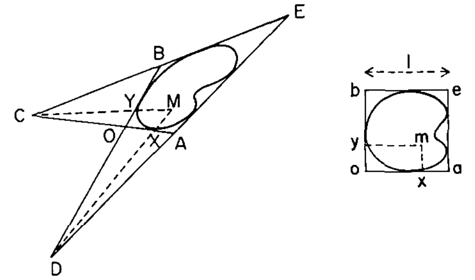
Stability

Will successive applications of the transformation lead to eventual deterioration of the pattern to the point of irrecognizability, or will consecutive iterations converge to a stable configuration which preserves the essential characteristics of the original pattern? Although we have not been able to prove any property of consistency for our method for defining the fiducial marks, experimental evidence shows that very little change occurs in the pattern after the initial transformation. Figure 6 shows the sequence of patterns generated from the numeral 8; the input for each pass is the pattern generated in the last pass. Some numerical results reported below support our contention that a single pass is sufficient for recognition applications.

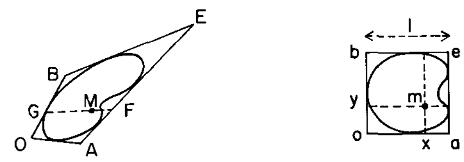
Quantitative Justification

It is clear that one cannot hope for an objective evaluation of a normalization or preprocessing method without at least some implicit assumptions regarding the nature of the recognition network used for the final classification. To benefit from the properties of projective normalization, we must, for example, assume that subsequent stages of the classifier extract geometrical (as opposed to topographic or topological) properties of the pattern. We have chosen, in lieu of simulating one of the well-known classification techniques, to base our assessment on the correlations between the patterns regarded as vectors in binary space.

It will be generally agreed that a pattern recognition problem may be considered easy if the patterns of the



$$\begin{cases} ox = \dfrac{OX.AC}{OA(OC+OX)} \\ oy = \dfrac{OY.BD}{OB(OD+OY)} \end{cases}$$



$$\begin{cases} ox = \dfrac{GM}{GF} \\ oy = \dfrac{OG}{OB} = \dfrac{AF}{AE} \end{cases}$$

Fig. 4. An approximation to geometric projection. The true geometric projection is shown on top, while the transformation used to approximate it is shown below. m is the point with coordinates (x, y) into which a point M of the original pattern is transformed. X and Y are the projections of M on the sides of the original quadrilateral towards the vertices of the completed quadrilateral, C and D. G and F, in the bottom diagram, are on a straight line with M and partition the sides OB and AE of the original quadrilateral into equal ratios. The equations defining the transformation are stated in terms of the distances between specified points.

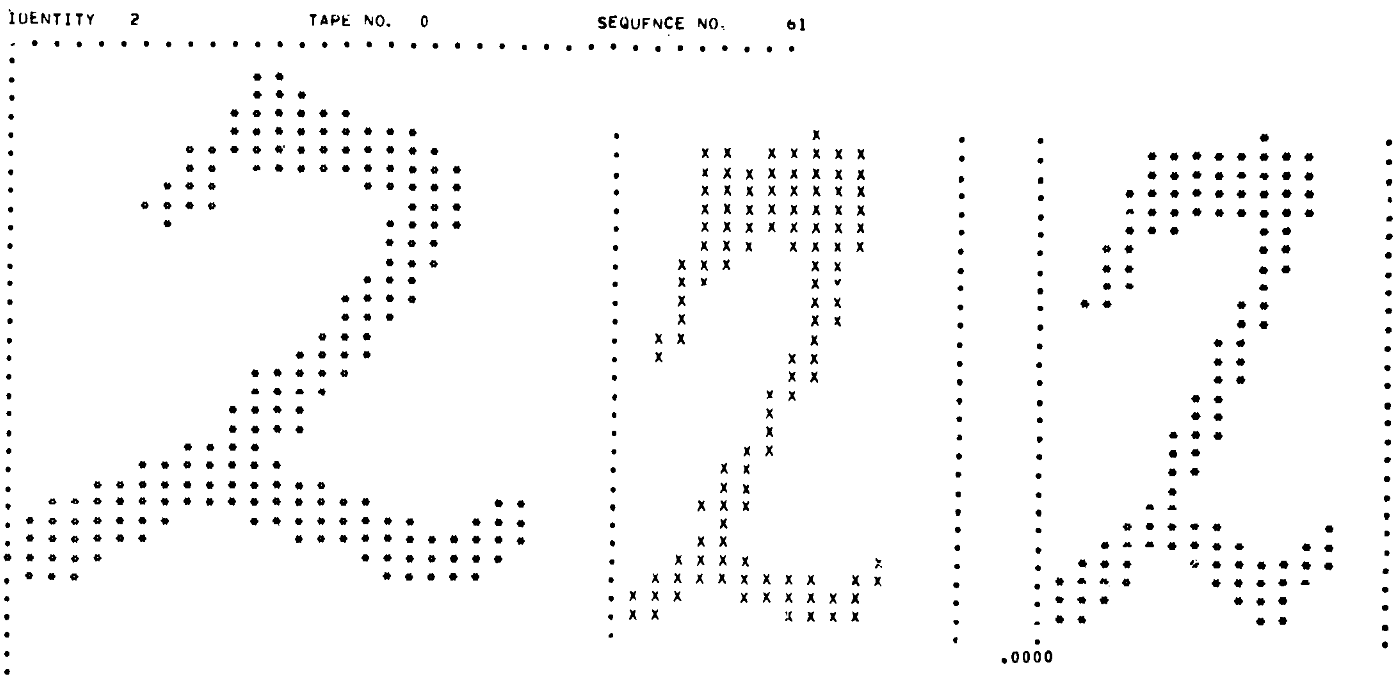


Fig. 5. Comparison of true and approximate methods of projection. The original character is on the left, the true geometric projection in the center, and the approximation on the right.

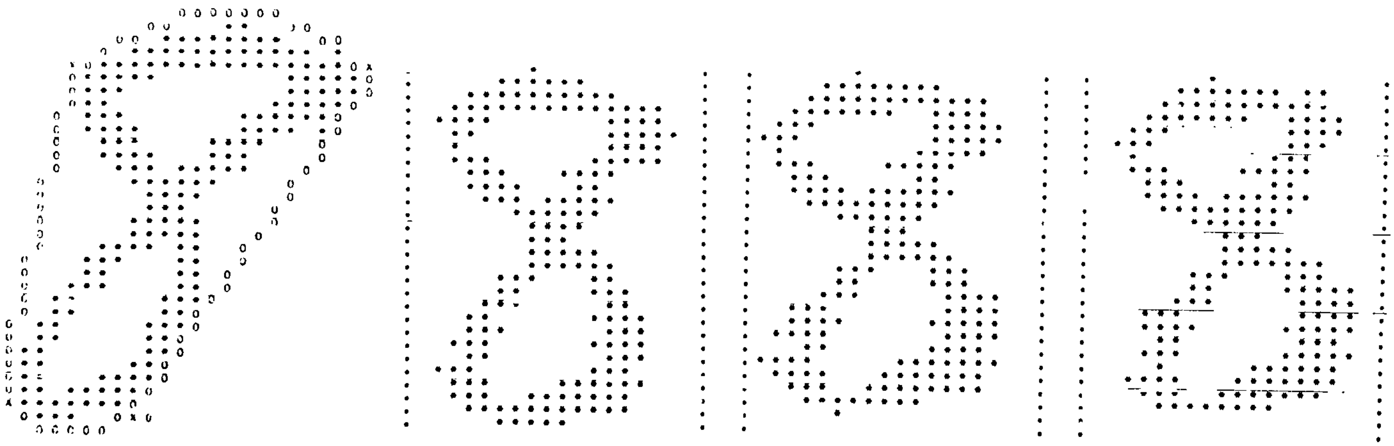


FIG. 6. Iterative application of the transformation. The greatest change is brought about by the first transformation, which is deemed to be sufficient for most practical requirements. Successive projections leave the pattern almost unaltered, as indeed they should.

same class show a high degree of correlation among themselves, and patterns of different class show a low degree of correlation. Conversely, a problem is hard if patterns in different classes tend to have significant overlap while patterns in the same class show little resemblance. By means of this criterion of difficulty we hope to avoid any more dependence than necessary on a specific recognition algorithm.

For a measure of the correlation between two patterns P_i and P_j we use

$$d_{ij} = P_i \cdot P_j / (|P_i| \cdot |P_j|)^{\frac{1}{2}}$$

when $|P_i|$ is the number of "black" bits in pattern P_i.

Figure 7 shows the fraction of pattern pairs having a given d_{ij}, for patterns in the same class and for patterns in different classes. Each pair of curves refers to a different method of normalization. The first pair (7a) is height *or* width normalization—each character is expanded or contracted until its largest dimension just fits into the 24×18 box. The second pair (7b) is independent height *and* width normalization. The third pair (7c) is the linear transformation described in [2]; it is designed to remove shear distortion by diagonalizing the (2×2) moment matrix of each pattern [2]. The fourth pair of curves (7d) is our approximate geometric projection with the fiducial marks derived from the slope of the convex hull. The fifth pair (7e) also represents geometric projection, but here the fiducial marks are the vertices of the quadrilateral of the largest possible area which may be *inscribed* in the convex hull.

Each of these sets of curves is based on 500,000 character pairs. Even with a machine language subroutine on a fairly fast computer (CDC 3400) for computing the correlations, the evaluation takes far longer than does the actual processing of the characters.

The shaded areas under the intersections of the curves correspond to the expected probability of confusion between two classes if the unknown samples are classified by comparing them to a randomly selected representative of each class. With this simple rule we would see about 28 percent of the characters misclassified with size normalization only, and about 13 percent with the projective method based on the slope of the contour.

For a more detailed analysis of the strengths and shortcomings of our normalization scheme, we would like to look at its performance on specific class pairs. Rather than plot all 45 pairs of curves, we have computed a standard statistical figure of merit, the ratio of the difference of the means to the sum of the variances, for each pair of classes (see Table I). These figures are tabulated only for the projective normalization; while the values for the other methods are appreciably lower, their rank ordering is similar.

Figure 8 shows the distribution of the correlation between an original character and its transformed counterpart, and between the transformed character and the result of yet another transformation.

Conclusions

A fairly general class of transformations, that of geometric projections, was applied to the normalization of handprinted numerals. In evaluating the performance of this method in reducing the internal variability of each class, an overall figure of merit of 2.82 was obtained versus 2.50 for its nearest rival, linear normalization through moments.

We know that a recognition rate of 95 percent was reached on the same data with a more sophisticated recognition network (weighted N-tuples) but with height normalization only [3]. We have, however, no way of estimating the potential contribution of geometric normalization to the performance of this network.

It is not surprising to see from the pairwise figure of merit table that the most closely resembling classes, after normalization, are "4" and "9." This is precisely the pair which gives us most trouble in interpreting figures.

Although the heuristic method we used to find the fiducial marks which determine the transformation is not readily amenable to analysis, it is clear from Figure 7 that

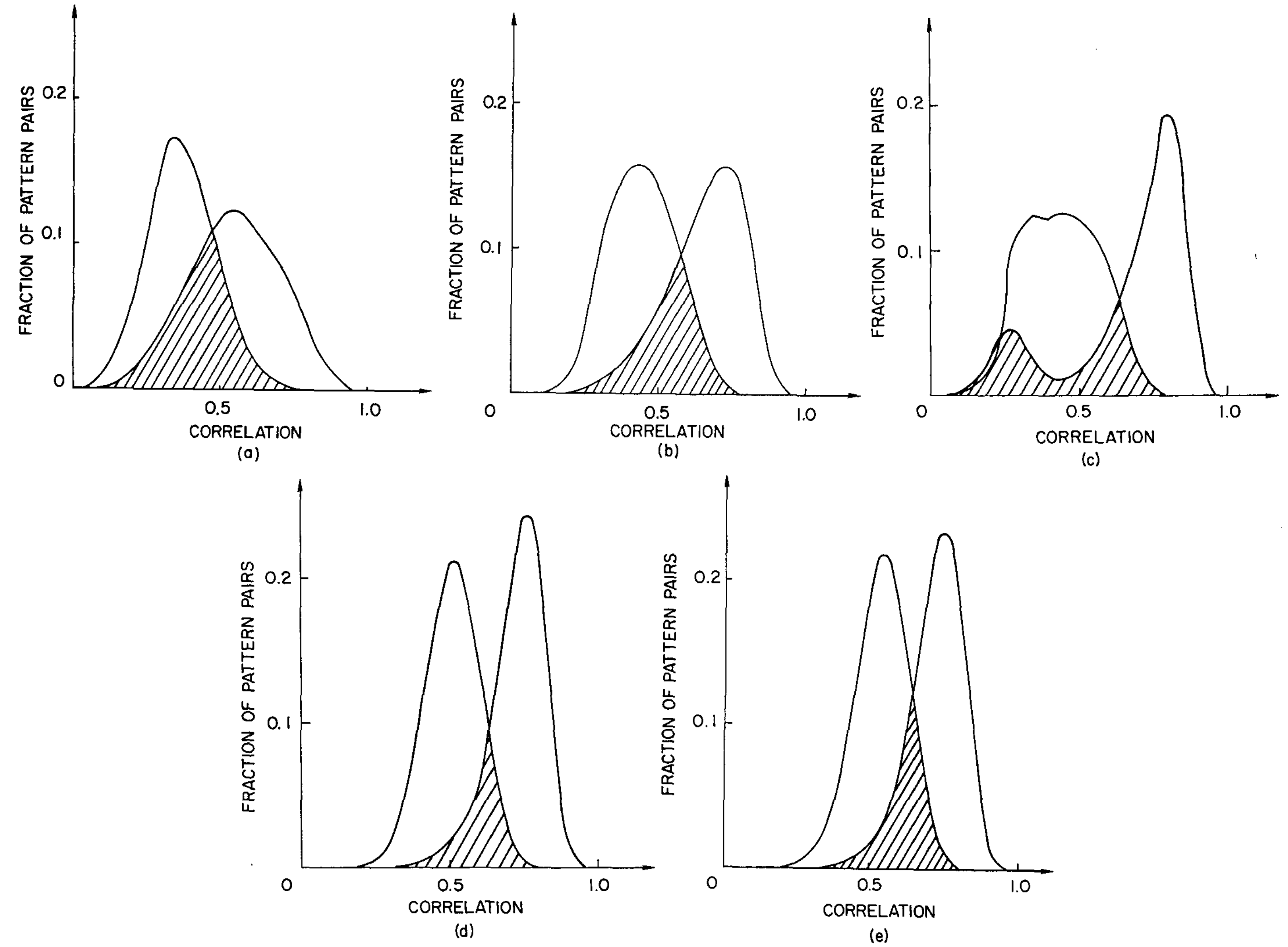


Fig. 7. Quantitative evaluation of different methods of normalization. The distribution of the correlation coefficient defined in the text is shown for pairs of characters in the same class and in different classes. (a) Size normalization. (b) Similarity transformation. (c) Linear transformation including shear removal. (d) Geometric projection based on slopes. (e) Geometric projection based on inscribed quadrilateral.

TABLE I. Value of the Figure of Merit for Each Pair of Classes After Geometric Projection

1	2	3	4	5	6	7	8	9	J/I
6.5835	8.2016	5.0567	6.0037	3.1163	2.0925	6.5160	4.1649	4.2236	0
	2.7659	3.3388	3.3945	5.5722	7.1884	2.8830	2.8962	2.1596	1
		1.0597	3.3512	2.2770	4.4011	1.0763	0.6490	2.5325	2
			2.0897	0.4397	4.8652	1.2015	0.9106	1.6230	3
				1.2421	2.7119	0.9652	0.9468	0.2772	4
					1.0432	1.9303	0.4228	1.3433	5
						6.2880	2.4601	4.9108	6
							1.5689	0.9028	7
								0.7520	8

the rule is stable in the sense that successive applications of the transformation introduce no degradation in the characters. Indeed, the variation between the successively transformed characters, after the initial transformation, is much less than the average variation between the members of the same class.

The transformation of an arbitrary outline inscribed in a quadrilateral into a figure in a square is a rather pretty ruler-and-compass construction, as described in [6]. In programming this transformation for the handprinted characters, quantization noise introduced through the coarse quantization of both the original and the target characters proved to be the greatest source of difficulty. In a production operation, the transformation should probably be approximated even further through table look-up operations which would be less vulnerable to the effects of quantization than the computation of intersections of straight lines.

As mentioned earlier, it is thought that the geometric projection would be particularly appropriate in a three-dimensional milieu where the point of observation is

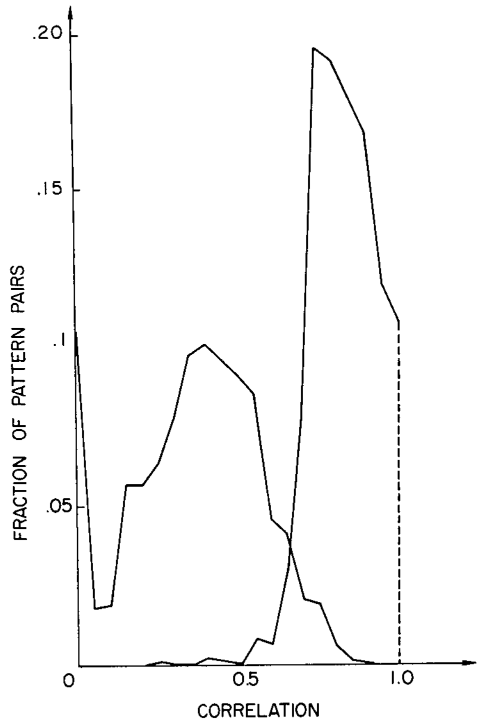


FIG. 8. Stability of transformation. Distribution of the correlation coefficient on first (left curve) and second (right curve) application of the geometric projection.

unknown. Such an application would be most welcome for further development of this approach.

Acknowledgments. The use of the geometric transformation for the normalization of handprinted characters was first suggested to us by Glen Shelton, Jr., of the IBM Thomas J. Watson Research Center. We have continually benefited from conversations with him and with R. G. Casey of the same laboratory; their experience with character recognition systems and scaling transformations in particular saved us much toil and trouble. We are also grateful for programming assistance to K. Abdali and to G. Basque.

RECEIVED JULY, 1969 REVISED NOVEMBER, 1969

REFERENCES

1. UGAWA, K., TORIWAKI, J., AND SUGINO, K. Normalization and recognition of two-dimensional patterns and linear distortion by moments. *Electron. Comm. Jap.* (June 1967), 34–46.
2. CASEY, R. G. Moment normalization of handprinted characters. Submitted to a technical journal.
3. BAKIS, R., HERBST, N., AND NAGY, G. An experimental study of machine recognition of hand-printed characters. *IEEE Trans. SSC-4*, 2, (July 1968), 119–132.
4. MUNSON, J. H. The recognition of handprinted text. In *Pattern Recognition*, L. Kanal (Ed.), MDI Publications, Wayne, Pa., 1968.
5. American Society of Photogrammetry. *Manual of Photogrammetry.* Washington, D. C., 1952.
6. TUONG, N. L'emploie des projections coniques dans la reconnaissance des formes. Tech. Rep. No. 1, Comput. Sci. Dep., Université de Montreal, Montreal, Que., Canada, Mar. 1969.

Full Table Quadratic Searching for Scatter Storage

A. COLIN DAY
University College London, London, England

The quadratic residue search method for hash tables avoids much of the clustering experienced with a linear search method. The simple quadratic search only accesses half the table. It has been shown that when the length of the table is a prime of the form $4n + 3$, where n is an integer, the whole table may be accessed by two quadratic searches plus a separate access for the original entry point. A search method is presented which is computationally simple, has all the advantages of the quadratic search, and yet accesses all the table in one sweep.

KEY WORDS AND PHRASES: quadratic residue, search method, hash tables, scatter storage, dictionary look-up, quadratic search, searching, hashing, hash code, clustering, collisions
CR CATEGORIES: 3.7, 3.73, 3.74

1. Introduction

Scatter storage, or hash coding, is a means of storing a table of items (e.g. identifier names in a compiler) in order to conserve storage and yet permit rapid searching of the table for a particular datum. There are two aspects of scatter storage which are of special interest. The first is the question of transforming (or *hashing*) the new datum into an initial address in the table. This is commonly performed by dividing the datum by the length of the table, and using the remainder as an index within the table. The second aspect is the choice of the action to be followed when a collision occurs, i.e. when two items hash to the same initial address. The various actions adopted are called *search methods*.

The simplest search method is the linear one. This is to step on through the table using a fixed step, and regarding the table as circular. There is no value to be gained in having any other step than 1 (providing that the initial hashing gives a satisfactory spread of values). This means that if the table is of length p, and the initial hash address is k, then locations $k, k + 1, k + 2, \cdots, k + p - 1[\text{mod } p]$ are searched. Morris [1] has pointed out that this is the worst strategy in use, as clustering of items easily results.

Maurer [2] has suggested a quadratic search. This involves searching successive addresses $(k + bi + ci^2)[\text{mod } p]$ for $i = 1, 2 \cdots, (p - 1)/2$. As Radke [3] points out, nothing is gained by choosing $b \neq 0$ or $c \neq 1$, so let us assume that the function is $(k + i^2)[\text{mod } p]$ for $i = 1, 2 \cdots, (p - 1)/2$. Only $(p - 1)/2$ locations are searched, but clustering is greatly reduced. For this method to work, p must be a prime number.

Radke [3] has suggested a method whereby the whole