# Low-dimensional procedure for the characterization of human faces

L. Sirovich and M. Kirby

*Division of Applied Mathemetics, Brown University, Providence, Rhode Island 02912*

Received August 25, 1986; accepted November 10, 1986

A method is presented for the representation of (pictures of) faces. Within a specified framework the representation is ideal. This results in the characterization of a face, to within an error bound, by a relatively low-dimensional vector. The method is illustrated in detail by the use of an ensemble of pictures taken for this purpose.

## 1. INTRODUCTION

The present investigation is concerned with the general problem of characterizing, identifying, and distinguishing individual patterns drawn from some well-defined class of patterns. Both the intuition for and the application of the methods comes from the particular problem of face identification. For this reason we use a terminology particular to this case, although the generality of approach will be apparent. Also, although a goal could be the investigation of the human ability to distinguish faces, the treatment given below can only be regarded as a paradigm for such a task. This is said since we offer no experimental procedure for verifying or refuting that our method bears in any way on our faculties for face recognition. However, a small speculation appears in Section 7.

The treatment presented here is based on a method known as the Karhunen–Loeve expansion in pattern recognition[1,2] and as factor or principal-component analysis in the statistical literature.[3] The applications of this procedure, especially in the analysis of signals in the time domain, is extensive, and no attempt is made to cite these studies. In brief, we demonstrate that any particular face can be economically represented in terms of a *best* coordinate system that we term *eigenpictures*. These are the eigenfunctions of the averaged covariance of the ensemble of faces. To give some idea of the data compression gained from this procedure, we first observe that a fairly acceptable picture of a face can be constructed from the specification of gray levels at $2^{14}$ pixel locations. Instead of this, we show, through actual construction, that roughly 40 numbers giving the admixture of eigenpictures characterize a face to within 3% error. Thus, in principle, any collection of faces could be classified by storing a small collection of numbers for each face and a small set of standard pictures known as eigenpictures.

## 2. FORMULATION

An individual face or picture $\varphi(\mathbf{x})$ is a scalar function of position $\mathbf{x} = (x_1, x_2)$. It furnishes the gray level $\varphi$ of the picture at each location $\mathbf{x}$. In the case treated here, a picture will be a full face recorded according to a normalization to be specified later. Since one theme of this paper is concerned with data management and reduction it is appropriate to be specific about the way in which a face is actually *captured*. Individual faces were recorded by a *frame grabber* (see Section 5) that stored each picture in digitalized form

$$\varphi(\mathbf{x}) \approx \varphi_{ij} = (\varphi)_{ij}. \tag{1}$$

For purposes of exposition, it sometimes is convenient to regard the matrix of gray scales $\varphi_{ij}$ as a vector $\varphi$, e.g., as the concatenation of rows of $\varphi_{ij}$. In a typical case, the picture was divided into $128 \times 128$ picture elements, or pixels, and a gray scale was determined at each pixel. An alternative and essentially equivalent way to digitalize a picture is through a Fourier transform

$$\varphi(\mathbf{x}) \approx \sum_{|m|,|n| \le N} a_{mn} \exp[2\pi i(nx_1 + mx_2)], \tag{2}$$

in which case the finite matrix $a_{nm}$ is the approximating form of the continuous picture.

We consider an ensemble of pictures $\{\varphi^{(n)}\}$, $n = 1, \ldots, M$, with $M$ assumed to be large enough. $\varphi^{(n)}$ will be used to denote a particular picture either in its continuous form or in one of its approximate forms. The average face is denoted by

$$\bar{\varphi} = \langle \varphi \rangle = \frac{1}{M} \sum_{n=1}^{M} \varphi^{(n)}. \tag{3}$$

It seems reasonable to assume that an efficient procedure for recognizing and storing pictures concentrates on departures from the mean. With this in mind, the deviation or departure from the mean

$$\phi^{(n)} = \varphi^{(n)} - \bar{\varphi} \tag{4}$$

is defined. We refer to a $\phi$ as a caricature. To make matters more concrete, Fig. 1 shows the average of the ensemble used by us, and Fig. 2 shows the comparison between a sample member of the ensemble and the corresponding caricature. (Details of the procedures for obtaining these pictures are given in Section 5.)

## 3. AN OPTIMAL REPRESENTATION

For the moment, it is convenient to consider each caricature as a vector $\phi^{(n)}$ and also to consider the ensemble of vectors

0740-3232/87/030519-06$02.00 

Fig. 1. Average face based on an ensemble of 115 faces. In this, as in the other plates, we have refrained from filtering out the high frequencies produced by the digitization. A pleasanter picture can be had by the usual trick of squinting or otherwise blurring the picture.

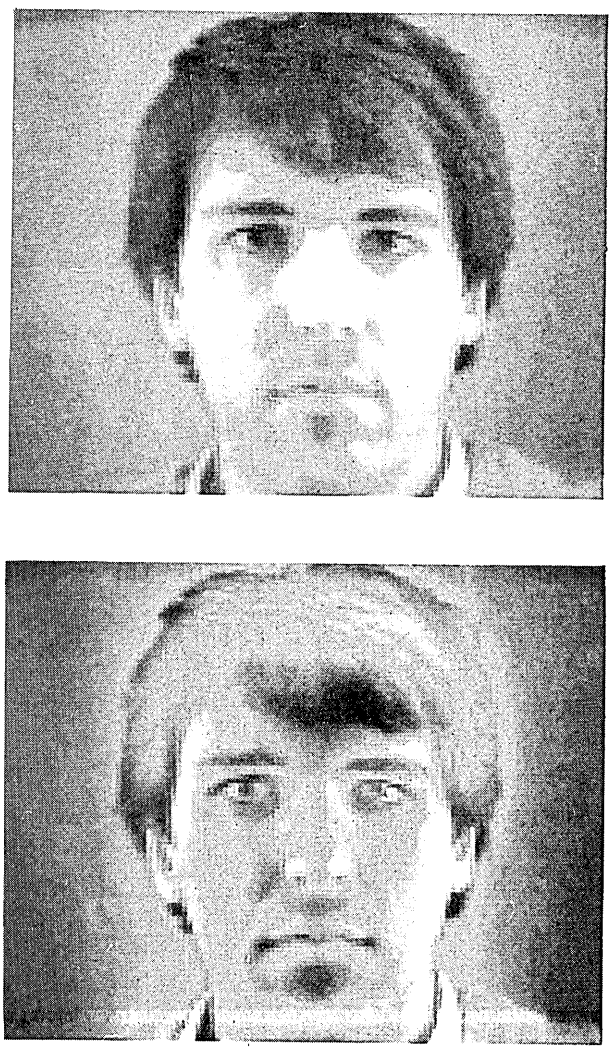

Fig. 2. Sample face on top and its caricature below it.

$\{\phi^{(n)}\}$. In the case actually considered the dimensionality of the space, $(128)^2 = 2^{14}$, is quite large. There is nothing particularly natural about this coordinate system, nor is there anything natural about the coordinate system corresponding to the Fourier representation [Eq. (2)].

We can start the development with the premise that only a relatively small number of dimensions should be necessary for pictures to be identified. This assertion is rooted first in the idea that humans are able to store and recognize enormous numbers of faces and, second, since recognition is *instantaneous* it is conceivable that we do it by processing picture information by anything so elaborate as the digital or Fourier methods just discussed. To place this assertion in geometrical terms, the view is that the endpoints of the vectors of the ensemble $\{\phi^{(n)}\}$, as $M$ becomes large, lie in a relatively low-dimensional space. To use a current idea, this asserts that the fractal dimension of the space of these endpoints is small.[4] To demonstrate this directly would require an $M$ that is several orders of magnitude larger than the nominal $2^{14}$ and is thus not feasible. (In Section 7 we indicate a somewhat more reasonable approach to this issue.) We now seek a natural coordinate system for representing the ensemble. In a sense to be described this will also be an optimal coordinate system. The treatment given below parallels the Karhunen–Loeve method[1–3] and is given in order to make our exposition self-contained.

The members of the ensemble $\{\phi^{(n)}\}$ are regarded as having been suitably normalized (see Section 5). We seek a system of orthonormal vectors $\{\mathbf{u}^{(n)}\}$,

$$(\mathbf{u}^{(n)}, \mathbf{u}^{(m)}) = \delta_{mn}, \tag{5}$$

under the usual Euclidean inner product that is optimal by the following criteria: Choose $\mathbf{u}^{(1)}$ such that

$$\lambda_1 = \frac{1}{M}\sum_{n=1}^{M} (\mathbf{u}^{(1)}, \phi^{(n)})^2 \tag{6}$$

is a maximum, subject to the condition that $[\mathbf{u}^{(1)}, \mathbf{u}^{(1)}] = 1$. (The coefficient of $1/M$ is inserted for later convenience.) This forces the unit vector $\mathbf{u}^{(1)}$ to be *central* to the ensemble $\{\phi^{(n)}\}$. Otherwise said, on average the members of $\{\phi^{(n)}\}$ have their greatest component in the direction $\mathbf{u}^{(1)}$. In general, the $k$th vector, $\mathbf{u}^{(k)}$, is chosen such that

$$\lambda_k = \frac{1}{M}\sum_{n=1}^{M} (\mathbf{u}^{(k)}, \phi^{(n)})^2 = \langle (\mathbf{u}^{(k)}, \phi)^2 \rangle \tag{7}$$

is a maximum, subject to the side conditions

$$(\mathbf{u}^{(k)}, \mathbf{u}^{(l)}) = \delta_{kl}, \qquad l \leqslant k. \tag{8}$$

For an alternative view of the problem just posed, consider the matrix

$$\mathbf{C} = \frac{1}{M}\sum_{n=1}^{M} \phi^{(n)}\phi^{(m)}, \tag{9}$$

where each term of the sum signifies a dyadic product. The matrix $\mathbf{C}$ is symmetric and nonnegative, and its eigenvalues and orthonormal eigenvectors are just

$$\mathbf{C}\mathbf{u}^{(n)} = \lambda^{(n)}\mathbf{u}^{(n)}. \tag{10}$$

In fact one recognizes conditions (7) and (8) as characterizing the extremal properties of the eigenvalues of the matrix $\mathbf{C}$.

**C**, which we may also write as

$$\mathbf{C} = \langle \phi \phi \rangle, \tag{11}$$

will be recognized as the ensemble average of the two-point correlation of the caricatures. More exactly, it is the discrete (in space) version of

$$\mathbf{C}(\mathbf{x}, \mathbf{y}) = \langle \phi(\mathbf{x})\phi(\mathbf{y}) \rangle = \frac{1}{M} \sum_{n=1}^{M} \phi^{(n)}(\mathbf{x})\phi^{(n)}(\mathbf{y}), \tag{12}$$

the actual two-point spatial correlation function.

Under the limit $M \uparrow \infty$ the symmetric nonnegative kernel $\mathbf{C}(\mathbf{x}, \mathbf{y})$ falls into a standard mathematical framework. Namely, since it is symmetric and square integrable, it follows from Mercer's theorem[5] that

$$\mathbf{C}(\mathbf{x}, \mathbf{y}) = \sum_{n=1}^{\infty} \lambda_n u^{(n)}(\mathbf{x}) u^{(n)}(\mathbf{y}), \tag{13}$$

where $\{u_n\}$ are the orthonormal eigenfunctions and $\{\lambda_n\}$ the corresponding eigenvalues. It then also follows that

$$\phi(\mathbf{x}) = \sum_{n=1}^{\infty} a_n u^{(n)}(\mathbf{x}) \tag{14}$$

in the $L^2$ sense.

## 4. EIGENPICTURES

As we have seen, the optimal representation of the ensemble of caricatures $\{\phi^{(n)}\}$ is equivalent to determining the eigenvectors of **C** [Eq. (10)]. Since the matrix **C** in our specific calculation is $2^{14} \times 2^{14}$ this problem is beyond the power of currently envisaged computers. However, if the number in the ensemble $M$ is less than the dimension of **C**, then **C** is singular and cannot be of order greater than $M$. (For the case actually considered by us, $M = 115$; see Section 5.) The analysis is simplified and follows from standard methods in linear algebra.[6]

We can proceed with the calculation of a typical eigenvector **u** by writing

$$\mathbf{u} = \sum_{k=1}^{M} a_k \phi^{(k)}, \tag{15}$$

which on substitution into Eq. (10) results in the simpler problem

$$\sum_{n=1}^{M} L_{mn} a_n = \lambda a_m, \tag{16}$$

where

$$L_{mn} = (\phi^{(m)}, \phi^{(n)}). \tag{17}$$

is a nonnegative symmetric matrix of dimension $M$. This procedure results in the determination of just $M$ of the eigenvectors of **C**. The remaining eigenvectors belong to the null space of this degenerate matrix. It is interesting to observe that the **u** that we term eigenpictures are formed by admixtures of members of the ensemble [Eq. (15)].

Another observation is that in this format the degree of digitalization of a picture, within limits, plays no role. For example, in the limit of a continuously formed picture, we have Eq. (12). Then the eigenfunction problem is

$$\int \mathbf{C}(\mathbf{x}, \mathbf{y}) u(\mathbf{y}) d\mathbf{y} = \lambda u(\mathbf{x}). \tag{18}$$

This is solved in the same way, viz., by writing

$$u(\mathbf{x}) = \sum_{n=1}^{M} a_n \phi^{(n)}(\mathbf{x}). \tag{19}$$

We are again led to Eq. (16) but with

$$L_{mn} = \int \phi^{(m)}(\mathbf{x}) \phi^{(n)}(\mathbf{x}) d\mathbf{x}. \tag{20}$$

The dimension of **L** remains the same, but the entries change with the degree of graininess.

## 5. PROCEDURES AND RESULTS

In order to examine the worth of the procedures just described, we assembled a file of 115 pictures. Individuals were drawn mainly from the undergraduate male population at Brown University. Also, since our initial goal was to demonstrate feasibility, we endeavored to create a relatively homogeneous population, viz., smooth-skinned caucasian males; but see Section 7. Beyond this, no other selection procedure was used. Passing students were simply asked to give a moment of their time to have their pictures taken.

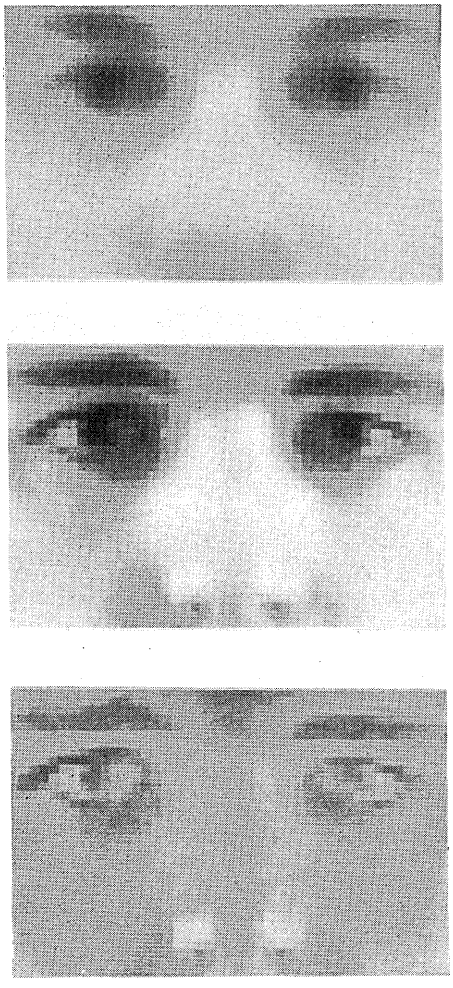

Fig. 3. Cropped faces: upper, the average; middle, a sample face; and bottom, its caricature.

An individual face was video recorded and digitized at $2^7 \times 2^7$ pixels with $2^8$ gray level by means of an IVS-100 image processor. Faces were lined up by a cross-hair overlay display that appeared on a video monitor. The vertical line passed through the symmetry line of the face and the horizontal line through the pupils of the eyes. Field depth was adjusted so that facial width was the same for images. Since these steps were all adjusted by eye, this contributed to the general error level. The pictures were taken under background-lighting conditions. Since the lighting varied with the time of day, this too was a source of error. To some extent, this error was diminished by a normalization procedure that we now describe.

A face, or for that matter any object, can be regarded as a pointwise map of reflectivities, say, $r(\mathbf{x})$. Under a uniform illumination, say, $I$, the face is given by

$$\tilde{\varphi}(\mathbf{x}) = Ir(\mathbf{x}). \tag{21}$$

For a variety of reasons, it is important to normalize a picture so that a reference portion of a face is at a standard level of illumination. If we denote a reference point by $\mathbf{x}_0$ and standard light level by $I_0$, then we take the normalized picture to be

$$\varphi(\mathbf{x}) = \frac{I_0}{\tilde{\varphi}(\mathbf{x}_0)} \tilde{\varphi}(\mathbf{x}). \tag{22}$$

In actual practice we took the reference portion to be small, high cheek areas below each eye and averaged the light level over these two patches. This procedure provides a specific light-level normalization. In addition, it provides the basis for the future identification of a picture.

The average face was computed according to Eq. (3) and is

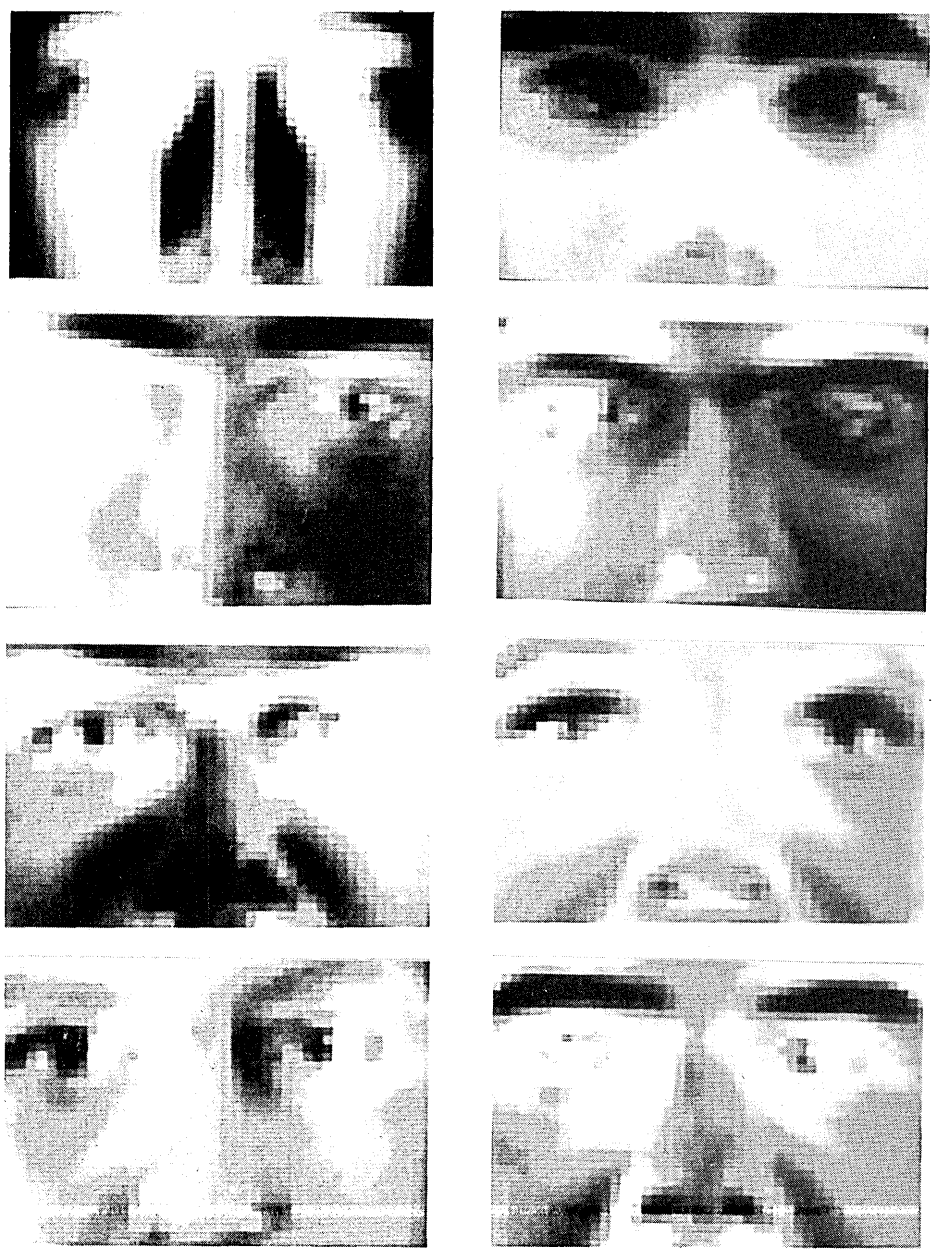

Fig. 4. First eight eigenpictures starting at upper left, moving to the right, and ending at lower right.

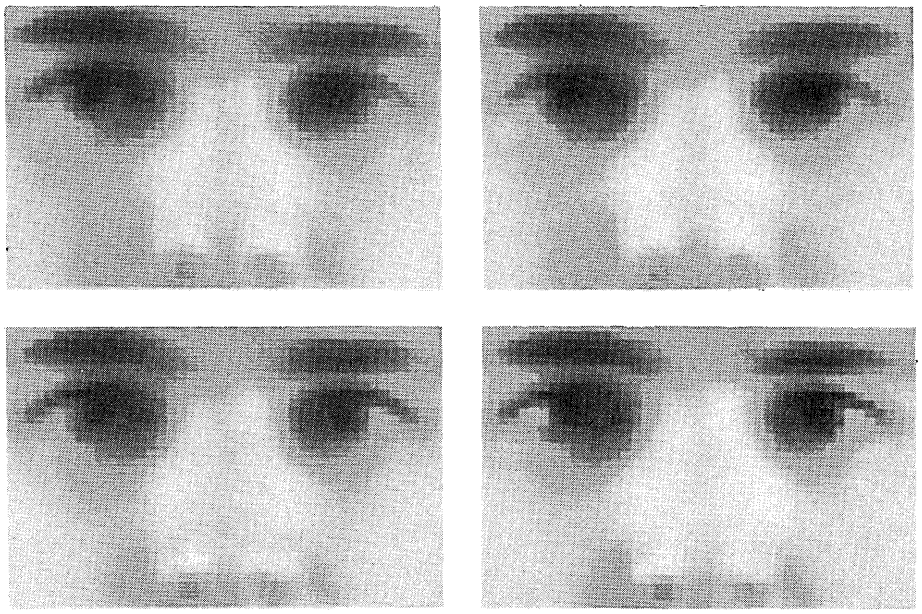

Fig. 5. Approximation to the exact picture (middle panel of Fig. 3) using 10, 20, 30, and 40 eigenpictures.

thus the pixel-by-pixel average of gray levels. The result is shown in Fig. 1. The caricature as defined by Eq. (4) was formed for each picture and, as illustrated in Fig. 2, a picture and its caricature appear virtually the same to us. A possible explanation is that our own visual apparatus does a similar subtraction.

For several reasons we mainly considered a cut-down version of the problem. Pictures were cropped to include only the eyes and nose. This is illustrated in Fig. 3, where the average, a member of the ensemble, and the corresponding caricature are shown. The correlation matrix was then formed and all the *eigenpictures* determined. The first eight eigenpictures are shown in Fig. 4. For purposes of illustration these have been slightly *doctored*. Since the eigenpictures have negative entries we have added to each a *pedestal* or background for purposes of viewing. (Since the multiplication of an eigenpicture by any constant is still an eigenpicture, this gives additional variation to their presentation.)

## 6. EIGENPICTURE CONSTRUCTION

We recall that each eigenpicture, according to Eq. (15), is an admixture of members of the ensemble. In addition, these eigenpictures, according to the criteria of Section 4, are the optimal set by which to represent a picture. To examine this assertion, we consider the fit of a typical picture by eigenpictures. For any member of the population we can write

$$\varphi = \bar{\varphi} + \sum_{n=1}^{M} a_n \mathbf{u}^{(n)}, \tag{23}$$

where

$$a_n = (\mathbf{u}^{(n)}, \varphi - \bar{\varphi}). \tag{24}$$

We next consider how good a partial fit is, viz., to what degree

$$\varphi \approx \bar{\varphi} + \sum_{n=1}^{N} a_n \mathbf{u}^{(n)} = \varphi^N \tag{25}$$

is a good fit for various values of $N$. In Fig. 5 we show the result of taking the partial terms $N = 10, 20, 30, 40$. This should be compared with the exact picture contained in Fig. 3.

A qualitative measure of the *goodness of fit* is given by

$$E_N = \|\varphi - \varphi^N\| / \|\varphi\|. \tag{26}$$

This is plotted (the ordinate is $100 \times E_N$) versus $N$ in Fig. 6 for the typical member of the set shown in Fig. 3. The dashed curve in this figure depicts the average error over a set of 10 members of the ensemble chosen at random.

From the form of eigenvalue $\lambda_n$, [Eq. (7)] it follows that $\sqrt{\lambda_n}$ measures the degree to which the population $\{\phi^{(n)}\}$ falls along the $n$th direction. It is conceptually convenient to put this in the form of a probability

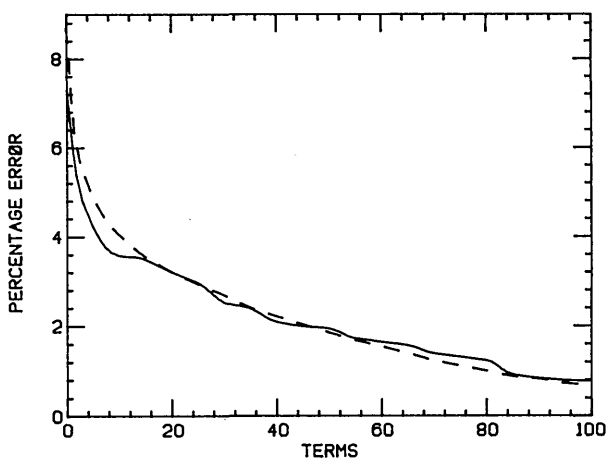


Fig. 6. Percent error versus number of eigenpictures used in the approximation. Solid curve is for picture shown in Fig. 2 (see also Fig. 5). Dashed curve is average over 10 different sample faces.

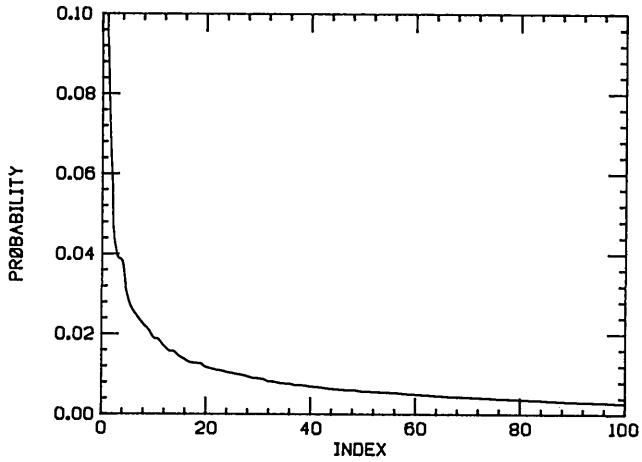


Fig. 7. Probability projection along the eigenpicture directions; see Eq. (27).

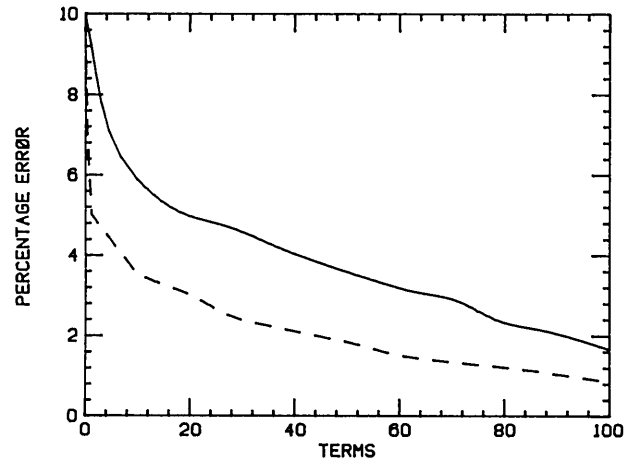


Fig. 8. Percent error versus number of eigenpictures for the full face shown in Fig. 2. The dashed curve is the corresponding cropped face and is the same as in Fig. 6.

$$p_n = \frac{\sqrt{\lambda_n}}{\sum_k \sqrt{\lambda_k}}. \tag{27}$$

This is plotted in Fig. 7, and, as the curve shows, the population lies overwhelmingly along the principal eigenvectors.

## 7. DISCUSSION

Thus far we have focused our attention on the ensemble of cropped pictures, as illustrated in Fig. 3, and, in addition, restricted our calculations to members of the ensemble. It is of interest to discuss what occurs when we depart from both of these restrictions.

To discuss the second point first, we took pictures of three randomly selected people not from the population and applied our fitting procedure to their cropped faces. In one case lighting conditions were deliberately chosen to be poor and quite different from those used in gathering the ensemble. Since the error was down to 7.8% after 40 terms, we regarded this as a success. It also lends further justification to the normalization condition (22) discussed earlier. The other two nonensemble faces were of females, and at $N = 40$ the error was down to 3.9% in one instance and down to 2.4% in the other. Since the ensemble was made up only of males, this strengthens confidence in the method presented and incidentally shows that the cropped features are, relatively speaking, gender independent (as long as makeup is not applied).

We turn next to the question of resolving a full face in terms of the corresponding eigenpicture development. It should be apparent that the number of eigenpictures needed to fit a picture, to within some error bound, increases as the number of features increases. This is illustrated in Fig. 8, which contrasts the error curve of the cropped face with the analogous curve that we calculated for the full face, as depicted in Figs. 1 and 2. We have also considered an intermediate version in which the hair and background are removed from a picture. As might be expected, this curve falls between those shown in Fig. 8.

Although we have refrained from imagining by what means we, as humans, process information of this sort, one suggestion can be made. If a face is compartmentalized, even beyond the cropping used by us (say, eyes, nose, mouth), then each can be fitted by relatively few eigenpictures. That is, if a face is identified by its parts, then an economical scheme results. (Family members are often described as having another's eyes, mouth, etc., which would seem to support this idea.)

A seemingly different, but actually related, idea pertains to the dimension of the manifold of all faces, a notion already mentioned in Section 3. As mentioned there, a space of $2^{14}$ dimensions, in which one supplies gray levels at each of the corresponding pixels, is sufficient to construct an adequate likeness. On the basis of our construction, we can greatly reduce this estimate. In fact, the procedure presented here suggests that fewer than 100 eigenpictures are necessary to fit a picture. Correspondingly, fewer than 100 dimensions are needed to provide a likeness. Our method can be regarded as giving an estimate on the upper bound of the number of dimensions of the (fractal) set in which the space of *all* pictures fits. Further, if the coordinate system based on the eigenpictures is now used in an actual calculation of the dimension of this set, far fewer pictures would be required. This calculation, in addition to being of general interest, might even give a clue to how we humans do the job. Unfortunately, we do not have a sufficiently large ensemble of faces at present for this purpose.

## ACKNOWLEDGMENT

The research reported here was supported in part by the National Science Foundation under grant MCS-77-08598.

## REFERENCES

1. K. Fukunaga, *Introduction to Statistical Pattern Recognition* (Academic, New York, 1972).
2. R. B. Ash, and M. F. Gardner, *Topics in Stochastic Processes* (Academic, New York, 1975).
3. N. Ahmed, and M. H. Goldstein, *Orthogonal Transforms for Digital Signal Processing* (Springer-Verlag, New York, 1975).
4. B. B. Mandelbrot, *The Fractral Geometry of Nature* (Freeman, San Francisco, Calif., 1982).
5. F. Riesz, and B. Sz.-Nagy, *Functional Analysis* (Ungar, New York, 1955).
6. R. Bellman, *Introduction to Matrix Analysis* (McGraw-Hill, New York, 1960).