International Journal of Computer Vision 43(3), 167–188, 2001

Face Recognition Using the Discrete Cosine Transform

ZIAD M. HAFED AND MARTIN D. LEVINE
Center for Intelligent Machines, McGill University, 3480 University Street, Montreal, Canada H3A 2A7
zhafed@cim.mcgill.ca
levine@cim.mcgill.ca

Abstract. An accurate and robust face recognition system was developed and tested. This system exploits the feature extraction capabilities of the discrete cosine transform (DCT) and invokes certain normalization techniques that increase its robustness to variations in facial geometry and illumination. The method was tested on a variety of available face databases, including one collected at McGill University. The system was shown to perform very well when compared to other approaches.

Keywords: Face recognition, discrete cosine transform, Karhunen-Loeve transform, geometric normalization, illumination normalization, feature extraction, data compression

1. Introduction

Face recognition by humans is a high level visual task for which it has been extremely difficult to construct detailed neurophysiological and psychophysical models. This is because faces are complex natural stimuli that differ dramatically from the artificially constructed data often used in both human and computer vision research. Thus, developing a computational approach to face recognition can prove to be very difficult indeed. In fact, despite the many relatively successful attempts to implement computer-based face recognition systems, we have yet to see one which combines speed, accuracy, and robustness to face variations caused by 3D pose, facial expressions, and aging. The primary difficulty in analyzing and recognizing human faces arises because variations in a single face can be very large, while variations between different faces are quite small. That is, there is an inherent structure to a human face, but that structure exhibits large variations due to the presence of a multitude of muscles in a particular face. Given that recognizing faces is critical for humans in their everyday activities, automating this process would be very useful in a wide range of applications including security, surveillance, criminal identification, and video compression. This paper discusses a new computational approach to face recognition that, when combined with proper face localization techniques, has proved to be very efficacious.

This section begins with a survey of the face recognition research performed to date. The proposed approach is then presented along with its objectives and the motivations for choosing it. The section concludes with an overview of the structure of the paper.

1.1. Background and Related Work

Most research on face recognition falls into two main categories (Chellappa et al., 1995): feature-based and holistic. Feature-based approaches to face recognition basically rely on the detection and characterization of individual facial features and their geometrical relationships. Such features generally include the eyes, nose, and mouth. The detection of faces and their features prior to performing verification or recognition makes these approaches robust to positional variations of the faces in the input image. Holistic or global approaches to face recognition, on the other hand, involve encoding the entire facial image and treating the resulting facial “code” as a point in a high-dimensional space. Thus, they assume that all faces are constrained to particular positions, orientations, and scales.

Feature-based approaches were more predominant in the early attempts at automating the process of face recognition. Some of this early work involved the use of very simple image processing techniques (such as edge detection, signatures, and so on) for detecting faces and their features (see, for example, Sakai et al., 1969; Kelly, 1970). In Sakai et al. (1969), an edge map was first extracted from an input image and then matched to a large oval template, with possible variations in position and size. The presence of a face was then confirmed by searching for edges at estimated locations of certain features like the eyes and mouth. Kelly (1970) used an improved edge detector involving heuristic planning to extract an accurate outline of a person's head from various backgrounds.

More recently, Govindaraju et al. (1990) proposed a technique for locating a face in a cluttered image that employed a deformable template similar to the ones used in Yuille et al. (1989). They based their template on the outline of the head and allowed it to deform according to certain spring-based models. This approach performed quite well when tested on a small data set, but it sometimes gave rise to false alarms (Govindaraju et al., 1990). Other recent approaches have used hierarchical coarse-to-fine searches with template-based matching criteria (Burt, 1989; Craw et al., 1992; Shepherd, 1985).

Once a face has been located, its features must be computed. Early examples of this are the work of Kanade (1973) and Harmon (Harmon and Hunt, 1977) who worked with facial profile features. An interesting recent discussion of feature-based methods, which compares them to holistic approaches, is found in Brunelli and Poggio (1993).

A successful holistic approach to face recognition uses the Karhunen-Loeve transform (KLT). This transform exhibits pattern recognition properties that were largely overlooked initially because of the complexity involved in its computation (Chellappa et al., 1995). Kirby and Sirovich (1990) originally proposed the KLT to characterize faces. This transform produces an expansion of an input image in terms of a set of basis images or the so-called "eigenimages." Turk and Pentland (1991) proposed a face recognition system based on the KLT in which only a few KLT coefficients were used to represent faces in what they termed "face space." Each set of KLT coefficients representing a face formed a point in this high-dimensional space. The system performed well for frontal mug shot images (Turk and Pentland, 1991). Specifically, it was tested on a database of 16 individuals, but with several images per person. These images covered changes in scale, orientation, and lighting. The authors reported 96% correct classification over lighting variations, 85% over orientation variations, and 64% over size variations.

The KLT does not achieve adequate robustness against variations in face orientation, position, and illumination (as seen in the above results). That is why it is usually accompanied by further processing to improve its performance. For example, in Akamatsu et al. (1991), operations were added to the KLT method to standardize faces with respect to position and size. Also, in Pentland et al. (1994), the authors still used the KLT, but now on particular features of a face. These features became part of the "feature space," and a distance-to-feature-space (DFFS) metric was used to locate them in an image (such localization could serve as a pre-processing stage for later normalization, cropping, and classification). A similar idea of using 'local' information was presented in Lades et al. (1993). An artificial neural network, which employed the so-called dynamic link architecture (DLA), was used to achieve distortion-invariant recognition. Local descriptors of the input images were obtained using Gabor-based wavelets. By conveying frequency, position, and orientation information, this approach performed well on relatively large databases.

Yet another holistic approach to face recognition is that based on linear discriminant analysis (LDA) (see Swets and Weng, 1996; Belhumeur et al., 1997). In this approach, Fisher's linear discriminant (Duda and Hart, 1973) is used (on the space of feature vectors obtained by the KLT) to obtain the most discriminating features of faces, rather than the most expressive ones given by the KLT alone (Swets and Weng, 1996). In both Swets and Weng (1996) and Belhumeur et al. (1997), LDA resulted in better classification than in the case of the KLT being applied alone, especially under varying pose and illumination conditions.

As can be seen, there are merits to both feature-based and holistic approaches to face recognition, and it seems that they may both be necessary to meet the two main objectives of a face recognition system: accuracy and robustness. Holistic approaches may be accurate for simple frontal mug shots, but they must be accompanied by certain feature-based techniques to make them more robust. In fact, this may be true for humans as well. Both holistic information and feature information are essential for human recognition of faces. It is possible that global descriptions serve as a front-end for more detailed feature-based perception (Bruce, 1990).

1.2. *Approach and Motivation*

This paper investigates an alternative holistic method for face recognition and compares it to the popular KLT approach. The basic idea is to use the discrete cosine transform (DCT) as a means of feature extraction for later face classification. The DCT is computed for a cropped version of an input image containing a face, and only a small subset of the coefficients is maintained as a feature vector. This feature vector may be conceived of as representing a point in a high-dimensional "face" space as in Turk and Pentland (1991). Classification is based on a simple Euclidean distance measure. To improve performance, various normalization techniques are invoked prior to recognition to account for small perturbations in facial geometry and illumination.

The main merit of the DCT is its relationship to the KLT. As is known, the KLT is an optimal transform based on various performance criteria (Rosenfeld and Kak, 1976). However, it is a statistical transform that is only defined after specifying the characteristics of the data set it is being applied to. Of the deterministic discrete transforms, the DCT best approaches the KLT (Jain, 1989). Thus, it is expected that it too will exhibit desirable pattern recognition capabilities. If this is shown to be the case, then the use of the DCT in face recognition becomes of more value than the KLT because of its computational speed.[1] In fact, because of the popularity of the JPEG image storage format (which is based on the DCT (Pennebaker and Mitchell, 1993)), large efforts have been concentrated on developing fast algorithms for computing the DCT (Rao and Yip, 1990). Furthermore, the KLT is not only more computationally intensive, but it must also be redefined every time the statistics of its input signals change. Therefore, in the context of face recognition, the eigenvectors of the KLT (eigenfaces) should *ideally* be recomputed every time a new face is added to the training set of known faces (Turk and Pentland, 1991).

This paper compares the DCT to the KLT in order to justify the use of the first in face recognition. The mathematical relationship between the two transforms is briefly described, and certain face recognition tests are performed to support the hypothesis that the DCT is indeed suitable for such an application.

1.3. *Overview of the Paper*

Following this introduction, Section 2 presents the mathematical definition of the discrete cosine transform as well as its relationship to the KLT. Then, Section 3 discusses the basics of a face recognition system using the discrete cosine transform. It details the proposed algorithm and discusses the various parameters that may affect its performance. It also explains the pre-processing steps involved prior to the use of the DCT in recognition in order to improve its performance. Section 4 highlights the proposed system's performance based on experimental results. Finally, the paper ends with conclusions and suggestions for future work.

2. The Discrete Cosine Transform

Data compression is essential for both biological and computer signal processing. In fact, at the retinal level, only approximately 1 million signals (out of almost 130 million from the photoreceptors) are projected to the lateral geniculate nucleus (LGN) for further processing, resulting in data compression of the order of 100:1 (Sekuler and Blake, 1994). By the time biological signals arrive at the higher visual centers of the brain, they are transformed into signals conveying contrast (magnitude), phase, frequency, and orientation information, all of which are attributes of Fourier analysis. As will be seen in this section, data compression is the main feature of the discrete cosine transform. Also, since the DCT is related to the discrete Fourier transform (Rao and Yip, 1990), it can be computed efficiently. It is these two properties of the DCT that we seek for face recognition.

2.1. *Definition*

Ahmed, Natarajan, and Rao (1974) first introduced the discrete cosine transform (DCT) in the early seventies. Ever since, the DCT has grown in popularity, and several variants have been proposed (Rao and Yip, 1990). In particular, the DCT was categorized by Wang (1984) into four slightly different transformations named DCT-I, DCT-II, DCT-III, and DCT-IV. Of the four classes Wang defined, DCT-II was the one first suggested by Ahmed et al., and it is the one of concern in this paper.

Given an input sequence $u(n)$ of length N, its DCT, $v(k)$, is obtained by the following equation:

$$v(k) = \alpha(k) \sum_{n=0}^{N-1} u(n) \cos\left(\frac{(2n+1)\pi k}{2N}\right)$$
$$0 \le k \le N-1 \qquad (2.1a)$$

where

$$\alpha(0) = \sqrt{\frac{1}{N}}, \alpha(k) = \sqrt{\frac{2}{N}} \quad 1 \le k \le N-1 \qquad (2.1b)$$

Alternatively, we can think of the sequence $u(n)$ as a vector and the DCT as a transformation matrix applied to this vector to obtain the output $v(k)$. In this case, the DCT transformation matrix, $C = \{c(k,n)\}$, is defined as follows:

$$c(k,n) = \begin{cases} \frac{1}{\sqrt{N}} & k=0, \ 0 \le n \le N-1 \\ \sqrt{\frac{2}{N}} \cos\left(\frac{(2n+1)\pi k}{2N}\right) & 1 \le k \le N-1, \\ & 0 \le n \le N-1 \end{cases} \qquad (2.2)$$

where k and n are the row and column indices, respectively. Using Eq. (2.2), the DCT of the sequence $u(n)$ (or vector $\mathbf{u}$) is simply

$$\mathbf{v} = C\mathbf{u} \qquad (2.3)$$

The inverse discrete cosine transform permits us to obtain $u(n)$ from $v(k)$. It is defined by:

$$u(n) = \sum_{k=0}^{N-1} \alpha(k) v(k) \cos\left(\frac{(2n+1)\pi k}{2N}\right)$$
$$0 \le n \le N-1 \qquad (2.4)$$

with $\alpha(k)$ as given in Eq. (2.1b). Using Eq. (2.3), the inverse discrete cosine transform, $\mathbf{u}$, of a vector $\mathbf{v}$ is obtained by applying the inverse of matrix C to $\mathbf{v}$. That is, the inverse discrete cosine transform is found from

$$\mathbf{u} = C^{-1}\mathbf{v} \qquad (2.5)$$

From these definitions, we observe that by applying the discrete cosine transform to an input sequence, we simply decompose it into a weighted sum of basis cosine sequences. This is obvious from Eq. (2.4) in which $u(n)$ is reconstructed by a summation of cosines which are weighted by the DCT coefficients obtained from Eq. (2.1) or (2.3). These basis sequences of the DCT are the rows of the matrix C.

2.2. *Compression Performance in Terms of the Variance Distribution*

The Karhunen-Loeve transform (KLT) is a statistically optimal transform based on a number of performance criteria. One of these criteria is the variance distribution of transform coefficients. This criterion judges the performance of a discrete transform by measuring its variance distribution for a random sequence having some specific probability distribution function (Rao and Yip, 1990). It is desirable to have a small number of transform coefficients with large variances such that all other coefficients can be discarded with little error in the reconstruction of signals from the ones retained. The error criterion generally used when reconstructing from truncated transforms is the mean-square error (MSE).

In terms of pattern recognition, it is noted that dimensionality reduction is perhaps as important an objective as class separability in an application such as face recognition. Thus, a transform exhibiting large variance distributions for a small number of coefficients is desirable. This is so because such a transform would require less information to be stored and used for recognition. In this respect, as well as others, the DCT has been shown to approach the optimality of the KLT (Pratt, 1991).

The variance distribution for the various discrete transforms is usually measured when the input sequence is a stationary first-order Markov process (Markov-1 process). Such a process has an auto-covariance matrix of the form shown in Eq. (2.6) and provides a good model for the scan lines of gray-scale images (Jain, 1989). The matrix in Eq. (2.6) is a Toeplitz matrix, which is expected since the process is stationary (Jain, 1989). Thus, the variance distribution measures are usually computed for random sequences of length N that result in an auto-covariance matrix of the form:

$$R = \begin{bmatrix} 1 & \rho & \rho^2 & .. & \rho^{N-1} \\ \rho & 1 & \rho & .. & \rho^{N-2} \\ . & . & . & .. & . \\ . & . & . & .. & . \\ \rho^{N-1} & \rho^{N-2} & . & .. & 1 \end{bmatrix}$$

$$\rho \equiv \text{correlation coeff.}$$
$$|\rho| < 1 \qquad (2.6)$$

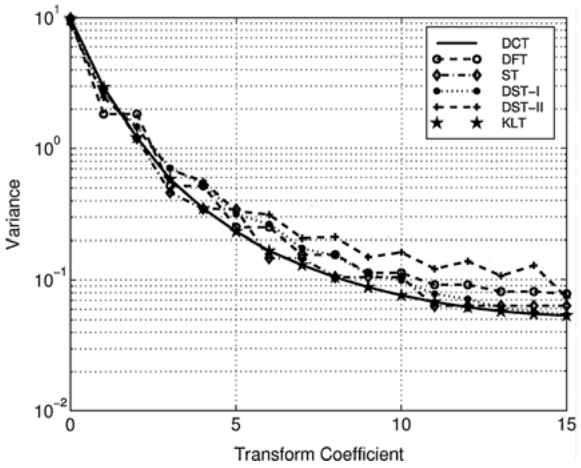

Figure 1. Variance distribution for a selection of discrete transforms for $N = 16$ and $\rho = 0.9$ (adapted from K.R. Rao and P. Yip, *Discrete Cosine Transform—Algorithms, Advantages, Applications*, New York: Academic, 1990). Data is shown for the following transforms: discrete cosine transform (DCT), discrete Fourier transform (DFT), slant transform (ST), discrete sine transform (type I) (DST-I), discrete sine transform (type II) (DST-II), and Karhunen-Loeve transform (KLT).

Figure 1 shows the variance distribution for a selection of discrete transforms given a first-order Markov process of length $N = 16$ and $\rho = 0.9$. The data for this curve were obtained directly from Rao and Yip (1990) in which other curves for different lengths are also presented. The purpose here is to illustrate that the DCT variance distribution, when compared to other deterministic transforms, decreases most rapidly. The DCT variance distribution is also very close to that of the KLT, which confirms its near optimality. Both of these observations highlight the potential of the DCT for data compression and, more importantly, feature extraction.

2.3. *Comparison with the KLT*

The KLT completely decorrelates a signal in the transform domain, minimizes MSE in data compression, contains the most energy (variance) in the fewest number of transform coefficients, and minimizes the total representation entropy of the input sequence (Rosenfeld and Kak, 1976). All of these properties, particularly the first two, are extremely useful in pattern recognition applications.

The computation of the KLT essentially involves the determination of the eigenvectors of a covariance matrix of a set of training sequences (images in the case of face recognition). In particular, given M training images of size, say, $N \times N$, the covariance matrix of interest is given by

$$\boldsymbol{C} = \boldsymbol{A} \cdot \boldsymbol{A}^T \tag{2.7}$$

where A is a matrix whose columns are the M training images (after having an average face image subtracted from each of them) reshaped into N^2-element vectors. Note that because of the size of A, the computation of the eigenvectors of C may be intractable. However, as discussed in Turk and Pentland (1991), because M is usually much smaller than N^2 in face recognition, the eigenvectors of C can be obtained more efficiently by computing the eigenvectors of another smaller matrix (see (Turk and Pentland, 1991) for details). Once the eigenvectors of C are obtained, only those with the highest corresponding eigenvalues are usually retained to form the KLT basis set. One measure for the fraction of eigenvectors retained for the KLT basis set is given by

$$\theta_\lambda = \frac{\sum_{l=1}^{M'} \lambda_l}{\sum_{l=1}^{M} \lambda_l} \tag{2.8}$$

where λ_l is the lth eigenvalue of C and M' is the number of eigenvectors forming the KLT basis set.

As can be seen from the definition of C in Eq. (2.7), the KLT basis functions are data-dependent. Now, in the case of a first-order Markov process, these basis functions can be found analytically (Rao and Yip, 1990). Moreover, these functions can be shown to be asymptotically equivalent to the DCT basis functions as ρ (of Eq. (2.6)) $\to 1$ for any given N (Eq. (2.6)) and as $N \to \infty$ for any given ρ (Rao and Yip, 1990). It is this asymptotic equivalence that explains the near optimal performance of the DCT in terms of its variance distribution for first-order Markov processes. In fact, this equivalence also explains the near optimal performance of the DCT based on a handful of other criteria such as energy packing efficiency, residual correlation, and mean-square error in estimation (Rao and Yip, 1990). This provides a strong justification for the use of the DCT for face recognition. Specifically, since the KLT has been shown to be very effective in face recognition (Pentland et al., 1994), it is expected that a deterministic transform that is mathematically related to it would probably perform just as well in the same application.

As for the computational complexity of the DCT and KLT, it is evident from the above overview that the KLT requires significant processing during training, since its basis set is data-dependent. This overhead in computation, albeit occurring in a non-time-critical off-line training process, is alleviated with the DCT. As for online feature extraction, the KLT of an $N \times N$ image can be computed in $O(M'N^2)$ time where M' is the number of KLT basis vectors. In comparison, the DCT of the same image can be computed in $O(N^2\log_2 N)$ time because of its relation to the discrete Fourier transform—which can be implemented efficiently using the fast Fourier transform (Oppenheim and Schafer, 1989). This means that the DCT can be computationally more efficient than the KLT depending on the size of the KLT basis set.[2]

It is thus concluded that the discrete cosine transform is very well suited to application in face recognition. Because of the similarity of its basis functions to those of the KLT, the DCT exhibits striking feature extraction and data compression capabilities. In fact, coupled with these, the ease and speed of the computation of the DCT may even favor it over the KLT in face recognition.

3. Face Recognition Using the Discrete Cosine Transform

3.1. Basic Algorithm

The face recognition algorithm discussed in this paper is depicted in Fig. 2. It involves both face normalization and recognition. Since face and eye localization is not performed automatically, the eye coordinates of the input faces need to be entered manually in order to normalize the faces correctly. This requirement is not a major limitation because the algorithm can easily be invoked after running a localization system such as the one presented in Jebara (1996) or others in the literature.

As can be seen from Fig. 2, the system receives as input an image containing a face along with its eye coordinates. It then executes both geometric and illumination normalization functions as will be described later. Once a normalized (and cropped) face is obtained, it can be compared to other faces, under the same nominal size, orientation, position, and illumination conditions. This comparison is based on features extracted using the DCT. The basic idea here is to compute the DCT of the normalized face and retain a certain subset of the DCT coefficients as a feature vector describing this face. This feature vector contains the low-to-mid frequency DCT coefficients, as these are the ones having the highest variance. To recognize a particular input face, the system compares this face's feature vector to the feature vectors of the database faces using a Euclidean distance nearest-neighbor classifier (Duda and Hart, 1973). If the feature vector of the probe is **v** and that of a database face is **f**, then the Euclidean distance between the two is

$$d = \sqrt{(f_0 - v_0)^2 + (f_1 - v_1)^2 + \cdots + (f_{M-1} - v_{M-1})^2} \tag{3.1}$$

where

$$\begin{aligned} \mathbf{v} &= [v_0 \quad v_1 \quad \ldots \quad v_{M-1}]^T \\ \mathbf{f} &= [f_0 \quad f_1 \quad \ldots \quad f_{M-1}]^T \end{aligned} \tag{3.2}$$

and M is the number of DCT coefficients retained as features. A match is obtained by minimizing d.

Note that this approach computes the DCT on the entire normalized image. This is different from the use of the DCT in the JPEG compression standard (Pennebaker and Mitchell, 1993), in which the DCT is computed on individual subsets of the image. The use of the DCT on individual subsets of an image, as in the JPEG standard, for face recognition has been proposed in Shneier and Abdel-Mottaleb (1996) and Eickeler et al. (2000).

Also, note that this approach basically assumes no thresholds on d. That is, the system described always assumes that the closest match is the correct match, and no probe is ever rejected as unknown. If a threshold q is defined on d, then the gallery face that minimizes d would only be output as the match when $d < q$. Otherwise, the probe would be declared as unknown. In this way, one can actually define a threshold to achieve 100% recognition accuracy, but, of course, at the cost of a certain number of rejections. In other words, the system could end up declaring an input face as unknown even though it exists in the gallery. Suitable values of q can be obtained using the so-called Receiver Operating Characteristic curve (ROC) (Grzybowski and Younger, 1997), as will be illustrated later.

3.2. Feature Extraction

To obtain the feature vector representing a face, its DCT is computed, and only a subset of the obtained coefficients is retained. The size of this subset is chosen such that it can sufficiently represent a face, but it can in fact

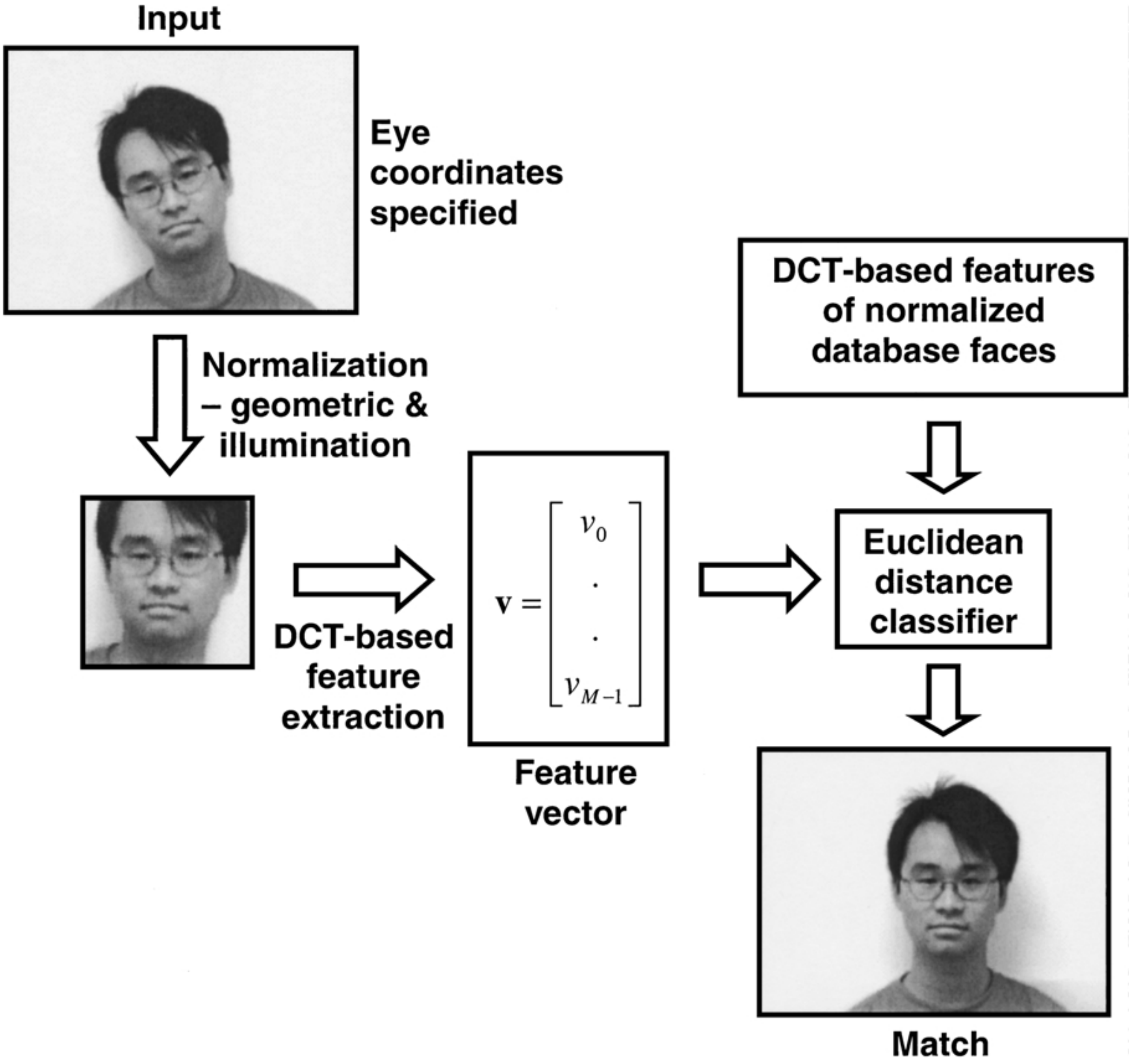
Input
Eye coordinates specified
Normalization – geometric & illumination
DCT-based feature extraction
$\mathbf{v} = \begin{bmatrix} v_0 \\ \cdot \\ \cdot \\ v_{M-1} \end{bmatrix}$
Feature vector
DCT-based features of normalized database faces
Euclidean distance classifier
Match

Figure 2. Face recognition system using the DCT.

be quite small, as will be seen in the next section. As an illustration, Fig. 3(a) shows a sample image of a face, and Fig. 3(b) shows the low-to-mid frequency 8 × 8 subset of its DCT coefficients. It can be observed that the DCT coefficients exhibit the expected behavior in which a relatively large amount of information about the original image is stored in a fairly small number of coefficients. In fact, looking at Fig. 3(b), we note that the DC term is more than 15,000 and the minimum magnitude in the presented set of coefficients is less than 1. Thus there is an order of 10,000 reduction in coefficient magnitude in the first 64 DCT coefficients. Most of the discarded coefficients have magnitudes less than 1. For the purposes of this paper, square subsets, similar to the one shown in Fig. 3(b), are used for the feature vectors.

It should be noted that the size of the subset of DCT coefficients retained as a feature vector may not be large enough for achieving an accurate reconstruction of the input image. That is, in the case of face recognition, data compression ratios larger than the ones necessary to render accurate reconstruction of input images are

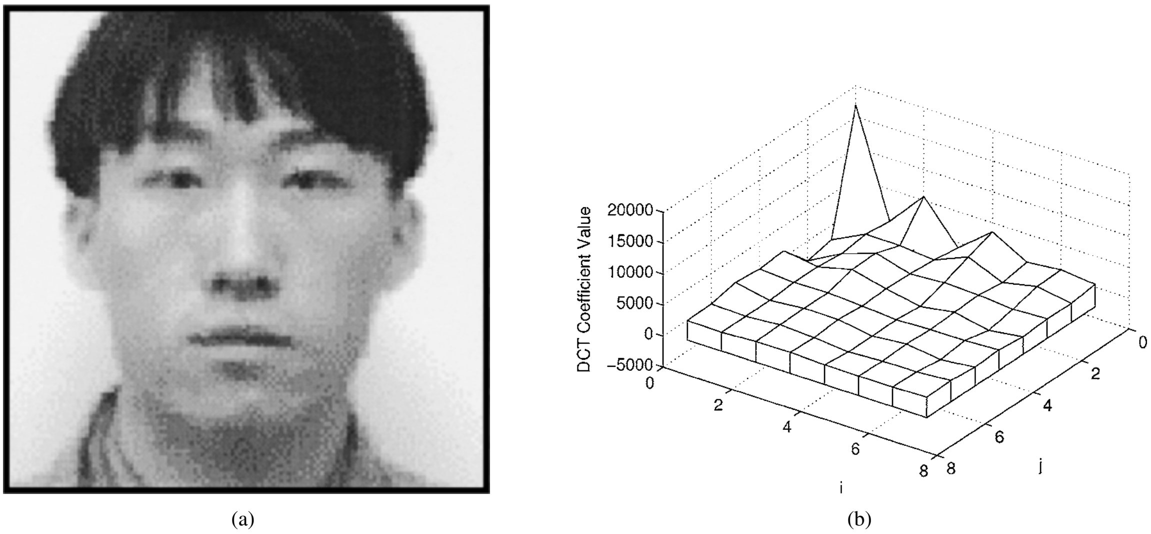

Figure 3. Typical face image (a) of size 128 × 128 and an 8 × 8 subset of its DCT (b).

encountered. This observation, of course, has no ramifications on the performance evaluation of the system, because accurate reconstruction is not a requirement. In fact, this situation was also encountered in Turk and Pentland (1991) where the KLT coefficients used in face recognition were not sufficient to achieve a subjectively acceptable facial reconstruction. Figure 4 shows the effect of using a feature vector of size 64 to reconstruct a typical face image. Now, it may be the case that one chooses to use more DCT coefficients to represent faces. However, there could be a cost associated with doing so. Specifically, more coefficients do not necessarily imply better recognition results, because by adding them, one may actually be representing more irrelevant information (Swets and Weng, 1996).

3.3. Normalization

Two kinds of normalization are performed in the proposed face recognition system. The first deals with geometric distortions due to varying imaging conditions.

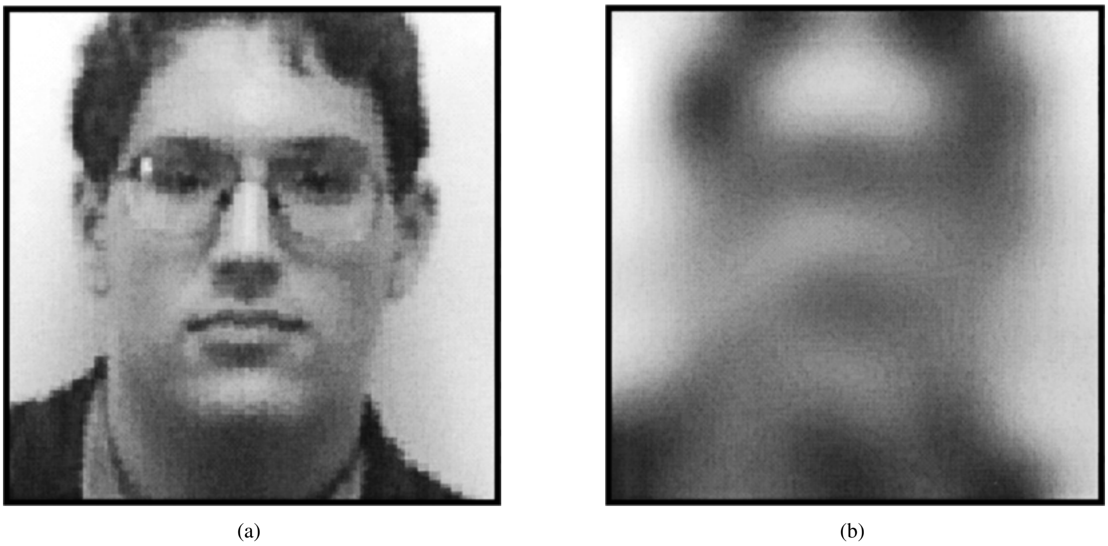

Figure 4. Effect of reconstructing a 128 × 128 image using only 64 DCT coefficients: (a) original (b) reconstructed.

That is, it attempts to compensate for position, scale, and minor orientation variations in faces. This way, feature vectors are always compared for images characterized by the same conditions. The second kind of normalization deals with the illumination of faces. The reasoning here is that the variations in pixel intensities between different images of faces could be due to illumination conditions. Normalization in this case is not very easily dealt with because illumination normalization could result in an artificial tinting of light colored faces and a corresponding lightening of dark colored ones. In the following two subsections, the issues involved in both kinds of normalization are presented, and the stage is set for various experiments to test their effectiveness for face recognition. These experiments and their results are detailed in Section 4.

3.3.1. Geometry. The proposed system is a holistic approach to face recognition. Thus it uses the image of a whole face and, as discussed in Section 1, it is expected to be sensitive to variations in facial scale and orientation. An investigation of this effect was performed in the case of the DCT to confirm this observation. The data used for this test were from the MIT database, which is described, along with the other databases studied, in a fair amount of detail in Section 4. This database contains a subset of faces that only vary in scale. To investigate the effects of scale on face recognition accuracy, faces at a single scale were used as the gallery faces, and faces from two different scales were used as the probes. Figure 5 illustrates how scale can degrade the performance of a face recognition system. In the figure, the term "Training Case" refers to the scale in the gallery images, and the terms "Case 1" and "Case 2" describe the two scales that were available for the probes. Figure 6 shows examples of faces from the training set and from the two cases of scale investigated. These results indicate that the DCT exhibits sensitivity to scale similar to that shown for the KLT (Turk and Pentland, 1991).

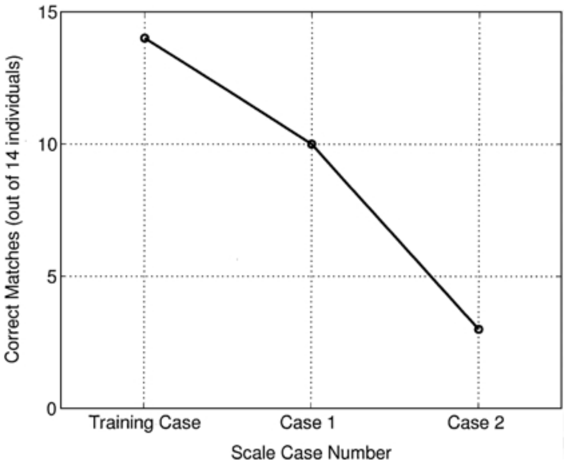

Figure 5. Effect of varying scale on recognition accuracy. 64 DCT coefficients were used for feature vectors, and 14 individuals of the MIT database were considered.

The geometric normalization we have used basically attempts to make all faces have the same size and same frontal, upright pose. It also attempts to crop face images such that most of the background is excluded. To achieve this, it uses the input face eye coordinates and defines a transformation to place these eyes in standard positions. That is, it scales faces such that the eyes are

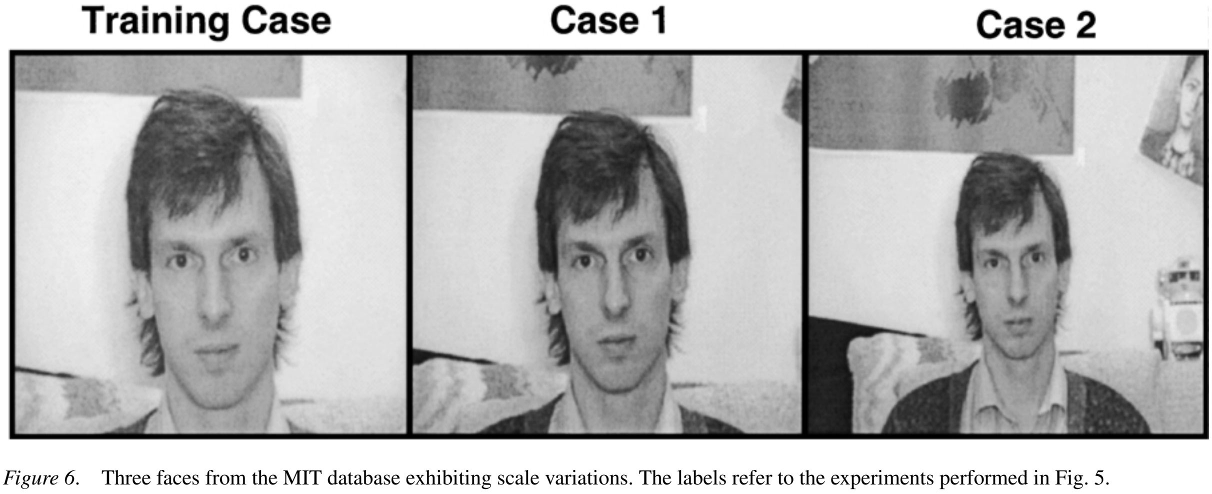

Figure 6. Three faces from the MIT database exhibiting scale variations. The labels refer to the experiments performed in Fig. 5.

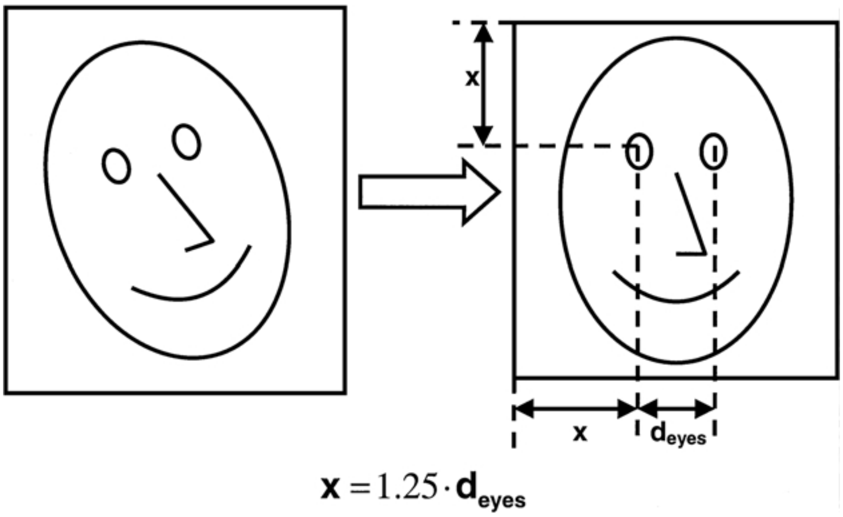

Figure 7. Geometric normalization and the parameters used. The final image dimensions are 128 × 128.

always the same distance apart, and it positions these faces in an image such that most of the background is excluded.

This normalization procedure is illustrated in Fig. 7, and it is similar to that proposed in Brunelli and Poggio (1993). Given the eye coordinates of the input face image, the normalization procedure performs the following three transformations: rotate the image so that the eyes fall on a horizontal line, scale the image (while maintaining the original aspect ratio) so that the eye centers are at a fixed distance apart (36 pixels), and translate the image to place the eyes at set positions within a 128×128 cropping window (see Fig. 7). Note that we only require the eye coordinates of input faces in order to perform this normalization. Thus no knowledge of individual face contours is available, which means that we cannot easily exclude the whole background from the normalized images. Since we cannot tailor an optimal normalization and cropping scheme for each face without knowledge of its contours, the dimensions shown in Fig. 7 were chosen to result in as little background, hair, and clothing information as possible, and they seemed appropriate given the variations in face geometry among people.

Another observation we can make about Fig. 7 is that the normalization performed accounts for only two-dimensional perturbations in orientation. That is, no compensation is done for three-dimensional (in depth) pose variations. This is a much more difficult problem to deal with, and a satisfactory solution to it has yet to be found. Of course, one could increase the robustness of a face recognition system to 3-D pose variations by including several training images containing such variations for a single person. The effect of doing this will be discussed in the next section. Also, by two-dimensional perturbations in orientation, we mean slight rotations from the upright position. These rotations are the ones that may arise naturally, even if people are looking straight ahead (see Fig. 8 for an example). Of course, larger 2-D rotations do not occur naturally and always include some 3-D aspect to them, which obviously 2-D normalization does not account for.

As for the actual normalization technique implemented, it basically consists of defining and applying a 2-D affine transformation, based on the relative eye positions and their distance. Figure 9 illustrates the result of applying such a transformation on a sample face image.

3.3.2. Illumination. Illumination variations play a significant role in degrading the performance of a face recognition system, even though Turk and Pentland indicate that the correlation between face images under different lighting conditions remains relatively high (Turk and Pentland, 1991). In fact, experience has shown that for large databases of images, obtained with different sensors under different lighting conditions, special care must be expended to ensure that recognition thresholds are not affected.

To compensate for illumination variations in our experiments, we apply Hummel's histogram modification technique (Hummel, 1975). That is, we simply choose

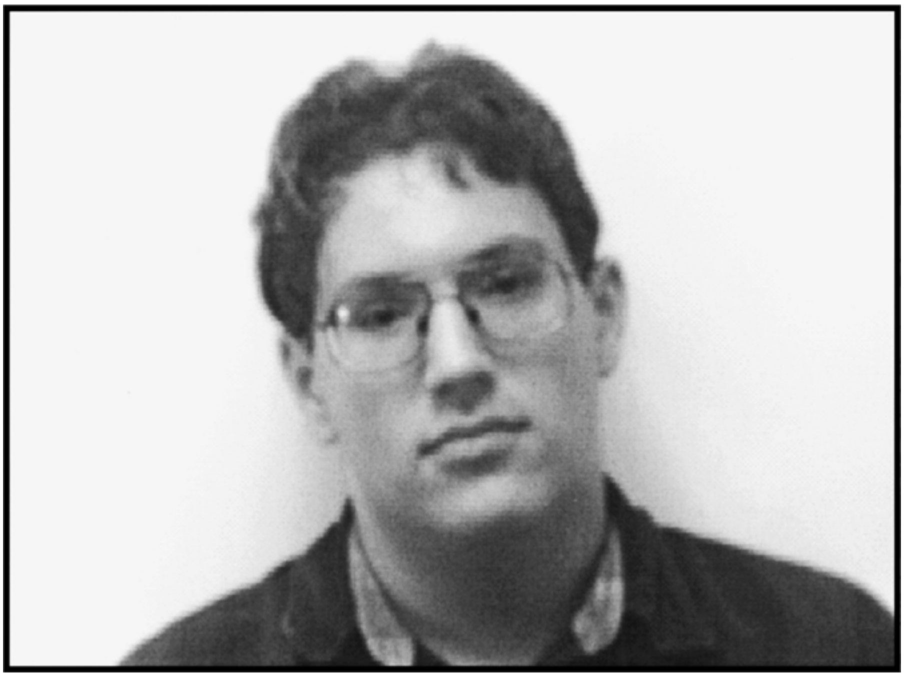

Figure 8. An example of naturally arising perturbations in face orientations.

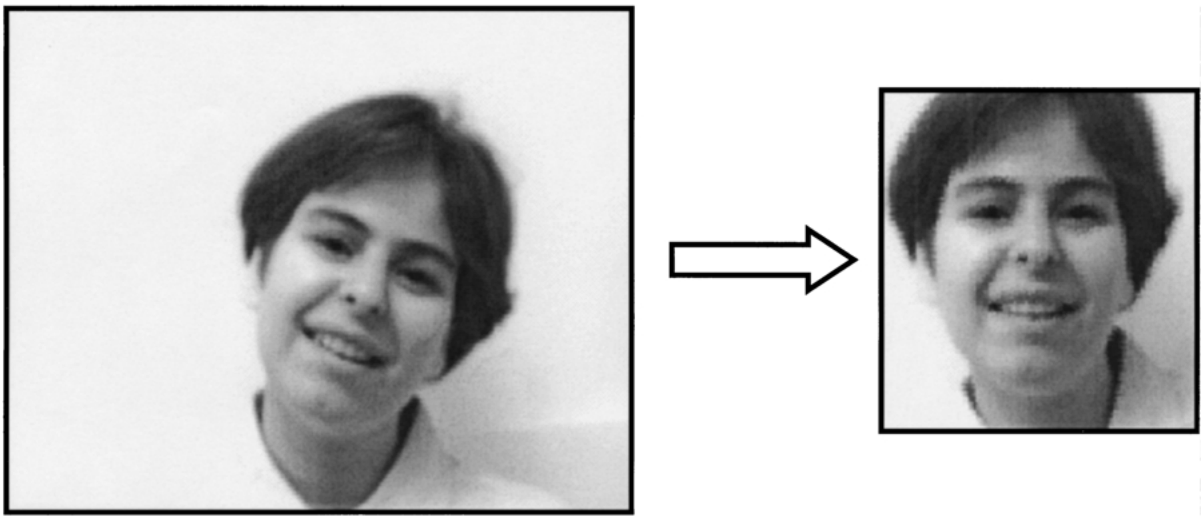

Figure 9. An illustration of the normalization performed on faces. Note the changes in scale, orientation, and position.

a target histogram and then compute a gray-scale transformation that would modify the input image histogram to resemble the target. It should be noted that another interesting approach to illumination compensation can be found in Brunelli (1997), in which computer graphics techniques are used to estimate and compensate for illuminant direction. This alleviates the need to train with multiple images under varying pose, but it also has significant computational costs.

The key issue in illumination compensation is how to select the target illumination. This is so because there could be tradeoffs involved in choosing such a target, especially if the face database contains a wide variety of skin tones. An extensive study of illumination compensation of faces for automatic recognition was done in conjunction with these experiments. The aim was to find an appropriate solution to this problem in order to improve the performance of our system. The results of this study are documented in an unpublished report available from the authors (Hafed, 1996).

The main conclusion that can be drawn from the study is that illumination normalization is very sensitive to the choice of target illumination. That is, if an average face is considered as a target, then all histograms will be mapped onto one histogram that has a reduced dynamic range (due to averaging), and the net result is a loss of contrast in the facial images. In turn, this loss of contrast makes all faces look somewhat similar, and some vital information about these faces, like skin color, is lost. It was found that the best compromise was achieved if the illumination of a single face is adjusted so as to compensate for possible non-uniform lighting conditions of the two halves of the same face. That is, no inter-face normalization is performed, and in this way, no artificial darkening or lightening of faces occurs due to attempts to normalize all faces to a single target. Of course, the results of illumination normalization really depend on the database being considered. For example, if the illumination of faces in a database is sufficiently uniform, then illumination normalization techniques are redundant.

4. Experiments

This section describes experiments with the developed face recognition system. These were fairly extensive, and the hallmark of the work presented here is that the DCT was put to the test under a wide variety of conditions. Specifically, several databases, with significant differences between them, were used in the experimentation.

A flowchart of the system described in the previous section is presented in Fig. 10. As can be seen, there is a pre-processing stage in which the face codes for the individual database images are extracted and stored for later use. This stage can be thought of as a modeling stage, which is necessary even for human beings: we perform a correlation between what is seen and what is already known in order to actually achieve recognition (Sekuler and Blake, 1994). At run-time, a test input is presented to the system, and its face codes are extracted. The closest match is found by performing a search that basically computes Euclidean distances and sorts the results using a fast algorithm (Silvester, 1993).

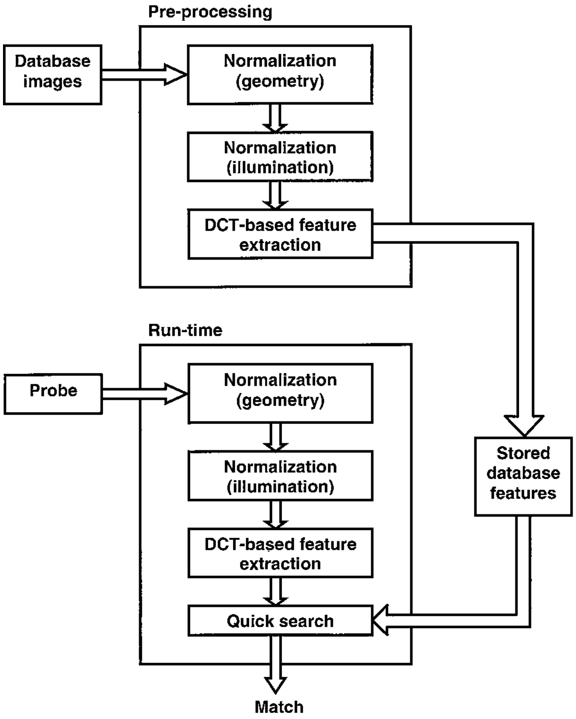

Figure 10. Implementation of face recognition system: the various modules used and the flowchart of operation.

This section begins with a brief overview of the various face databases used for testing the system; the differences among these databases are highlighted. Then the experiments performed and their results are presented and discussed.

4.1. Face Databases Considered

In order to establish the validity of the proposed face recognition algorithm empirically, it was tested on a variety of databases. As will be seen, there are significant differences among these databases, and this, in fact, was the motivation for considering all of them in evaluating our system. That is, the purpose was to show the consistency of the results for a range of databases that varied in the constraints imposed on the face images acquired.

4.1.1. The Achermann Database. The Achermann database was acquired at the University of Bern in Switzerland and contains 300 images of 30 individuals. For each individual in the database, a set of 10 images was taken with certain constrained 3-D pose variations.

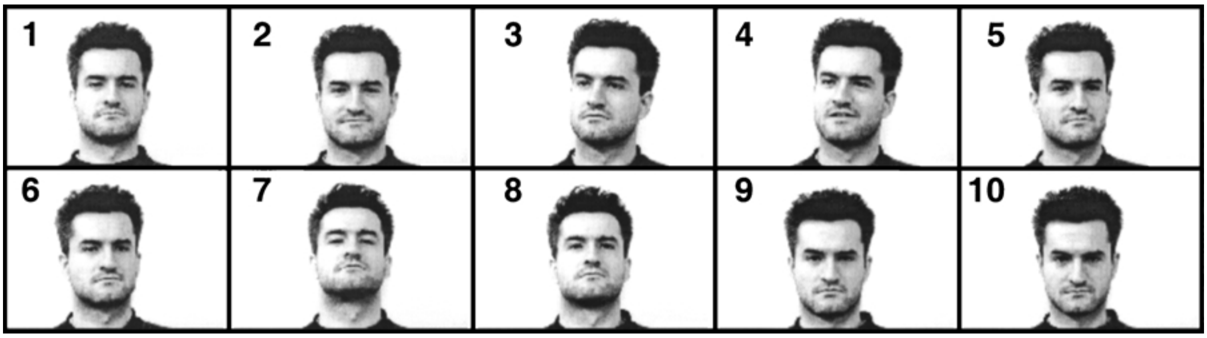

Figure 11. Views included in the Achermann face database.

Figure 11 shows these variations for a typical face in the database. Note that background and lighting conditions were uniform for all images. Also, note that this database permits the investigation of the sensitivity of the DCT to 3-D variations and the observation of the effects of increasing the number of training images per person on recognition accuracy. Finally, it should be mentioned that the database only contains males.

4.1.2. The Olivetti Database. The Olivetti database, as the name suggests, originated at the Olivetti Research Laboratory in England. It consists of 400 images of 40 individuals. Ten images were taken for each individual, and few constraints on facial expression and pose were imposed. Furthermore, some of the captured images were subject to illumination variations. Therefore, it is expected that this is a more difficult database to work with. However, the images do not include any backgrounds whatsoever. This database includes both males and females, and it can prove useful in investigating the effects of an increased number of training images per person. Figure 12 presents a sample set from this database.

4.1.3. The MIT Database. The MIT database used in this study consists of 432 images of 16 individuals. Twenty-seven images were obtained for each person in the database, and variations such as scale, orientation,

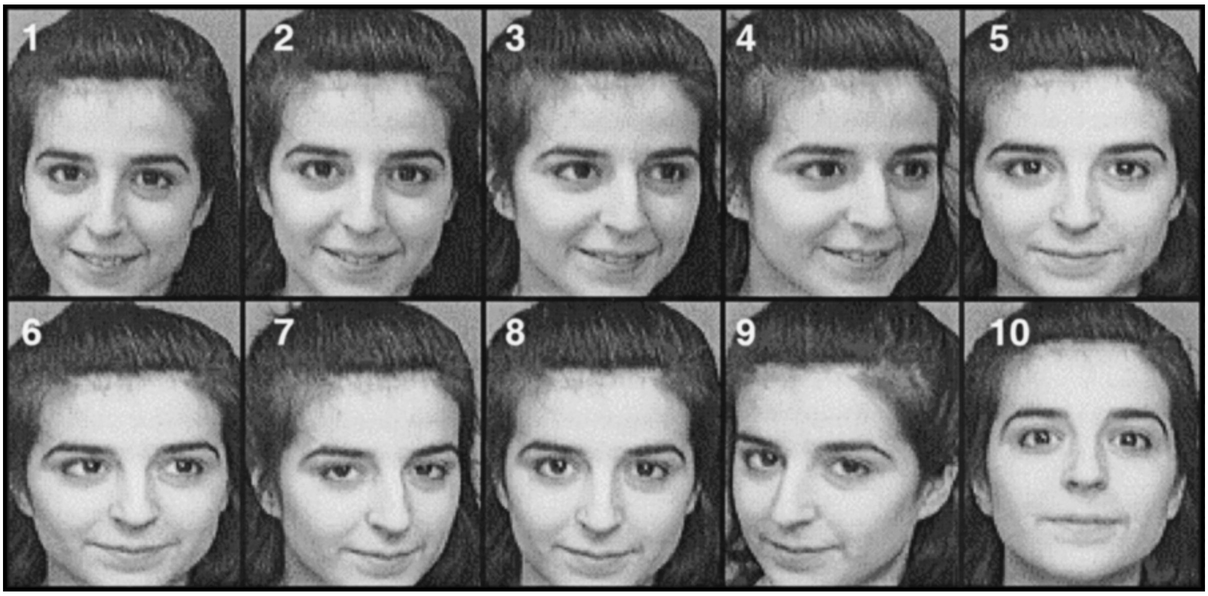

Figure 12. Views included in the Olivetti face database. Different people have different varieties of poses in this database.

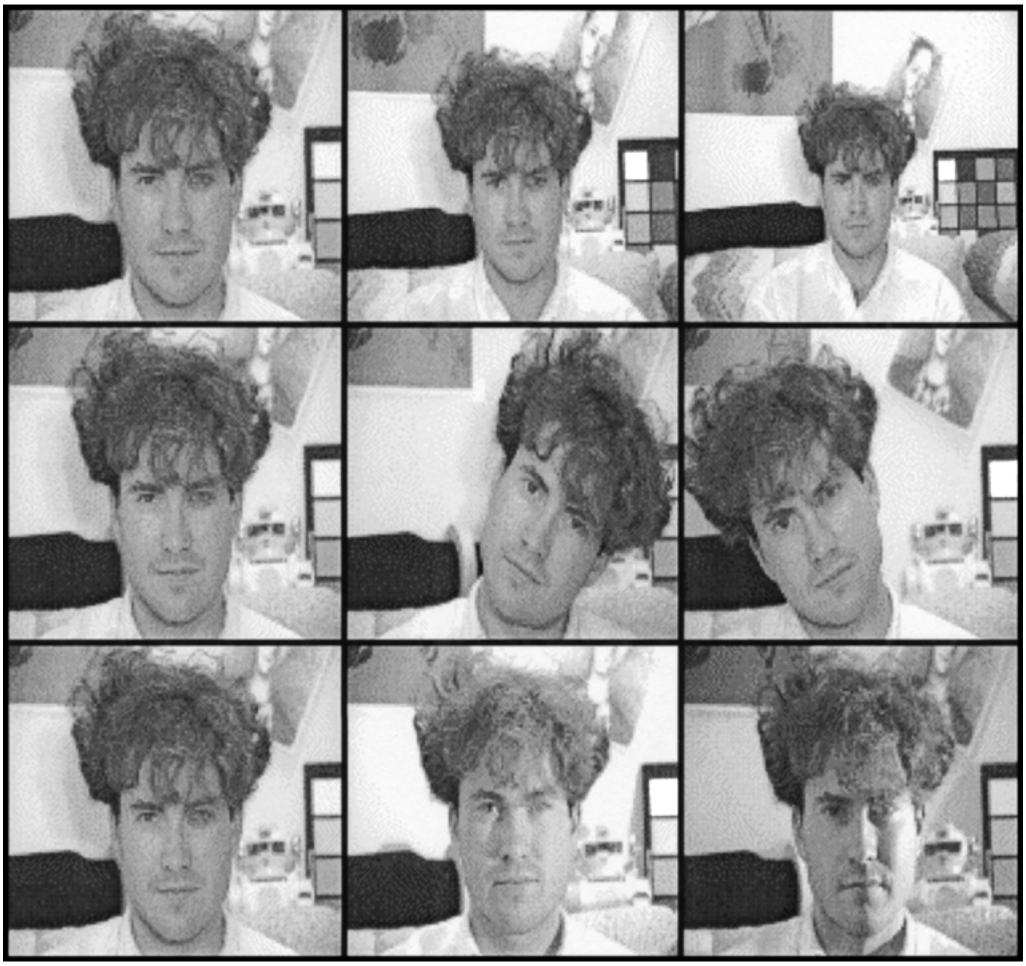

Figure 13. A subset of the various pictures of people in the MIT database. The top row shows variations in scale, the second row shows variations in orientation, and the third row shows variations in illumination. The remaining views in this database combine two or more of the variations shown here.

and lighting were included. Specifically, three cases of scale, three cases of orientation, and three cases of lighting conditions were considered. Then all combinations of these cases were taken. This database is useful for testing the efficacy of the normalization techniques described in the previous section. However, it is quite small and only includes males. Figure 13 shows sample faces from this database.

4.1.4. The CIM Database. The final database considered in this study was the CIM Face Database, which was obtained at the Center for Intelligent Machines (CIM) in McGill University. The database was collected for the purpose of this and other CIM projects during McGill University's recent 175th Anniversary Open House. It is a fairly large database and was designed to combine many of the features of the databases mentioned so far. Specifically, the database consists of 231 individuals for which 8 images per individual were taken. These 8 images covered variations in 2-D orientation, 3-D pose, and facial expression, as can be seen from Fig. 14. In fact, the CIM database combines the orientation variations of the MIT database with the 3-D pose changes of the Achermann database and the facial expression variations in the Olivetti database. It also includes people of various age, gender, and skin tone, and it thus poses a significant challenge to the DCT as well as to the normalization techniques used. An example of the variety encountered in the CIM Face Database is shown in Fig. 15. It should be noted that this database consists of approximately 70% males, 30% females, and 16% children.

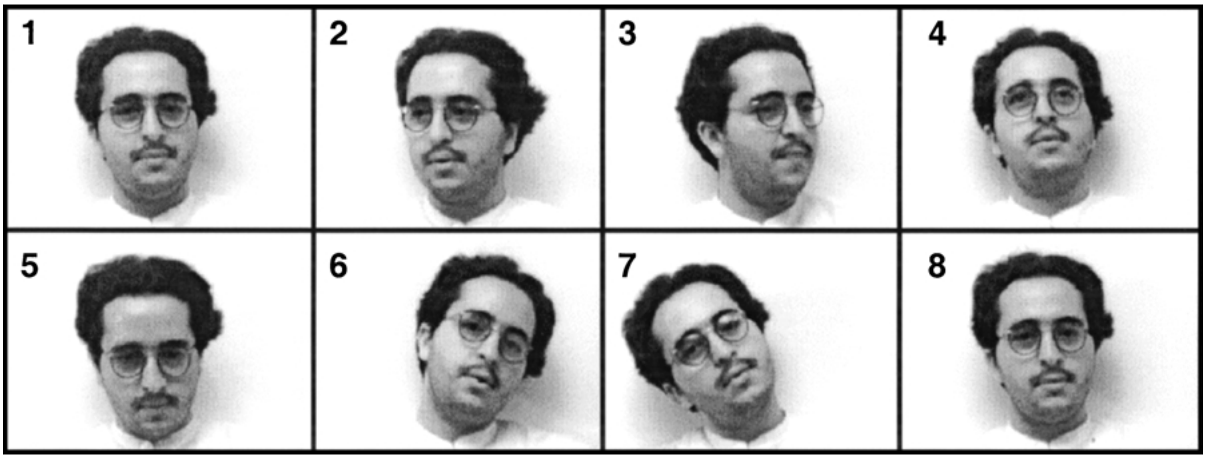

Figure 14. Views included in the CIM face database.

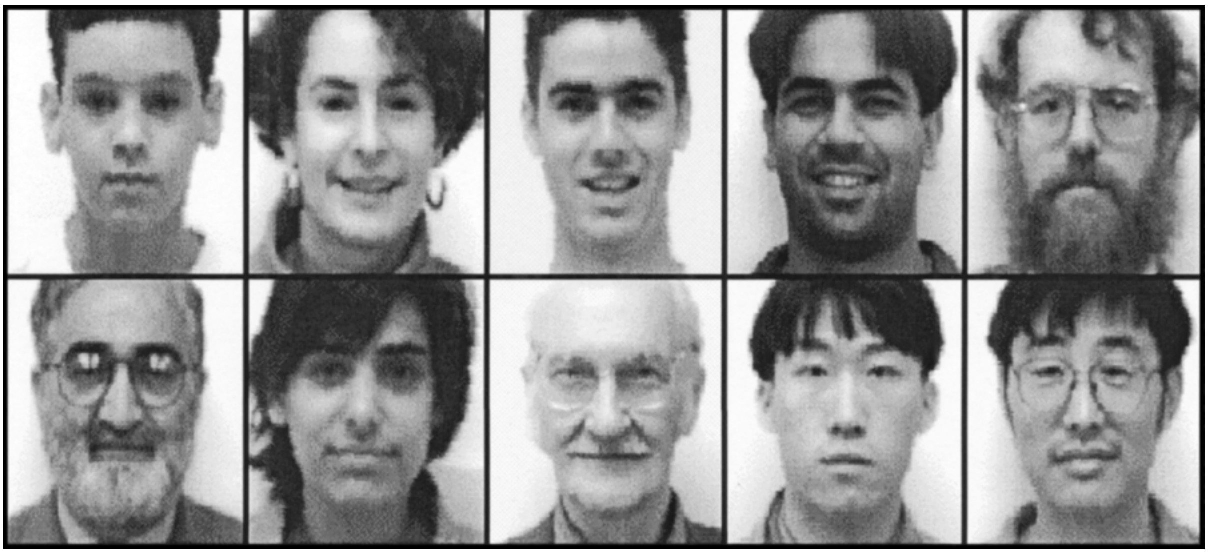

Figure 15. A sample set obtained from the CIM face database illustrating the variations of faces encountered in it. Facial expression variations also exist in the database.

4.2. Testing

In this section, various results are presented and discussed. We begin with the effects of the number of training images per person on recognition. Then, we consider varying the sizes of the feature vectors and observing the effects on recognition accuracy. Normalization is tested next, and finally, some general results are presented.

4.2.1. Number of Face Models Per Person. It is expected that the recognition accuracy of a face recognition system will improve as the number of face models per person increases. By face model we mean the image whose feature vector is stored in the system's database file or "memory." If a system has more than one face model for a particular person, then it "knows" more than one view of that person and thus can recognize him/her under more than one condition. Increasing the number of models per person is a simple and effective way of accounting for 3-D pose variations in faces. This solution may actually be likened to the situation in some biological vision systems where face selective cells have been shown to exist in the brain

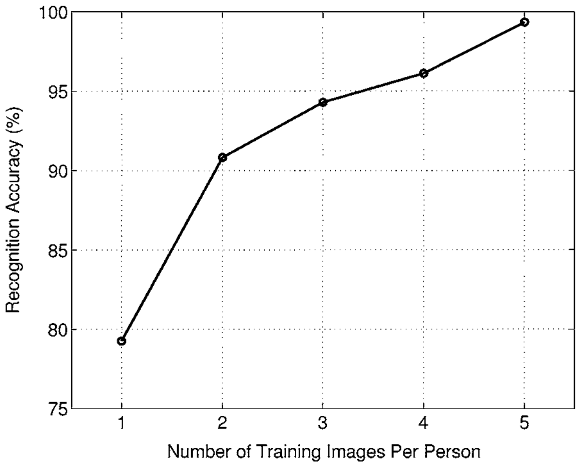

Figure 16. Effect of varying the number of training images/person on recognition accuracy for the Achermann database. 64 DCT coefficients were used for the feature vectors.

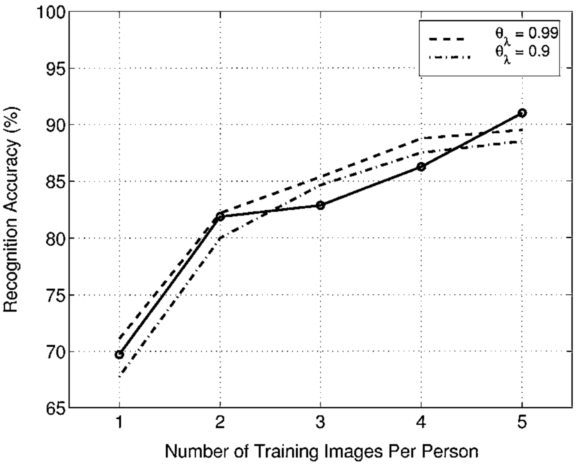

Figure 17. Effect of varying the number of training images/person on recognition accuracy for the Olivetti database. 49 DCT coefficients were used for the feature vectors. The two dashed curves are those obtained using the KLT with different values of θ_λ (Eq. (2.8)).

(Perrett et al., 1987). In fact, these face selective cells respond optimally to certain face poses.

Two experiments were performed to investigate the effects of varying the number of face models per person on recognition accuracy. These were performed *without* any normalization because none was necessary. In each of the databases considered, the faces were all of similar size and orientation, and they were obtained with the same background. The first experiment was performed on the Achermann database, and the results are presented in Fig. 16. Note that feature vectors of size 64 were used. Also note that the face models were chosen to be the odd numbered views in Fig. 11. So, for example, if two face models were used, then views 1 and 3 in Fig. 11 were selected and the remaining views were taken as the test inputs. That is, for this example, the system "memory" consisted of views 1 and 3 of all individuals in the database, and the probe set consisted of all remaining views. As can be seen, the expected increase in recognition accuracy is evident in Fig. 16.

Examining the figure, we note that the recognition accuracy is not very high when only one face model per person is used. This is expected because the single face model was a frontal image and the remaining test inputs exhibited large variations in 3-D pose. We also note that with 5 training images, the recognition rate becomes very high. Again this is expected because of the nature of the Achermann database. In other words, this is a relatively easy database to deal with because of the constraints put on the faces. All faces were of the same size, all rotations were controlled, and all facial expressions were standardized. In fact, looking at Fig. 11, we observe that the odd and even numbered views look almost identical.

Perhaps a more realistic data set to consider is the Olivetti database. A similar experiment was performed, and the results are highlighted in Fig. 17. In this experiment, 49 DCT coefficients were used for the feature vectors. Views 1 to 5 in Fig. 12 were used for the face models. That is, when 3 face models were included, for example, views 1 to 3 were considered to be training images and the remaining views were the test inputs.

As can be seen from Fig. 17, the same trend observed in Fig. 16 is present here. However, the overall performance is slightly inferior, as was predicted. In any case, it can be observed that very high recognition rates can be achieved with a small increase in the number of face models per person.

Figure 17 also indicates the performance of the KLT in the same experiment. The two curves shown are for different values of θ_λ (Eq. (2.8)). As can be seen, the performance of the DCT is comparable to that of the KLT. However, the larger feature vectors required to achieve the performance shown in the figure disadvantage the KLT. For example, when 4 training images are used and θ_λ is 0.99, the number of KLT coefficients in each feature vector is 138 (as opposed to 49 for the case of the DCT). To compare the performance of the DCT to other face recognition methods, the reader is referred to Table 3 in Eickeler et al. (2000). In this table, the authors report performance measures for various face recognition methods using the Olivetti database with five training images person. According to Fig. 17, the DCT achieves 91% accuracy under such conditions.

Note that, excluding the following sub-section, all other experiments in this paper were performed with only one face model per person.

4.2.2. Number of DCT Coefficients. In this section, we present the recognition accuracy of our system as a function of the number of DCT coefficients used. The experiment performed involved the Achermann database and basically consisted of a pair-wise matching of all faces in this database. That is, for each of the 300 images in the database, the closest match from the remaining 299 was found. This was repeated for various numbers of DCT coefficients, and the results are as shown in Fig. 18. It should be noted that this experiment utilizes the take-one-out method often used in pattern recognition applications to evaluate classifiers. We observe that the recognition accuracy becomes very high at certain points, where it actually exceeds 99%. This is true for two main reasons. First, pair-wise matching means that we are effectively using 9 face models per person and are thus accounting for all possible poses encountered in the test inputs. Second, the Achermann data were obtained under fairly strict conditions, and it was observed earlier that this database was relatively easy to deal with.

The reason Fig. 18 was generated was to show how the number of DCT coefficients used might have an effect on the performance of our system. We observe that there is a slight decrease in recognition accuracy as we go to higher numbers of coefficients. Also, note that only 64 DCT coefficients are enough to achieve good accuracy. This confirms the earlier discussion about whether accurate reconstruction of images is necessary for good performance, because in the case of the Achermann database, 64 coefficients are far from sufficient for an accurate reconstruction of the faces. Finally, experiments on other databases yielded curves similar to Fig. 18.

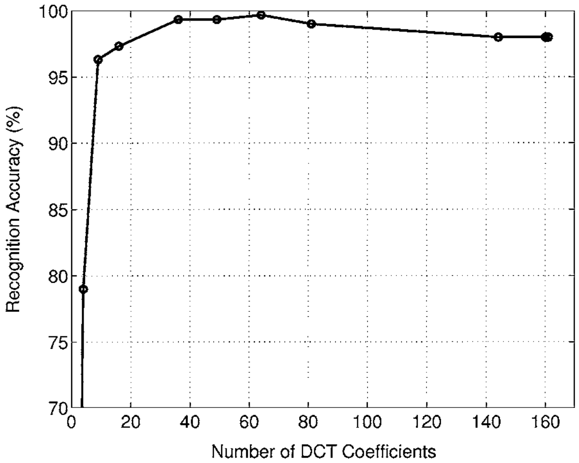

Figure 18. Effect of varying the number of DCT coefficients on recognition accuracy. The experiment involved pair-wise matching of the Achermann database faces.

4.2.3. Geometric Normalization. We now turn to the effects of the affine transformation discussed in Section 3. Namely, it was shown in Fig. 5 that scale variations could have detrimental effects on the performance of our face recognition system. This, in fact, is also true for orientation variations. In this section, we repeated the experiment described in Section 3.3.1 but with face normalization. The faces used are from the MIT database, normalized to yield images like the ones shown in Fig. 19. Figure 20 shows a tremendous improvement in the system's recognition rate.

The normalization technique proposed in Section 3.3.1 accounts for 2-D rotations of faces. However, one must be careful when dealing with such rotations, because for large angles of rotation, significant 3-D distortion becomes evident. That is, because of the nature of the human head and neck mechanism, we observe that large orientation changes also produce slight pose variations. Thus, a simple 2-D rotation will not fully re-orient a face to its frontal position. This is illustrated clearly in Fig. 21, where we can observe how the normalized image exhibits 3-D distortion. Of course, as was shown in Fig. 8, small-scale perturbations in orientation do arise naturally when people look straight ahead, and that is why the normalization technique used here is still necessary. We note that combining this normalization with multiple face models to account for 3-D distortions in pose would be effective for the large rotation angle exhibited in Fig. 21.

For the case of naturally arising perturbations in orientation, the normalization technique discussed above and in Section 3.3 was tested on 214 individuals in the CIM database. In this experiment, view 1 in Fig. 14 was used as the model for all people, and view 8 was used as the test view. That is, we compared frontal poses to other frontal poses, but we normalized to standardize scale, position, and the slight variations in orientation inherent in the database. We also used 49 DCT coefficients for the feature vectors. The recognition rate in

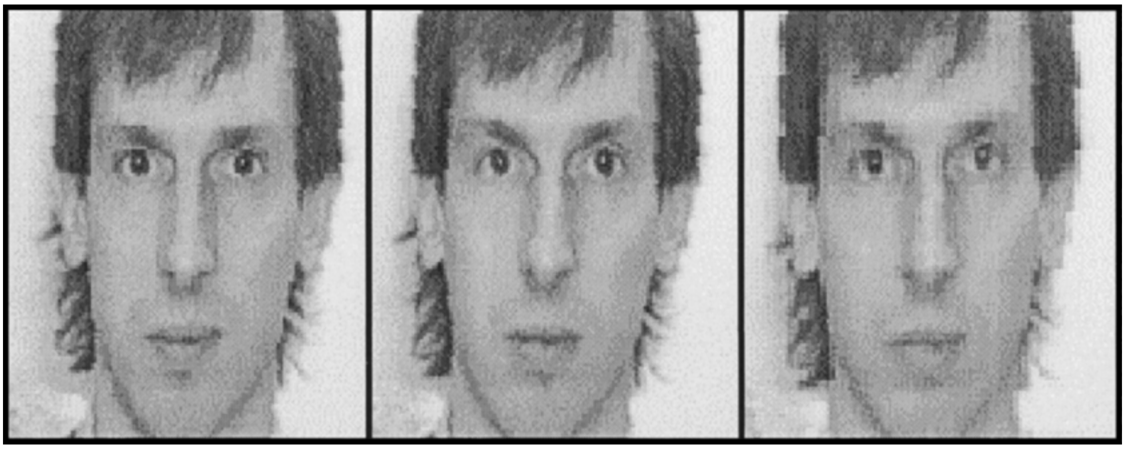

Figure 19. The faces of Fig. 6 after normalization. Note how normalization makes the images look almost identical.

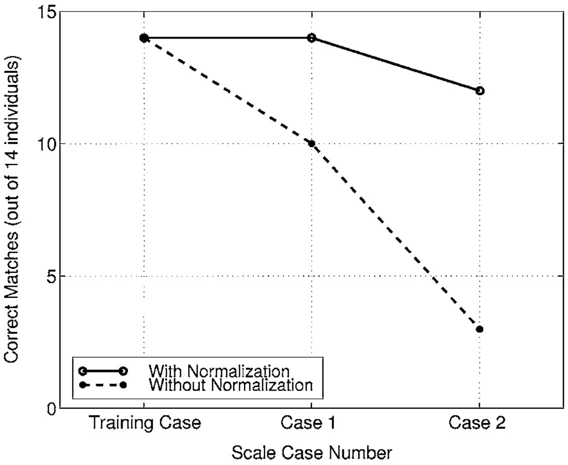

Figure 20. Effect of normalization on the recognition accuracy for the portion of the MIT database studied in Fig. 5.

this case was 84.58%. This experiment was also repeated using the KLT, and Table 1 summarizes our findings. As can be seen, the DCT outperforms the KLT when tested on the CIM database, and this may be attributed to the increased size of this database. That is, whereas the KLT performed fairly well and on par with the DCT for the Olivetti database, its performance did not scale well with database size. This observation was also made in our other experiments, as will be seen shortly. Also, note that the best KLT performance is achieved when 158 coefficients are used to represent each face. With this number of coefficients, the DCT was found to be computationally more efficient than the KLT, as is expected based on the analysis of Section 2.3.

Table 1. A performance comparison between the DCT and KLT on the CIM face database.

Method	Number of coefficients	Recognition accuracy
DCT	49	**84.58%**
	20	73.36%
KLT	46	77.10%
	158	**77.57%**

Several other comments can be made about the performance of the DCT on the CIM database. First, the rate obtained here is better than the rates obtained in Figs. 16 and 17, when only one face model was used. This is to be expected because, in the case considered here, the probes were all frontal. No 3-D poses were input to the system in this case. Also, note that 49 DCT coefficients were now used instead of the 64 that were found to be optimal for the Achermann database in Fig. 18. This number was obtained experimentally by varying the number of DCT coefficients and studying the recognition rates. Finally, as mentioned earlier, there were very slight restrictions on facial expressions in the CIM database, especially for view 8. So, in this respect, the CIM database is closer to the Olivetti database than it is to the Achermann or MIT databases. In fact, the CIM and Olivetti databases are slightly more difficult than both the Achermann and MIT ones.

4.2.4. Further Results. In this section, we present additional experiments performed on the CIM Face Database. These experiments were intended to further

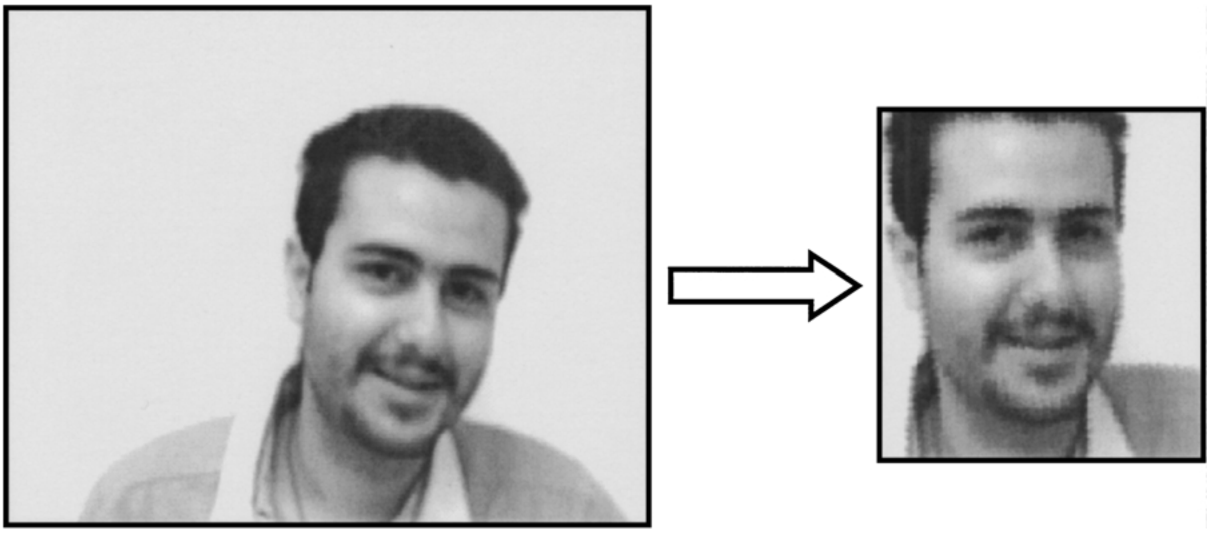

Figure 21. An example where 3-D pose distortions arise when large-scale head rotations (2-D) are attempted.

highlight the face recognition capabilities of the DCT, and the CIM database was chosen because of its size and variety. The results presented here show cumulative recognition accuracy as a function of rank for a variety of conditions. This format actually parallels that used in Phillips et al. (1996). The basic idea behind this format is to show that even if the closest match (rank 1) was not the correct match, the correct match almost always appears in the top, say, 50 matches (or ranks). That is, if a particular experiment results in a cumulative recognition accuracy of 90% at rank 20, then the correct match is among the closest 20 matches 90% of the time. Below, we also show how the ROC curve alluded to earlier in this paper can provide an estimate of the system performance in a verification scenario.

The first experiment performed was on 214 individuals in the CIM database, and involved frontal poses only. Geometric normalization was done to standardize scale, position, and orientation for all faces considered, and no illumination normalization was performed. This is because the faces in the CIM database are well illuminated, and experiments in Hafed (1996) suggested that illumination normalization for these faces was unnecessary. Finally, 49 DCT coefficients were used as feature vectors. Figure 22 shows the results of this experiment, as well as those obtained using the KLT. As can be observed, the results are as expected: there is an increase in the cumulative recognition accuracy with rank. We also notice the slightly inferior performance of the KLT when compared to the DCT. It should be noted that direct comparison of the results in Fig. 22 to those in Phillips et al. (1996) may not be very informative because they were not obtained on the same database.[3]

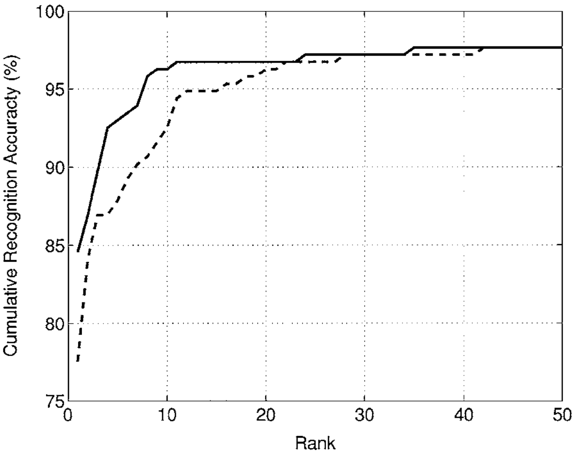

Figure 22. Cumulative recognition accuracy as a function of rank for the CIM face database. The dashed curve shows the cumulative recognition accuracy for the same database using the KLT.

The next experiment performed was the same as the one described above, but for adults only. The motivation for such an experiment was that some applications of face recognition, like automatic banking machines, for example, only involve adults, and an estimate of the performance of the DCT in such a case was desirable. The same experimental conditions as the ones above

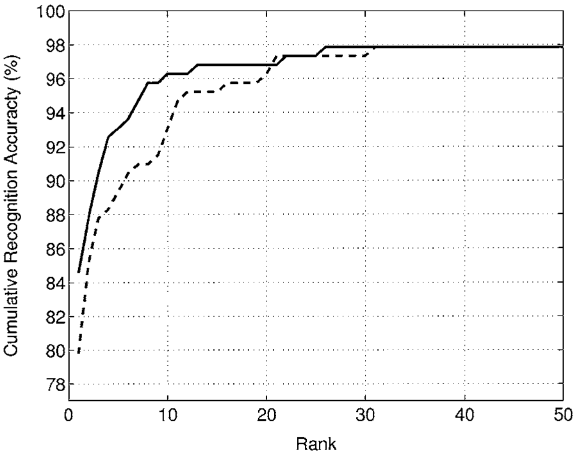

Figure 23. Cumulative recognition accuracy as a function of rank for the adults in the CIM face database. The dashed curve shows the cumulative recognition accuracy for the same database using the KLT.

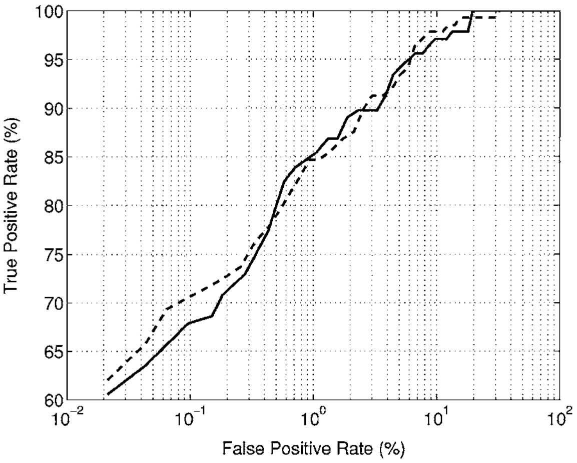

Figure 24. An ROC curve obtained for a subset of our CIM face database. The x-axis is shown in logarithmic scale to illustrate the tradeoffs involved in choosing a verification threshold. The dashed curve shows the ROC performance achieved using the KLT.

were replicated here, and Fig. 23 shows the results. As can be observed, the same trend is noticed with a slightly better performance at higher ranks. This is expected because the data set here is smaller. The DCT again outperforms the KLT in this experiment.

As mentioned earlier, adding a threshold for the distance measure between features permits rejection of unknown faces and verification of those that are known. In other words, in a face verification scenario, we are given an unknown face and a 'claimed' identity for that face. If the distance between this face's features and those of the database image, against which it is being verified is less than some threshold, the claim is accepted; otherwise, it is rejected. Obviously, an ideal threshold would be one that gives rise to 100% true positives (faces correctly accepted as known) and 0% false positives (faces incorrectly accepted as known). However, in practice, a tradeoff is bound to arise when choosing a threshold. That is, if the distribution of same-person distances overlaps with that of different-person distances, then non-zero false positive rates will necessarily arise. The problem then is to choose the optimum threshold that would meet a particular system's performance criteria. This is where ROC analysis is extremely helpful. As an example, Fig. 24 shows the ROC curve obtained for a subset of the CIM database. Note that the threshold value is implicit in this kind of analysis. However, it should be evident that as the threshold is increased, we travel upwards on the curve, towards increasing true positive and false positive rates. We observe the tradeoff between correct verification versus false acceptances. One can also observe the diminishing returns of threshold increases. In fact, beyond a certain point, threshold increases simply increase the number of errors in verification, without improving the desired correct performance (which in this case saturates at 100%). The dashed curve in the figure was obtained using the KLT. As can be seen, the performance of the DCT is very similar to the KLT in a verification scenario on the CIM database.

5. Conclusions and Comments

An alternative holistic approach to face recognition was investigated and tested. The approach was based on the discrete cosine transform, and experimental evidence to confirm its usefulness and robustness was presented. The mathematical relationship between the discrete cosine transform (DCT) and the Karhunen-Loeve transform (KLT) explains the near-optimal performance of the former. This mathematical relationship justifies the use of the DCT for face recognition, in particular, because Turk and Pentland have already shown earlier that the KLT performs well in this application (Turk and Pentland, 1991). Experimentally, the DCT was shown to perform very well in face recognition, just like the KLT.

Face normalization techniques were also incorporated in the face recognition system discussed here. Namely, an affine transformation was used to correct for scale, position, and orientation changes in faces.

It was seen that tremendous improvements in recognition rates could be achieved with such normalization. Illumination normalization was also investigated extensively. Various approaches to the problem of compensating for illumination variations among faces were designed and tested, and it was concluded that the recognition rate of our system was sensitive to many of these approaches. This was partly because the faces in the databases used for the tests were uniformly illuminated and partly because these databases contained a wide variety of skin tones. That is, certain illumination normalization techniques had a tendency to make all faces have the same overall gray-scale intensity, and they thus resulted in the loss of much of the information about the individuals' skin tones.

A complexity comparison between the DCT and the KLT is of interest. In the proposed method, training essentially means computing the DCT coefficients of all the database faces. On the other hand, using the KLT, training entails computing the basis vectors of the transformation. This means that the KLT is more computationally expensive with respect to training. However, once the KLT basis vectors have been obtained, it may be argued that computing the KLT coefficients for recognition is trivial. But this is also true of the DCT, with the additional proviso that the DCT may take advantage of very efficient computational algorithms (Rao and Yip, 1990). For example, with 158 basis vectors (which is the number that provided the best performance for the CIM database) and 128 × 128 images, the KLT computation required around five times more computational time than the DCT computation on a 200MHz personal computer.

As for the issue of multiple face models per person, it has been argued that this might be a simple way to deal with 3D facial distortions. In this regard, the KLT method is also not distortion-invariant, so it would suffer from similar speed degradation if it were to deal with face distortions in this manner. On the other hand, a method like that described in Wiskott and von der Malsburg (1995) is said to be distortion-invariant. This method performs relatively well, but being based on the Dynamic Link Architecture (DLA), it is not very efficient. Specifically, in this method, matching is dependent on synaptic plasticity in a self-organizing neural network. Thus, to recognize a face, a system based on this method has to first match this face to all models (through this process of map self-organization) and then choose the model that minimizes some cost function. Clearly, simulating the dynamics of a neural network for each model face in a database in order to recognize an input image is computationally expensive. Therefore, it seems that there remains a strong tradeoff between performance and complexity in many existing face recognition algorithms.

This paper has discussed a face recognition system using the DCT, which included both geometrical and illumination normalization techniques. Naturally, improvements to the proposed system can be envisioned. For example, the system lacks face localization capabilities. It would be desirable to add one of the many reported methods in the literature so that the system could be completely independent of the manual input of the eye coordinates. In fact, the DCT could be used to perform this localization. That is, frequency domain information obtained from the DCT could be used to implement template-matching algorithms for finding faces or eyes in images. Geometric normalization could also be generalized to account for 3-D pose variations in faces. As for illumination compensation, we have observed that light-colored faces were artificially tinted and darker colored faces brightened due to the choice of target face illumination used when applying histogram modification. Thus, being able to categorize individuals in terms of, perhaps, skin color could be used to define different target illuminations, independently tuned to suit various subsets of the population. For example, an average of Caucasian faces would not be very well suited to modify the illumination of black faces, and vice versa. This classification approach would have the advantage of reducing the sensitivity of the system to illumination normalization.

Finally, we can contemplate other enhancements similar to those attempted for the KLT method. For example, the DCT could be used as a first stage transformation followed by linear discriminant analysis, as in Belhumeur et al. (1997). Also, the DCT could be computed for local facial features in addition to the global computation proposed here. This, while moderately enlarging the size of the feature vectors, would most likely yield better performance.

Notes

1. This will be discussed in a later section.
2. See Sections 4 and 5 for more on this point.
3. The much-discussed FERET database is absolutely not available to researchers outside the USA. It is rather unfortunate that this database is being used as a de facto standard by some and that numerous papers are being published in the literature based on

experiments with it. Even the reviewers of this paper asked for comparisons with FERET and questioned the conclusions in this section. We were unable to perform such comparisons because of the restrictions on the usage of FERET. The US government, however, should be very pleased; none of the researchers we contacted was willing to let us have access to the FERET data since all had signed nondisclosure agreements!

References

Ahmed, N., Natarajan, T., and Rao, K. 1974. Discrete cosine transform. *IEEE Trans. on Computers*, 23(1):90–93.

Akamatsu, S., Sasaki, T., Fukamachi, H., and Suenaga, Y. 1991. A robust face identification scheme—KL expansion of an invariant feature space. In *SPIE Proc.: Intell. Robots and Computer Vision X. Algorithms and Techn.*, 1607:71–84.

Belhumeur, P.N., Hespanha, J.P., and Kriegman, D.J. 1997. Eigenfaces vs. fisherfaces: Recognition using class specific linear projection. *IEEE Trans. on Patt. Anal. and Mach. Intel.*, 19(7):711–720.

Bruce, V. 1990. Perceiving and recognizing faces. *Mind and Language*, 5:342–364.

Brunelli, R. and Poggio, T. 1993. Face recognition: Features versus templates. *IEEE Trans. on Patt. Anal. and Mach. Intel.*, 15(10):1042–1052.

Brunelli, R. 1997. Estimation of pose and illuminant direction for face processing. *Image and Vision Computing*, 15(10):741–748.

Burt, P. 1989. Multiresolution techniques for image representation, analysis, and 'smart' transmission. In *SPIE Proc.: Visual Comm. and Image Proc. IV.*, 1199:2–15.

Chellappa, R., Wilson, C., and Sirohey, S. 1995. Human and machine recognition of faces: A survey. In *Proc. IEEE*, 83(5):705–740.

Craw, I., Tock, D., and Bennett, A. 1992. Finding face features. In *Proc. 2nd Europe. Conf. on Computer Vision*, pp. 92–96.

Duda, R.O. and Hart, P.E. 1973. *Pattern Classification and Scene Analysis*. Wiley: New York, NY.

Eickeler, S., Müller, S., and Rigoll, G. 2000. Recognition of JPEG compressed face images based on statistical methods. *Image and Vision Computing*, 18(4):279–287.

Govindaraju, V., Srihari, S., and Sher, D. 1990. A Computational model for face location. In *Proc. 3rd Int. Conf. on Computer Vision*, pp. 718–721.

Grzybowski, M. and Younger, J. 1997. Statistical methodology: III. Receiver operating characteristic (ROC) curves. *Academic Emergency Medicine*, 4(8):818–826.

Hafed, Z. 1996. Illumination compensation of facial images for automatic recognition. Internal report for McGill University course 304–529A.

Harmon, L.D. and Hunt, W.F. 1977. Automatic recognition of human face profiles. *Computer Graphics and Image Proc.*, 6:135–156.

Hummel, R. 1975. Histogram modification techniques. *Computer Graphics and Image Proc.*, 4:209–224.

Jain, A. 1989. *Fundamentals of Digital Image Processing*. Prentice: Englewood Cliffs, NJ.

Jebara, T. 1996. 3D Pose estimation and normalization for face recognition. Honours Thesis, McGill University, Montreal, Quebec, Canada.

Kanade, T. 1973. Picture processing system by computer complex and recognition of human faces. Ph.D. Thesis, Department of Information Science, Kyoto University, Japan.

Kelly, M. 1970. Visual identification of people by computer. Stanford AI Proj., Stanford, CA, Tech. Rep. AI-130.

Kirby, M. and Sirvoich, L. 1990. Application of the Karhunen-Loeve procedure for the characterization of human faces. *IEEE Trans. on Patt. Anal. and Machine Intell.*, 12:103–108.

Lades, M., Vorbruggen, J., Buhmann, J., Lange, J., von der Malsburg, C., and Wurtz, R. 1993. Distortion invariant object recognition in the dynamic link architecture. *IEEE Trans. on Computers*, 42:300–311.

Oppenheim, A. and Schafer, R. 1989. *Discrete-Time Signal Processing*. Prentice: Englewood Cliffs, NJ.

Pennebaker, W. and Mitchell, J. 1993. *JPEG: Still Image Data Compression Standard*. VNR: New York, NY.

Pentland, A., Moghaddam, B., Starner, T., and Turk, M. 1994. View-based and modular eigenspaces for recognition. In *Proc. IEEE Computer Soc. Conf. On Computer Vision and Patt. Recog.*, pp. 84–91.

Perrett, D., Mistlin, A., and Chitty, A. 1987. Visual neurons responsive to faces. *Trends in Neuroscience*, 10:358–363.

Phillips, P., Rauss, P., and Der, S. 1996. FERET (Face recognition technology) recognition algorithm development and test results. Army Research Laboratory, Adelphi, MD, Tech. Rep. ARL-TR-995.

Pratt, W. 1991. *Digital Image Processing*. 2nd edition. Wiley: New York, NY.

Rao, K. and Yip, P. 1990. *Discrete Cosine Transform—Algorithms, Advantages, Applications*. Academic: New York, NY.

Rosenfeld, A. and Kak, A. 1976. *Digital Picture Processing*. Academic: New York, NY.

Sakai, T., Nagao, M., and Fujibayashi, S. 1969. Line extraction and pattern recognition in a photograph. *Patt. Recog.*, 1:233–248.

Sekuler, R. and Blake, R. 1994. *Perception*. 3rd edition. McGraw-Hill: New York, NY.

Shepherd, J. 1985. An interactive computer system for retrieving faces. In *Aspects of Face Processing*, H.D. Ellis, M.A. Jeeves, F. Newcombe, and A. Young (Eds.). Nijhoff: Dordrecht, pp. 398–409.

Shneier, M. and Abdel-Mottaleb, M. 1996. Exploiting the JPEG compression scheme for image retrieval. *IEEE Trans. on Patt. Anal. and Mach. Intell.*, 18(8):849–853.

Silvester, P. 1993. *Data Structures for Engineering Software*. Computational Mechanics: Southampton.

Swets, D. and Weng, J. 1996. Using discriminant eigenfeatures for image retrieval. *IEEE Trans. on Patt. Anal. and Mach. Intell.*, 18(8):831–836.

Turk, M. and Pentland, A. 1991. Face recognition using eigenfaces. In *Proc. Int. Conf. on Comp. Vision and Patt. Recog. (CVPR'91)*, pp. 586–591.

Wang, Z. 1984. Fast algorithms for the discrete W transform and for the discrete fourier transform. *IEEE Trans. Acoust., Speech, and Signal Proc.*, 32:803–816.

Wiskott, L. and von der Malsburg, C. 1995. Recognizing face by dynamic link matching. In *Proc. ICANN'95*, pp. 347–352.

Yuille, A., Cohen, D., and Hallinan, P. 1989. Feature extraction from faces using deformable templates. In *Proc. IEEE Computer Soc. Conf. on Computer Vision and Patt. Recog.*, pp. 104–109.