Discriminant analysis for recognition of human face images

Kamran Etemad and Rama Chellappa

Department of Electrical Engineering and Center for Automation Research, University of Maryland, College Park, Maryland 20742

Received January 29, 1996; revised manuscript received October 25, 1996; accepted February 14, 1997

The discrimination power of various human facial features is studied and a new scheme for automatic face recognition (AFR) is proposed. The first part of the paper focuses on the linear discriminant analysis (LDA) of different aspects of human faces in the spatial as well as in the wavelet domain. This analysis allows objective evaluation of the significance of visual information in different parts (features) of the face for identifying the human subject. The LDA of faces also provides us with a small set of features that carry the most relevant information for classification purposes. The features are obtained through eigenvector analysis of scatter matrices with the objective of maximizing between-class variations and minimizing within-class variations. The result is an efficient projection-based feature-extraction and classification scheme for AFR. Each projection creates a decision axis with a certain level of discrimination power or reliability. Soft decisions made based on each of the projections are combined, and probabilistic or evidential approaches to multisource data analysis are used to provide more reliable recognition results. For a medium-sized database of human faces, excellent classification accuracy is achieved with the use of very-low-dimensional feature vectors. Moreover, the method used is general and is applicable to many other image-recognition tasks. [S0740-3232(97)01008-9]

Key words: Face recognition, wavelet packets, discriminant eigenpictures, evidential reasoning, multisource soft decision integration.

1. INTRODUCTION

Inspired by the human's ability to recognize faces as special objects and motivated by the increased interest in the commercial applications of automatic face recognition (AFR) as well as by the emergence of real-time processors, research on automatic recognition of faces has become very active. Studies on the analysis of human facial images have been conducted in various disciplines. These studies range from psychophysical analysis of human recognition of faces and related psychovisual tests[1,2] to research on practical and engineering aspects of computer recognition and verification of human faces and facial expressions[3] and race and gender classification.[4,5]

The problem of AFR alone is a composite task that involves detection and location of faces in a cluttered background, facial feature extraction, and subject identification and verification.[6,7] Depending on the nature of the application, e.g., image acquisition conditions, size of database, clutter and variability of the background and foreground, noise, occlusion, and finally cost and speed requirements, some of the subtasks are more challenging than others.

The detection of a face or a group of faces in a single image or a sequence of images, which has applications in face recognition as well as video conferencing systems, is a challenging task and has been studied by many researchers.[7–12] Once the face image is extracted from the scene, its gray level and size are usually normalized before storing or testing. In some applications, such as identification of passport pictures or drivers' licenses, conditions of image acquisition are usually so controlled that some of the preprocessing stages may not be necessary.

One of the most important components of an AFR system is the extraction of facial features, in which attempts are made to find the most appropriate representation of face images for identification purposes. The main challenge in feature extraction is to represent the input data in a low-dimensional feature space, in which points corresponding to different poses of the same subject are close to each other and far from points corresponding to instances of other subjects' faces. However, there is a lot of within-class variation that is due to differing facial expressions, head orientations, lighting conditions, etc., which makes the task more complex.

Closely tied to the task of feature extraction is the intelligent and sensible definition of similarity between test and known patterns. The task of finding a relevant distance measure in the selected feature space, and thereby effectively utilizing the embedded information to identify human subjects accurately, is one of the main challenges in face identification. In this paper we focus on feature-extraction and face-identification processes.

Typically, each face is represented by use of a set of gray-scale images or templates, a small-dimensional feature vector, or a graph. There are also various proposals for recognition schemes based on face profiles[13] and isodensity or depth maps.[14,15] There are two major approaches to facial feature extraction for recognition in computer vision research: holistic template matching-based systems and geometrical local feature-based schemes and their variations.[7]

In holistic template-matching systems each template is a prototype face, a facelike gray-scale image, or an abstract reduced-dimensional feature vector that has been obtained through processing the face image as a whole. Low-dimensional representations are highly desirable for large databases, fast adaptation, and good generalization. On the basis of these needs, studies have been performed on the minimum acceptable image size and the smallest number of gray levels required for good recognition results.[6] Reduction in dimensionality can also be achieved by using various data-compression schemes. For example, representations based on principal-component analysis[16–20] (PCA) and singular-value decomposition have been studied and used extensively for various applications. It has also been shown that the nonlinear mapping capability of multilayer neural networks can be utilized and that the internal and hidden representations of face patterns, which typically are of much lower dimensionality than the original image, can be used for race and gender classification.[4,5] Some of the most successful AFR schemes are based on the Karhunen–Loève transform[17,19] (KLT), which yield the so-called eigenfaces. In these methods the set of all face images is considered a vector space, and the eigenfaces are simply the dominant principal components of this face space; they are computed as eigenvectors of the covariance matrix of data.

In geometrical feature-based systems, one attempts to locate major face components or feature points in the image.[21–24] The relative sizes of and distances between the major face components are then computed. The set of all normalized size and distance measurements constitute the final feature vectors for classification. One can also use the information contained in the feature points to form a geometrical graph representation of the face that directly shows sizes and relative locations of major face attributes.[22] Most geometrical feature-based systems involve several steps of window-based local processing, followed by some iterative search algorithms, to locate the feature points. These methods are more adaptable to large variations in scale, size, and location of the face in an image but are more susceptible to errors when face details are occluded by objects, e.g., glasses, or by facial hair, facial expressions, or variations in head orientation. Compared with template and PCA-based systems, these methods are computationally more expensive. Comparative studies of template versus local feature-based systems can be found in Refs. 4, 7, and 25. There are also various hybrid schemes that apply the KLT or the template matching idea to face components and use correlation-based searching to locate and identify facial feature points.[4,19] The advantage of performing component-by-component matching is improved robustness against head orientation changes, but its disadvantage is the complexity of searching for and locating face components.

The human audiovisual system, as a powerful recognition model, takes great advantage of context and auxiliary information. Inspired by this observation, one can devise schemes that can consistently incorporate context and collateral information, when and if they become available, to enhance its final decisions. Incorporating information such as race, age, and gender, obtained through independent analysis, improves recognition results.[19] Also, since face recognition involves a classification problem with large within-class variations, caused by dramatic image variation in different poses of the subject, one has to devise methods of reducing or compensating for such variability. For example,

1. For each subject, store several templates, one for each major distinct facial expression and head orientation. Such systems are typically referred to as view-based systems.
2. Use deformable templates along with a three-dimensional model of a human face to synthesize virtual poses and apply the template matching algorithm to the synthesized representations.[26]
3. Incorporate such variations into the process of feature extraction.

In this paper we take the third approach and keep the first method as an optional stage that can be employed depending on the complexity of the specific task. Our approach is a holistic linear discriminant analysis (LDA)-based feature extraction for human faces followed by an evidential soft-decision integration for multisource data analysis. This method is a projection-based scheme of low complexity that avoids any iterative search or computation. In this method both off-line feature extraction and on-line feature computation can be done at high speeds, and recognition can be done in almost real time. Our experimental results show that high levels of recognition performance can be achieved with low complexity and a small number of features.

The organization of this paper is as follows. In Section 2 we provide an objective study of multiscale features of face images in terms of their discrimination power. In Section 3 we propose a holistic method of projection-based discriminant facial feature extraction through LDA of face images. We also make a comparative study of the features obtained with the proposed scheme and the features used in compression-based methods such as PCA. In Section 4 we address the task of classification and matching through multisource data analysis and combining soft decisions from multiple imprecise information sources. Finally, we propose a task-dependent measure of similarity in the feature space that is based on the reliability of the basic decisions, to be used at the identification stage.

2. LINEAR DISCRIMINANT ANALYSIS OF FACIAL IMAGES

As highly structured two-dimensional patterns, human face images can be analyzed in the spatial and the frequency domains. These patterns are composed of components that are easily recognized at high levels but are loosely defined at low levels of our visual system.[2,27] Each of the facial components (features) has a different discrimination power for identifying a person or the person's gender, race, and age. There have been many studies of the significance of such features that used subjective psychovisual experiments.[1,2]

Using objective measures, in this section we propose a computational scheme for evaluating the significance of different facial attributes in terms of their discrimination potential. The results of this analysis can be supported by subjective psychovisual findings. To analyze any representation V, where V can be the original image, its spatial segments, or transformed images, we provide the following framework.

First, we need a training set composed of a relatively large group of subjects with diverse facial characteristics. The appropriate selection of the training set directly determines the validity of the final results. The database should contain several examples of face images for each subject in the training set and at least one example in the test set. These examples should represent different frontal views of subjects with minor variations in view angle. They should also include different facial expressions, different lighting and background conditions, and examples with and without glasses. It is assumed that all images are already normalized to $m \times n$ arrays and that they contain only the face regions and not much of the subjects' bodies.

Second, for each image and subimage, starting with the two-dimensional $m \times n$ array of intensity values $I(x, y)$, we construct the lexicographic vector expansion $\phi \in R^{m \times n}$. This vector corresponds to the initial representation of the face. Thus the set of all faces in the feature space is treated as a high-dimensional vector space.

Third, by defining all instances of the same person's face as being in one class and the faces of different subjects as being in different classes for all subjects in the training set, we establish a framework for performing a cluster separation analysis in the feature space. Also, having labeled all instances in the training set and having defined all the classes, we compute the within- and between-class scatter matrices as follows:

$$S_w^{(V)} = \sum_{i=1}^{L} \Pr(C_i)\Sigma_i , \tag{1}$$

$$S_b^{(V)} = \sum_{i=1}^{L} \Pr(C_i)(\mu - \mu_i)(\mu - \mu_i)^T. \tag{2}$$

Here S_w is the within-class scatter matrix showing the average scatter Σ_i of the sample vectors (V) of different classes C_i around their respective mean, vectors μ_i:

$$\Sigma_i = E[(V - \mu_i) \times (V - \mu_i)^T | C = C_i]. \tag{3}$$

Similarly, S_b is the between-class scatter matrix, representing the scatter of the conditional mean, vectors (μ_i) around the overall mean vector μ. $\Pr C_i$ is the probability of the ith class. The discriminatory power of a representation can be quantified by using various measures. In this paper we use the separation matrix, which shows the combination of within- and between-class scatters of the feature points in the representation space. The class separation matrix and a measure of separability can be computed as

$$S^{(V)} = S_w^{-1} S_b \tag{4}$$

$$J_V = \mathrm{sep}(V) = \mathrm{trace}(S^{(V)}) \tag{5}$$

J_V is our measure of the discrimination power (DP) of a given representation V. As mentioned above, the representation may correspond to the data in its original form (e.g., a gray-scale image), or it can be based on a set of abstract features computed for a specific task.

For example, through this analysis we are able to compare the DP's of different spatial segments (components) of a face. We can apply the analysis to segments of the face images such as the areas around the eyes, mouth, hair, and chin or combinations of them. Figure 1 shows a separation analysis for horizontal segments of the face images in the database. The results show that the DP's of all segments are comparable and that the area between the nose and the mouth has more identification information than other parts. Figure 2 shows that the DP of the whole image is significantly larger than the DP's of its parts.

Using wavelet transforms[28–30] as multiscale orthogonal representations of face images, we can also perform a comparative analysis of the DP's of subimages in the wavelet domain. Different components of a wavelet decomposition capture different visual aspects of a grayscale image. As Fig. 3 shows, at each level of decomposition there are four orthogonal subimages corresponding to

- **LL:** the smoothed, low-frequency variations.
- **LH:** sharp changes in the horizontal direction, i.e., vertical edges.
- **HL:** sharp changes in the vertical direction, i.e., horizontal edges.
- **HH:** sharp changes in nonhorizontal, nonvertical directions, i.e., other edges.

We applied the LDA to each subimage of the wavelet transform (WT) of the face and estimated the DP of each

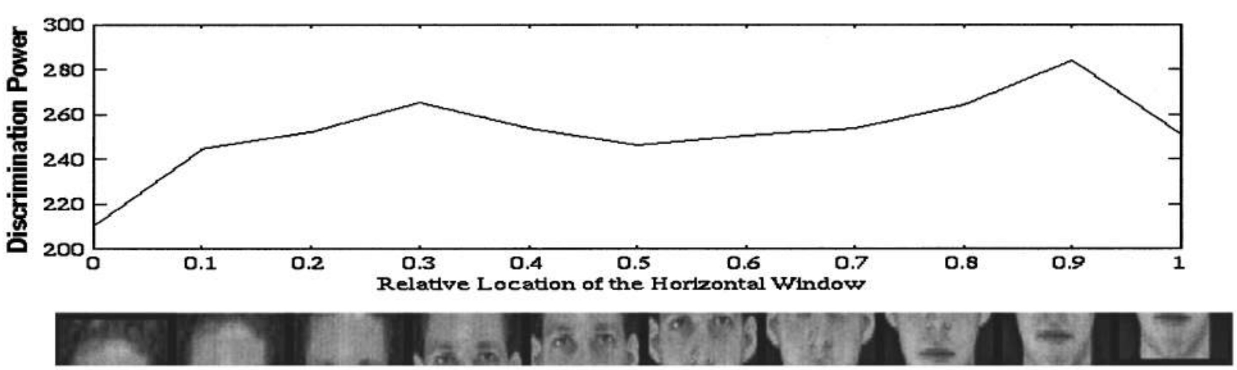

Fig. 1. Variation of the discrimination power of horizontal segments of the face defined by a window of fixed height sliding from top to bottom of the image.

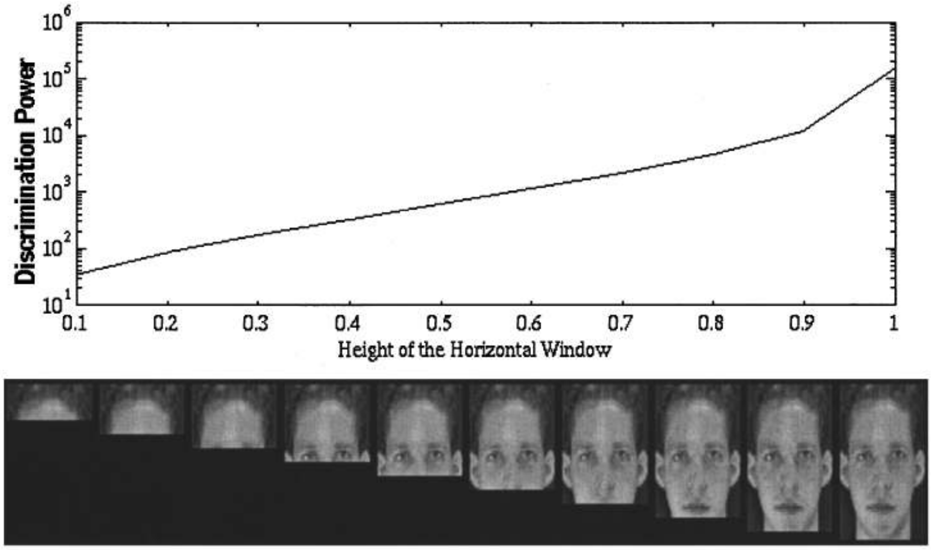

Fig. 2. Variation of the discrimination power of a horizontal segment of the face that grows in height from top to bottom of the image.

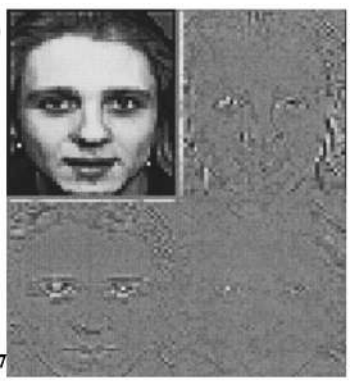

Fig. 3. Different components of a wavelet transform that capture sharp variations of the image intensity in different directions and have different discrimination potentials. The numbers represent the relative discrimination power.

subband. Figure 3 compares the separations obtained by using each of the subbands. Despite their equal sizes, different subimages carry different amounts of information for classification; the low-resolution components are the most informative. The horizontal edge patterns are almost as important as the vertical edge patterns, and their relative importance depends on the scale. Finally, the least important component in terms of face discrimination is the fourth subband, i.e., the slanted edge patterns. These results are consistent with our intuition and also with subjective psychovisual experiments.

One can also apply this idea to the study of the importance of facial components for gender or race classification from images.

3. DISCRIMINANT EIGENFEATURES FOR FACE RECOGNITION

In this section we propose a new algorithm for face recognition that makes use of a small yet efficient set of discriminant eigentemplates. The analysis is similar to the method suggested by Pentland and colleagues,[18,19] which is based on PCA. The fundamental difference is that in our system eigenvalue analysis is performed on the separation matrix rather than on the covariance matrix.

Human face images as two-dimensional patterns have a lot in common and are spectrally quite similar. Therefore, considering the face image as a whole, one expects to see important discriminant features that have low energies. These low-energy discriminant features may not be captured in a compression-based feature-extraction scheme such as PCA, or even in multilayer neural networks, which rely on minimization of the average Euclidean error. In fact, there is no guarantee that the error incurred by applying the compression scheme, despite its low energy, does not carry significant discrimination information. Also, there is no reason to believe that for a given compression-based feature space, feature points corresponding to different poses of the same subject will be closer (in Euclidean distance) to one another than to those of other subjects. In fact, it has been argued and experimentally shown that ignoring the first few eigenvectors, corresponding to the top principal components, can lead to a substantial increase in recognition accuracy.[19,31] Therefore the secondary selection from PCA vectors is based on their discrimination power. But one could ask, why do we not start with criteria based on discrimination rather than on representation from the beginning to make the whole process more consistent?

The PCA approach provides us with features that capture the main directions along which face images differ the most, but it does not attempt to reduce the within-class scatter of the feature points. In other words, since no class membership information is used, examples of the same class and of different classes are treated the same way. LDA, however, uses the class membership information and allows us to find eigenfeatures and therefore representations in which the variations among different faces are emphasized, while the variations of the same face that are due to illumination conditions, facial expression, orientation etc., are de-emphasized.

According to this observation, and on the basis of the results that follow, we believe that for classification purposes LDA-based feature extraction seems to be an appropriate and logical alternative to PCA or any other compression-based system that tries to find the most compact representation of face images. Concurrently but independently of our studies, LDA has been used by Swets and Weng[32,33] to discriminate human faces from other objects.

To capture the inherent symmetry of basic facial features and the fact that a face can be identified from its mirror image, we can use the mirror image of each example as a source of information.[17] Also, by adding noisy but identifiable versions of given examples, we can expand our training data and improve the robustness of the feature extraction against a small amount of noise in the input. Therefore for each image in the database we include its mirror image and one of its noisy versions, as shown in Fig. 4. Let Φ denote the face database, i.e.,

$$\Phi = \{\Phi_s : s = 1, 2, \ldots, N_S\}, \tag{6}$$

$$\Phi_s = \{\phi_i{}^s, \tilde{\phi}_i{}^s, (\phi_i{}^s + \nu) : i = 1, 2, \ldots, N_E, \quad \nu = [N(0, \sigma^2)]^{m \times n}\}, \tag{7}$$

where $\tilde{\phi}_i{}^s$ and $\phi_i{}^s + \nu$ are mirror images and noisy versions, respectively, of ϕ_i, the ith example of subject s in the data base Φ. Also, N_S is the number of subjects and N_E is the number of examples per subject in the initial database. Following our earlier observations, and having determined the separation matrix, we perform an eigenvalue analysis of the separation matrix $S^{(\Phi)}$ on the augmented database:

$$\text{eig}\{S^{(\Phi)}\} = \{(\lambda_i, u_i), \quad i = 1, \ldots, N_S - 1, \lambda_i > \lambda_{i+1}\}. \tag{8}$$

To reduce the computational cost for large data-set sizes, one can use the following equality[32,34]:

$$S_b u_i = \lambda_i S_w u_i . \tag{9}$$

This shows that the u_i's and λ_i's are generalized eigenvectors of $\{S_b, S_w\}$. From this equation the λ_i's can be computed as the roots of the characteristic polynomial

$$|S_b - \lambda_i S_w| = 0, \tag{10}$$

and then the u_i's can be obtained by solving

$$(S_b - \lambda_i S_w) u_i = 0 \tag{11}$$

Fig. 4. For each example in the database we add its mirror image and a noisy version.

only for the selected largest eigenvectors.[32] Note that the dimensionality m of the resulting set of feature vectors is $m < \text{rank}(S) = \min(n, N_S - 1)$. Now define

$$\Lambda^{(m)} = \{\lambda_i, i = 1, \ldots, m < N_S - 1\}, \tag{12}$$

$$U^{(m)} = \{u_i, i = 1, \ldots, m < N_S - 1\}, \tag{13}$$

so that $\Lambda^{(m)}$ and $U^{(m)}$ represent the set of m largest eigenvalues of $S^{(\Phi)}$ and their corresponding eigenvectors. Considering $U^{(m)}$ as one of the possible linear transformations Ω from $\mathbf{R}^n$ to $\mathbf{R}^m$, with $m < n$, one can show that

$$\Omega = \{U : X \subset \mathbf{R}^n \rightarrow U^T X = Y \subset \mathbf{R}^m, m < n\}, \tag{14}$$

$$U^{(m)} = \text{argmin}_{U \in \Omega}\{J_X - J_{U^T X}\}, \tag{15}$$

where $J_X = \text{tr}(S^{(X)})$ and $J_Y = \text{tr}(S^{(Y)})$ are separabilities computed over the X and $Y = U^T X$ spaces, respectively. This means that $U^{(m)}$ minimizes the drop $|\text{sep}(X) - \text{sep}(U^T X)|$ in classification information incurred by the reduction in the feature space dimensionality, and no other $\mathbf{R}^n$ to $\mathbf{R}^m$ linear mapping can provide more separation than $U^{(n)}$ does.

Therefore the optimal linear transformation from the initial representation space in $\mathbf{R}^n$ to a low-dimensional feature space in $\mathbf{R}^m$, which is based on our selected separation measure, results from projecting the input vectors ϕ onto m eigenvectors corresponding to the m largest eigenvalues of the separation matrix $S^{(\Phi)}$. These optimal vectors can be obtained from a sufficiently rich training set and can be updated if needed.

The columns of $U^{(m)}$ are the eigenvectors corresponding to the m largest eigenvalues; they represent the directions along which the projections of the face images within the database show the maximum class separation.

Each face image in the database is represented, stored, and tested in terms of its projections onto the selected set of discriminant vectors, i.e., the directions corresponding to the largest eigenvalues of $S^{(F)}$:

$$\forall \phi_i{}^s \in \Phi_s, \forall u \in U^{(m)} : \psi_i{}^s(u) = \langle \phi_i{}^s, u \rangle, \tag{16}$$

$$\Psi^s = \{\Psi_i{}^s(u) : \forall u \in U^{(m)}, \quad I = 1, \ldots, N_S\}. \tag{17}$$

Although all images of each subject are considered in the process of training, only one of them needs to be saved, as a template for testing. If a view-based approach is taken, one example for each distinct view has to be stored. Since only the projection coefficients need to be saved, for each subject we retain the example that is closest to the mean of the corresponding cluster in the feature space. Storing the projection coefficients instead of the actual images is highly desirable when large databases are used. Also, applying this holistic LDA to multiscale representations of face images, one can obtain multiscale discriminant eigentemplates. For example, one can apply LDA to each component of the WT of face images and select the most discriminant eigentemplates obtained from various scales. This approach is more complex because it requires the WT computation of each test example, but in some applications it may be useful, for example, when the DP of the original representation is not captured in the first few eigenvectors or when the condition of $m < N_{\text{classes}} - 1$) becomes restrictive, e.g., in gender classification.

4. MULTISOURCE SOFT-DECISION INTEGRATION

A number of different approaches have been proposed for analyzing information obtained from several sources.[34–36] The simplest method is to form an extended data (feature) vector, containing information from all the sources and treat this vector as the vector output of a single source. Usually, in such systems all similarities and distances are measured in the Euclidean sense. This approach can be computationally expensive; it is successful only when all the sources have similar statistical characteristics and comparable reliabilities. In our application this assumption is not valid, and therefore a more intelligent alternative approach has to be taken.

Each projection of the input pattern onto a discriminant vector u_i creates a decision axis with a certain level of reliability and discrimination power. The level of significance or reliability α_i of the decisions based on u_i is directly related to the class separation along that axis that is equal to the corresponding (normalized) eigenvalue in the LDA:

$$\forall(\lambda_i, u_i) \in (\Lambda^{(m)} \times U^{(m)}) : \alpha_i = \frac{\lambda_i}{\sum_{i=1}^{m} \lambda_i}. \tag{18}$$

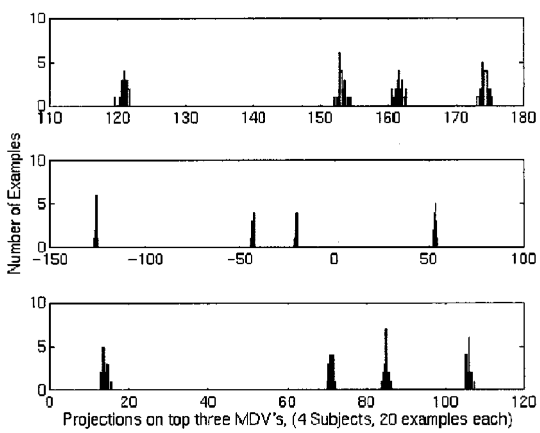

Fig. 5. Distribution of projection coefficients along three discriminant vectors with different levels of discrimination power for several poses from four different subjects.

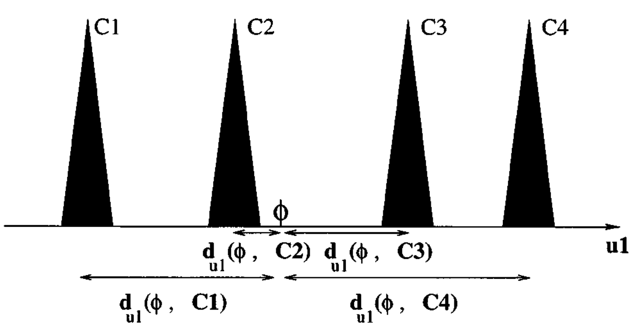

Fig. 6. Raw distances between each test example and the known clusters along each discriminant axis reuslt in the soft decision along that axis.

Figure 5 shows the distribution of projection coefficients onto three discriminant vectors. For any test vectorized face image ϕ, we project the image onto each of the top discriminant vectors u. On the basis of the distances between the resulting coefficients $\phi(u)$ and those of the existing templates $\psi_u{}^s$ stored in the database, we estimate the level of similarity of the input image to each known subject (see Fig. 6):

$$\forall u \in U^{(m)} : \phi(u) = \langle \phi, u \rangle, \tag{19}$$

$$\forall s \in \tilde{S} : d_u(\phi, s) = |\phi(u) - \psi_u{}^s|, \tag{20}$$

$$\pi_u(\phi, s) = 1 - \frac{\mathrm{d}_u(\phi, s)}{\sum_{s \in \tilde{S}} \mathrm{d}_u(\phi, s)}, \tag{21}$$

where $\pi_u(\phi, s)$ reflects the relative level of similarity between input ϕ and subject s according to source u, which has reliability α_u.

Having determined our decision axis and the reliabilities, we can apply a probabilistic or an evidential scheme of multisource data analysis to combine the soft decisions made on the basis of the individual imprecise sources to obtain a more precise and reliable final result. The normalized similarity measures (π's) indicate the proportions of evidence suggested by different sources. They can be interpreted as the so-called basic masses of evidence or they can be used as rough estimates of posterior probabilities given each measurement. From this stage on, a probabilistic or an evidential reasoning approach can be taken to combine basic soft decisions. A comparative study of various probabilistic and evidential reasoning schemes is given in Ref. 35.

Similarly working with distances as dissimilarity measures, one can combine a basic soft decision and incorporate the reliability of each source to define a reasonable measure of distance in the feature space. Although the most common measure used in the literature is Euclidean distance, as a more reasonable measure we suggest a weighted-mean absolute square distance, with the weights based on the discrimination powers. In other words,

$$\delta_u(\phi, s) = \frac{\mathrm{d}_u(\phi, s)}{\sum_{s \in \tilde{S}} \mathrm{d}_u(\phi, s)} \tag{22}$$

$$D(\phi, s) = \sum_{u \in U^{(m)}} [\delta_u(\phi, s) \times \alpha_u]. \tag{23}$$

Therefore for a given input ϕ the best match s^0 and its confidence measure is

$$s^0 = \mathrm{argmin}_{s \in \tilde{S}} \{D(\phi, s)\}, \tag{24}$$

$$\mathrm{conf}(\phi, s^0) = 1 - \frac{D(\phi, s^0)}{D(\phi, s')}, \tag{25}$$

where s' is the second-best candidate and conf stands for confidence measure. In this framework, incorporating collateral information or prior knowledge and expectations from context becomes very easy and logical. All we need to do is to consider each of them as an additional source of information corresponding to a decision axis

Table 1. Summary of Recognition Rates

Task	No. of Examples	No. of Features	Recognition Rate (%) (Training Set)	Recognition Rate (%) (Test Set)
Face recognition	2000	4	100	99.2
Gender classification	400	1	100	95

with a certain reliability and include it in the decision process. See Table 1 for a summary of recognition rates.

5. EXPERIMENTS AND RESULTS

In our experiments, to satisfy the requirements mentioned in Section 2 we used a mixture of two databases. We started with the database provided by Olivetti Research Ltd.[37] This database contains 10 different images of each of 40 different subjects. All the images were taken against a homogeneous background, and some were taken at different times. The database includes frontal views of upright faces with slight changes in illumination, facial expression (open or closed eyes, smiling or nonsmiling), facial details (glasses or no glasses), and some side movements. Originally we chose this database because it contained many instances of frontal views for each subject. Then, to increase the size of the database, we added some hand-segmented face images from the FERRET database.[38] We also included mirror-image and noisy versions of each face example to expand the data set and improve the robustness of recognition performance to image distortions. The total number of images used in training and in testing were approximately 1500 and 500, respectively. Each face was represented by a 50×60 pixel 8-bit gray-level image, which for our experiments was reduced to 25×30. The database was divided into two disjoint training and test sets. Using this composite database, we performed several tests on gender classification and face recognition.

The first test was on gender classification with use of a subset of the database containing multiple frontal views of 20 males and 20 females of different races. LDA was applied to the data, and the most discriminant template was extracted. Figure 7 shows this eigentemplate and the distribution of projection coefficients for all images in the set. As Fig. 7 shows, with only one feature very good separation can be achieved. Classification tests on a disjoint test set also gave 95% accuracy. Also, applying this discriminant template to a set of new faces from individuals outside the training set reduced the accuracy to ~92%.

As mentioned above, one can also apply LDA to wavelet transforms of face images and extract the most discriminant vectors of each transform component and combine multiscale classification results by using the proposed method of soft-decision integration.

We then applied LDA to a database of 1500 faces, with 60 classes corresponding to 60 individuals. Figure 8 shows the discriminatory power of the top 40 eigenvectors chosen according to PCA and LDA. As Fig. 8 depicts, the classification information of the principal components does not decrease monotonically with their energy; in other words, there are many cases in which a low-energy component has a higher discriminatory power than a high-energy component. The figure also shows that the top few discriminant vectors from LDA contain almost all the classification information embedded in the original image space.

Figure 9 shows the separation of clusters for ten poses of four different individuals, obtained with use of the two most discriminatory eigenvectors or eigenpictures. As Figure 9 indicates, the differences among classes (individuals) are emphasized, while the variations of the same face in different poses are deemphasized. The separation is achieved despite all the image variations resulting from the various poses of each subject. Figure 10 shows the distribution of clusters for 200 images of 10 subjects in the best two-dimensional discriminant feature space and in the best two-dimensional PCA-based space.

For each test face example, we first projected it onto the

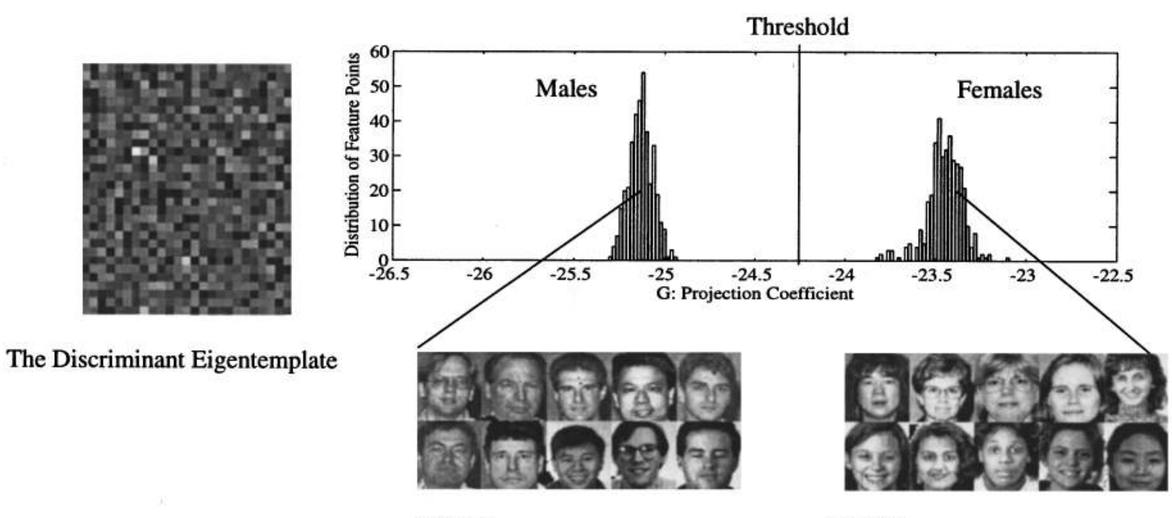

Fig. 7. Distribution of feature points for male and female examples in the database.

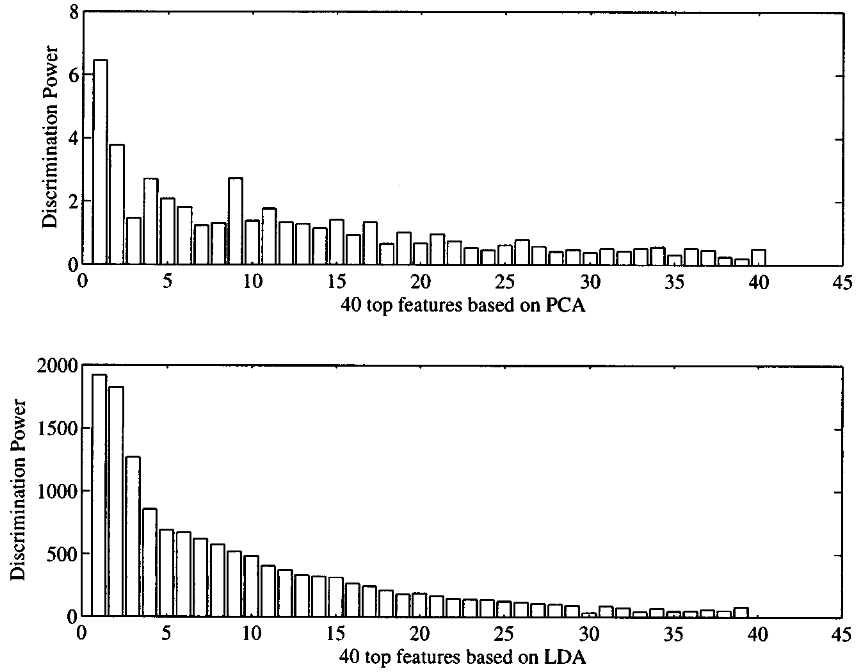

Fig. 8. Comparison of DP's of the top 40 selected eigenvectors based on PCA and LDA.

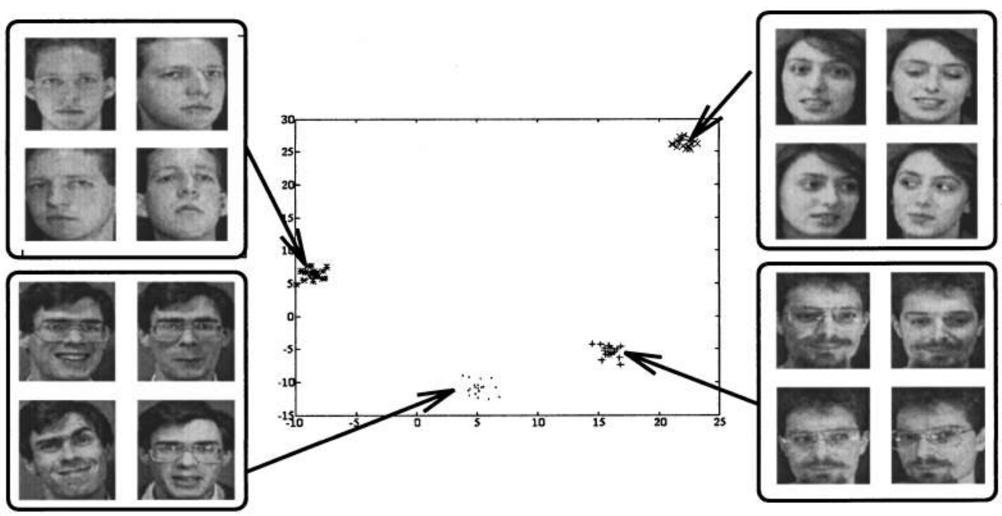

Fig. 9. Separation of clusters in the selected two-dimensional feature space. Four clusters correspond to variations of the faces of four different subjects in the database.

selected eigenvectors and found the distance from the corresponding point in the four-dimensional feature space to all of the previously saved instances. All distances were measured according to Eq. (23), and the best match was selected. For the given database, excellent (i.e., 99.2%) accuracy was achieved on the test set.

To evaluate the generalization of the feature extraction beyond the original training and test sets, we tested the classification results on pictures of new individuals, none of whom was present in the training set. Because of our limitations in terms of data availability, we could use only ten new subjects with two pictures per subject: one saved in the database as a template and two for testing. As expected, the application of the projection templates to these completely new faces resulted in a reduction in classification accuracy to $\sim$90%.

This reduction was expected, considering the fact that we did not have a very large training set. Extracting discriminant facial features from a large training set with diverse examples should improve the generalization and performance of the system on recognition of subjects outside the training set.

The simplicity of our systems, the size of the database, and the robustness of the results to small variations of the pose and the noise show that our suggested scheme is a good alternative approach to face recognition. It provides highly competitive results at much lower complexity with the use of low-dimensional feature sizes.

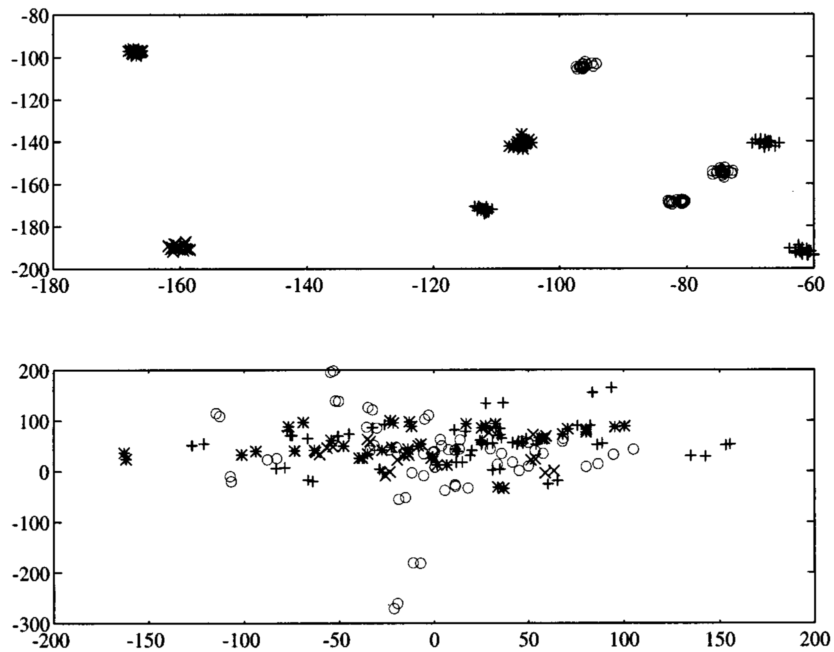

Fig. 10. Cluster separation in the best two-dimensional feature space. Top, based on LDA; bottom, based on PCA.

6. CONCLUSIONS

The application of LDA to study the discriminatory power of various facial features in spatial and wavelet domain is presented. Also, an LDA-based feature extraction for face recognition is proposed and tested. A holistic projection-based approach to face feature extraction is taken in which eigentemplates are the most discriminant vectors derived from LDA of face images in a rich enough database. The effectiveness of the proposed LDA-based features is compared with that of PCA-based eigenfaces. For classification a variation of evidential reasoning is used, in which each projection becomes a source of discriminating information, with reliability proportional to its discrimination power. The weighted combination of similarity or dissimilarity scores suggested by all projection coefficients is the basis for membership values.

Several results on face recognition and gender classification are presented, in which highly competitive recognition accuracies are achieved with a small number of features. The feature extraction can be applied to WT representation of images to provide a multiscale discriminant framework. In such cases the system becomes more complex at the expense of improving separability and performance. The proposed feature extraction combined with soft classification seems to be a promising alternative to other face-recognition systems.

The support of this research by the Advanced Research Projects Agency (ARPA Order No. C635) and the U.S. Office of Naval Research under contract N00014-95-1-0521 is gratefully acknowledged.

REFERENCES

1. R. Baron, "Mechanisms of human facial recognition," Int. J. Man–Machine Studies **15**, 137–178 (1981).
2. G. Davies, H. Ellis, and E. J. Shepherd, *Perceiving and Remembering Faces* (Academic, New York, 1981).
3. Y. Yacoob and L. S. Davis, "Computing spatio-temporal representations of human faces," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition* (IEEE Computer Society, Los Alamitos, Calif., 1994), pp. 70–75.
4. R. Brunelli and T. Poggio, "HyperBF networks for gender classification," in *Proceedings of the DARPA Image Understanding Workshop* (Defense Advanced Research Projects Agency, Arlington, Va., 1992), pp. 311–314.
5. B. A. Golomb and T. J. Sejnowski, "SEXNET: A neural network identifies sex from human faces," in *Advances in Neural Information Processing Systems 3*, D. S. Touretzky and R. Lipmann, eds. (Morgan Kaufmann, San Mateo, Calif., 1991), pp. 572–577.
6. A. Samal and P. Iyengar, "Automatic recognition and analysis of human faces and facial expressions: a survey," Pattern Recog. **25**, 65–77 (1992).
7. R. Chellappa, C. L. Wilson, and S. Sirohey "Human and machine recognition of faces, a survey," Proc. IEEE **83**, 705–740 (1995).
8. V. Govindaraju, S. N. Srihari, and D. B. Sher, "A computational model for face location," in *Proceedings of the Third International Conference on Computer Vision* (IEEE Computer Society Press, Los Alamitos, Calif., 1990), pp. 718–721.
9. A. Shio and J. Sklansky, "Segmentation of people in motion," in *Proceedings of the IEEE Workshop on Visual Motion* (Institute of Electrical and Electronics Engineers, Piscataway, N.J., 1991), pp. 325–332.
10. G. Yang and T. S. Huang, "Human face detection in a scene," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (Institute of Electrical and Electronics Engineers, Piscataway, N.J., 1993), pp. 453–458.
11. A. Pentland and B. Moghaddam, "Probabilistic visual learning for object detection," in *Proceedings of the International Conference on Computer Vision* (IEEE Computer Society Press, Los Alamitos, Calif., 1995), pp. 786–793.
12. K. Sung and T. Poggio, "Example-based learning for view based human face detection," *Proceedings of the IEEE Im-*

age Understanding Workshop (Institute of Electrical and Electronics Engineers, Piscataway, N.J., 1994), pp. 843–850.

13. C. Wu and J. Huang, "Huang face profile recognition by computer," Pattern Recog. **23**, 255–259 (1990).
14. G. Gordon, "Face recognition based on depth maps and surface curvature," in *Geometric Methods in Computer Vision*, B. C. Vemuri, ed., Proc. SPIE **1570**, 234–247 (1991).
15. O. Nakamura, S. Mathur, and T. Minami, "Identification of human faces based on isodensity maps," Pattern Recog. **24**, 263–272 (1991).
16. Y. Cheng, K. Liu, J. Yang, and H. Wang, "A robust algebraic method for human face recognition," in *Proceedings of the 11th International Conference on Pattern Recognition* (IEEE Computer Society Press, Los Alamitos, Calif., 1992), pp. 221–224.
17. M. Kirby and L. Sirovich, "Application of the Karhunen–Loève procedure for the characterization of human faces," IEEE Trans. Pattern. Anal. Mach. Intell. **12**, 103–108 (1990).
18. M. A. Turk and A. P. Pentland, "Face recognition using eigenfaces," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition* (IEEE Computer Society, Los Alamitos, Calif., 1991), pp. 586–591.
19. A. Pentland, B. Moghaddam, T. Starner, and M. Turk, "View-based and modular eigenspaces for face recognition," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition* (IEEE Computer Society, Los Alamitos, Calif., 1994), pp. 84–91.
20. L. Sirovich and M. Kirby, "Low-dimensional procedure for the characterization of the human face," J. Opt. Soc. Am. A **4**, 519–524 (1987).
21. I. Craw, D. Tock, and A. Bennett, "Finding face features," in *Proceedings of the Second European Conference on Computer Vision* (Springer-Verlag, Berlin, 1992), pp. 92–96.
22. B. S. Manjunath, R. Chellappa, and C. v. d. Malsburg, "A feature based approach to face recognition," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition* (IEEE Computer Society, Los Alamitos, Calif., 1992), pp. 373–378.
23. M. Lades, J. Vorbruggen, J. Buhmann, J. Lange, C.v.d. Malsburg, and R. Wurtz, "Distortion invariant object recognition in the dynamic link architecture," IEEE Trans. Comput. **42**, 300–311 (1993).
24. M. Seibert and A. Waxman, "Recognizing faces from their parts," in *Sensor Fusion IV: Control Paradigms and Data Structures*, P. S. Schenker, ed., Proc. SPIE **1616**, 129–140 (1991).
25. A. Rahardja, A. Sowmya, and W. Wilson, "A neural network approach to component versus holistic recognition of facial expressions in images," in *Intelligent Robots and Computer Vision X: Algorithms and Techniques*, D. P. Casasent, ed., Proc. SPIE **1607**, 62–70 (1991).
26. A. Yuille, D. Cohen, and P. Hallinan, "Feature extraction from faces using deformable templates," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition* (IEEE Computer Society, Los Alamitos, Calif., 1989), pp. 104–109.
27. D. Marr, *Vision* (Freeman, San Francisco, Calif., 1982).
28. R. R. Coifman and M. V. Wickerhauser, "Entropy based algorithms for best basis selection," IEEE Trans. Inf. Theory **38**, 713–718 (1992).
29. S. G. Mallat, "A theory for multi-resolution signal decomposition, the wavelet representation," in IEEE Trans. Pattern. Anal. Mach. Intell. **11**, 674–693 (1989).
30. I. Daubechies, "Orthonormal basis of compactly supported wavelets," Commun. Pure Appl. Math. **41**, 909–996 (1988).
31. A. O'Toole, H. Abdi, K. Deffenbacher, and D. Valentin, "Low-dimensional representation of faces in higher dimensions of the face space," J. Opt. Soc. Am. A **10**, 405–410 (1993).
32. D. L. Swets and J. J. Weng, "SHOSLIF-O: SHOSLIF for object recognition (phase I)," in Tech. Rep. CPS 94-64 (Michigan State University, East Lansing, Mich., 1994).
33. D. L. Swets, B. Punch, and J. J. Weng, "Genetic algorithms for object recognition in a complex scene," in *Proceedings of the IEEE International Conference on Image Processing* (Institute of Electrical and Electronics Engineers, Piscataway, N.J., 1995), pp. 595–598.
34. K. Fukunaga, "Statistical Pattern Recognition" (Academic, New York, 1989).
35. T. Lee, J. A. Richards, and P. H. Swain, "Probabilistic and evidential approaches for multisource data analysis," IEEE Trans. Geosci. Remote Sens. **GE-25**, 283–293 (1987).
36. P. L. Bogler "Shafer–Dempster reasoning with applications to multisensor target identification systems," IEEE Trans. Syst. Man Cybern. **17**, 968–977 (1987).
37. F. Samaria and A. Harter, "Parameterization of a stochastic model for human face identification," in *Second IEEE Workshop on Applications of Computer Vision* (Institute of Electrical and Electronics Engineers, Piscataway, N.J., 1994).
38. P. Rauss, P. J. Phillips, M. Hamilton, and A. T. DePersia, "FERET (Face Recognition Technology) Program," in *25th AIPR Workshop: Emerging Applications of Computer Vision,* D. Schaefer and W. Williams, eds., Proc. SPIE **2962**, 253–263 (1996).