**M.I.T Media Laboratory Perceptual Computing Section Technical Report No. 301**
**Appeared in:** *Automatic Systems for the Identification and Inspection of Humans, SPIE Vol. 2277, July 1994*

# Face Recognition using View-Based and Modular Eigenspaces

**Baback Moghaddam and Alex Pentland**
Perceptual Computing Group, The Media Laboratory
Massachusetts Institute of Technology
20 Ames St., Cambridge, MA 02139

## Abstract

In this paper we describe experiments using *eigenfaces* for recognition and interactive search in the FERET face database. A recognition accuracy of 99.35% is obtained using frontal views of 155 individuals. This figure is consistent with the 95% recognition rate obtained previously on a much larger database of 7,562 "mugshots" of approximately 3,000 individuals, consisting of a mix of all age and ethnic groups. We also demonstrate that we can automatically determine head pose without significantly lowering recognition accuracy; this is accomplished by use of a *view-based* multiple-observer eigenspace technique. In addition, a modular eigenspace description is used which incorporates salient facial features such as the eyes, nose and mouth, in an *eigenfeature* layer. This modular representation yields slightly higher recognition rates as well as a more robust framework for face recognition. In addition, a robust and automatic feature detection technique using *eigentemplates* is demonstrated.

## 1 INTRODUCTION

In recent years considerable progress has been made on the problems of face detection and recognition, especially in the processing of "mug shots," i.e., head-on face pictures with controlled illumination and scale. The best results have been obtained for 2-D, view-based techniques based on either template matching (*e.g.,* [2]), combined feature-and-template matching (*e.g.,* [1]) or matching using "eigenfaces," *i.e.* template matching using the Karhunen-Loeve transformation of a set of face pictures (*e.g.,* [11, 12, 5]). However to date tests of these methods have been confined to datasets of only a few hundred images. For real-world applications, we must be able to reliably discriminate among thousands of individuals. Moreover, the problem of recognizing a human face from a *general* view remains largely unsolved, because transformations such as position, orientation, scale, and illumination cause the face's appearance to vary substantially. It is therefore important to ask if we can extend these successful 2-D, view-based recognition approaches to large databases with more general viewing conditions.

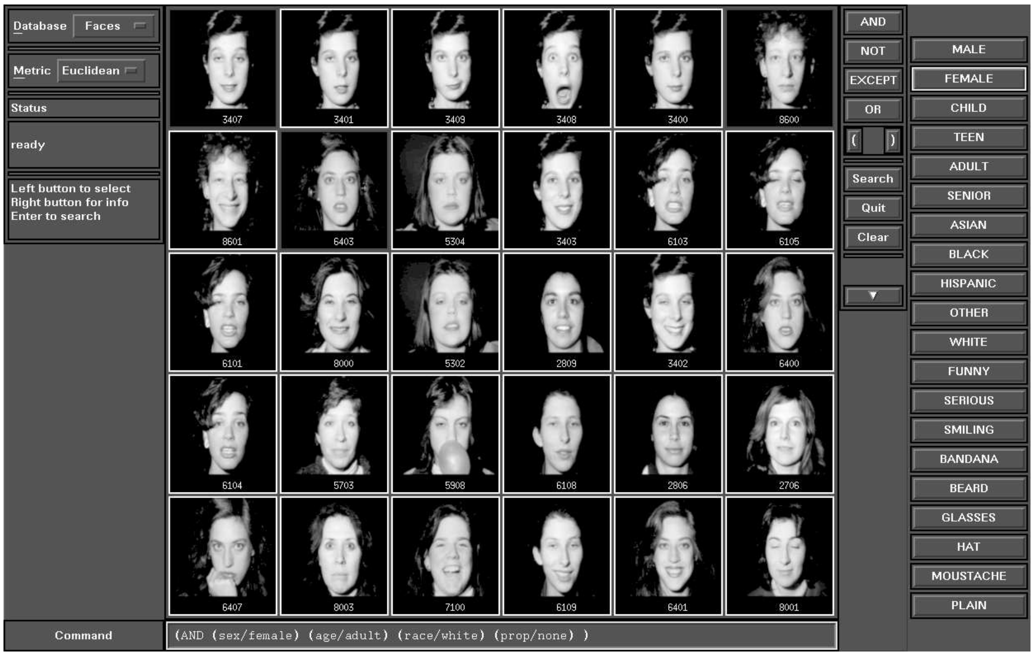
In this paper we first explore how the *eigenface* technique of Turk and Pentland [12] scales when applied to much larger recognition problems. We have recently extended the eigenface technique to a view-based and modular framework for automatic detection and recognition [9]. The view-based formulation allows us to automatically determine head orientation and scale. The modular description allows for the incorporation of important facial features such as eyes, nose and mouth. These extensions account for variations in head orientation, scale, hairstyle, and makeup, thus leading to a more robust face recognition system. Although the application reported in this paper is that of face recognition, the same technique can be applied to recognition and detection of most rigid, roughly convex objects. The general applicability of eigenvector decomposition methods for appearance-based 3D object recognition has recently been convincingly demonstrated by Murase and Nayar [7].

## 2 PHOTOBOOK: A database tool

To date, most face recognition experiments have had at most a few hundred faces. Thus how face recognition performance scales with the number of faces is almost completely unknown. In order to have an estimate of the recognition performance on much larger databases, we have conducted tests on a Media Lab database of $7,562$ images of approximately $3,000$ people. The eigenfaces for this database were approximated using a principal components analysis on a representative sample of 100 faces. Every image in the database was then encoded by a projection onto a 100-dimensional basis corresponding to the principal eigenvectors of the training data. Recognition and matching is then performed using a nearest-neighbor pattern matching. In addition, each image was then annotated (by hand) as to sex, race, approximate age, facial expression, and other salient features. Almost every person has at least two images in the database; several people have many images with varying expression, headwear, facial hair, etc.

Photobook is an X-windows browsing tool that allows the user to interactively search through image databases [8]. The user begins by selecting the types of faces they wish to examine; *e.g.,* FERET faces or caucasian males from the Media Lab database of 7,562 images. This subset selection is accomplished using an object-oriented database to search through the face image annotations. Photobook then presents the user with a screenful of the selected type of images, the rest of the images can be viewed by "paging" through the database. At any time the user can select a face from among those presented, and Photobook will then use the eigenvector description of that face to sort the entire set of faces in terms of their similarity to the selected face. Photobook then re-presents the user with the face images, now sorted by similarity to the selected face.

Figure 1(a) shows the typical results of such a similarity




(a)


(b)

Figure 1: (a) The face at the upper left was selected by the user; the remainder of the faces are the 20 most-similar faces found from among the entire 7,562 individuals in the database. Similarity decreases left to right, top to bottom. (b) An example using the FERET database; note the first four images are all of the same person.

search using the Media Lab face database. The face at the upper left of each set of images was selected by the user; the remainder of the faces are the next most-similar faces from among the entire $7,562$ Media Lab database. Similarity decreases left to right, top to bottom. Figure 1(b) shows the typical results of such a similarity search using the FERET database. The face at the upper left of each set of images was selected by the user; the remainder of the faces are the most-similar faces from the 575 frontal views in the FERET database. Note that the first four images (in the top row) are all of the same person (taken a month apart and exhibiting different hairstyles). Note that this database represents a more realistic application scenario where position, scale, lighting and background are not uniform. Consequently, an off-line pre-processing stage is used to correct for translation, scale, and contrast. Once the images are geometrically and photometrically normalized, they can be used in the standard eigenface technique. The entire searching and sorting operation takes less than one second on a standard Sun Sparcstation, because each face is described using only a very small number of eigenvector coefficients. Of particular importance is the ability to find the same person despite wide variations in expression, hairstyle, image size, and eyewear.

### 2.1 Recognition Accuracy

An early version of our face recognition system has obtained an accuracy of 95% on the Media Lab database of $7,562$ frontal images of $3,000$ people. Since this database has accurate registration and alignment, no normalization or pre-processing was used. This level of performance has proven that the eigenface technique does indeed scale favorably with larger databases.

To assess the recognition accuracy of our new system on the more challenging FERET database, we selected a subset consisting of the images of the 150 people for which all views were available. This subset of images includes the most recent imagery in which lighting and scale were approximately standardized. As in our previous work[9] we have used a *view-based* recognition paradigm for the multiple head orientations. This yields 5 separate eigensapces, one for every available view (frontal, half left, half right, profile left, profile right). Recognition and matching are performed in each space as with the standard eigenface technique. Figure 2 shows the recognition accuracies obtained with our system. Perhaps most important is the case of frontal view versus frontal view, which is the traditional "mugshot" situation. The accuracy of 99.4 corresponds to only one mistake in matching two frontal views of 150 people. Furthermore, this performance corresponds to completely automatic processing of the raw imagery. The front-end to our eigenface recognition system consists of several pre-processing stages which first detect and estimate the head location and scale, find the facial features and then normalize the geometry of the face. These stages correct for translation, scale, lighting, contrast, as well as slight rotations in the image plane.

Surprisingly, we found that the accuracy obtained by comparing left and right profiles, or left and right half views, is much lower than might have been expected. We believe there are two factors responsible for this decline in recognition accuracy: facial assymetry and a lack of consistency in head orientations in the left and right views. Although human faces are generally bilaterally symmetric, there are differences (confounded by hairstyle) which may be a problem in matching opposite views. The latter factor is merely a lack of calibration in the image acquisition which can not guarantee that a pair of left and right views are at the same angular offset from frontal.

Note that because of the lack of multiple images in all views, the off-frontal diagonal accuracies (profile vs. profile, half vs. half) were estimated in a manner different from the other entries. In these cases the data set is geometrically normalized according to ground truth data on facial feature locations (eyes, nose mouth, etc.). These normalized images are then treated as the training set, and matched with automatically normalized images. The slight variations between the manual and automatic alignment procedures will therefore simulate two different images for each person. The recognition rates obtained for this type of simulated test are shown in parentheses.

## 3 HEAD ORIENTATION

Our approach to automatically determining head orientation is to build a *view-based* set of $M$ separate eigenspaces, each capturing the variation of the $N$ individuals in a common view. The view-based eigenspace is essentially an extension of the eigenface technique to multiple sets of eigenvectors, one for each combination of scale and orientation. One can think of this architecture as a set of parallel "observers" each trying to explain the image data with their set of eigenvectors (see also Darrell and Pentland [3]).

In this view-based, multiple-observer approach, the first step is to determine the location and orientation of the target object by selecting the eigenspace which best describes the input image. This is accomplished by calculating the residual description error (the "distance-from-face-space" metric [12]) using each viewspace's eigenvectors. Once the proper viewspace is determined, the image is described using the eigenvectors of that viewspace, and then recognized. We have evaluated this approach using data similar to that shown in Figure 3. This data consists of 189 images consisting of nine views of 21 people. The nine views of each person were evenly spaced from $-90°$ to $+90°$ along the horizontal plane. Data were provided by Westinghouse Electronic Systems. The *interpolation* performance was tested by training on a subset of the available views $\{\pm 90°, \pm 45°, 0°\}$ and testing on the intermediate views $\{\pm 68°, \pm 23°\}$. The average recognition rate obtained was 92%.

## 4 EIGENFEATURES

The eigenface technique is easily extended to the description and coding of facial features, yielding "eigeneyes", "eigennoses" and "eigenmouths". Eye-movement studies indicate that these particular facial features represent important landmarks for fixation, especially in an attentive discrimination task. Therefore we should expect an improvement in recognition performance by incorporating an additional layer of description in terms of facial features. This can be viewed as either a modular or layered representation of a face, where a coarse (low-resolution) description of the whole head is augmented by additional (higher-resolution) details in terms of salient facial features.

| | Frontal | Half left | Half right | Profile left | Profile right |
|---|---|---|---|---|---|
| Frontal | 99 | ** | ** | ** | ** |
| Half left | ** | (87) | 38 | ** | ** |
| Half right | ** | 38 | (82) | ** | ** |
| Profile left | ** | ** | ** | (70) | 32 |
| Profile right | ** | ** | ** | 32 | (68) |

Figure 2: Percent correct recognitions for the FERET database.

Figure 3: Some of the images used to test accuracy at face recognition despite wide variations in head orientation. Average recognition accuracy was 92%, the orientation error had a standard deviation of 15°

This modularity in face description also has distinct advantages for face coding in teleconferencing. For example, a layered representation consisting of the face and eigenmouths has recently been implemented for low bit-rate transmission of visual telephony by Welsh and Shah [14]. In section 5, we will demonstrate the potential utility of eigenfeatures for face recognition.

## 4.1 Detection of facial features

An important pre-processing step in an eigenvector recognition system is that of registration. A face in an input image must first be located and registered in a standard-size frame before being processed. In addition to head detection and tracking, automatic detection of facial features is also an important component for face recognition. Over the years, various strategies for facial feature detection have been proposed, ranging from the early work of Kanade with edge-map projections [4], to more recent techniques using generalized symmetry operators [10] and multilayer perceptrons [13].

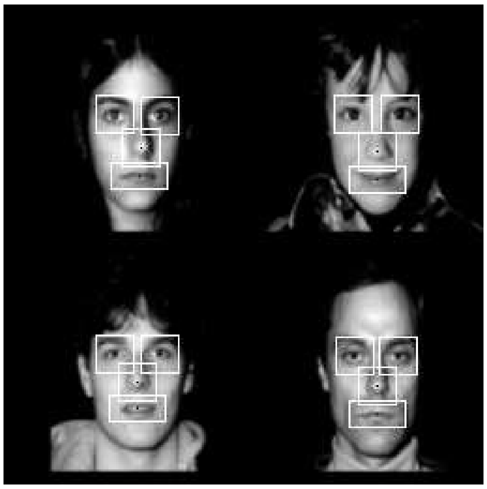
By far, the standard detection paradigm in computer vision is that of simple correlation or template matching. The eigenspace formulation, however, leads to a powerful alternative to simple template matching. The reconstruction error (or residual) of the principal component representation (referred to as the "distance-from-face-space" in the context of our earlier work [12]) is a an effective indicator of a match. The residual error is easily computed using the projection coefficients and signal energy. This detection strategy is equivalent to matching with *eigentemplates* and allows for a greater range of distortions in the input signal (including lighting, rotation and scale). In a statistical signal detection framework, the use of eigentemplates has been shown to yield superior performance in comparison with standard matched filtering [6].

In the eigenfeature representation the equivalent "distance-from-*feature*-space" (DFFS) can be effectively used for the detection of facial features. Given an input image, a feature distance-map is built by computing the DFFS at each pixel. When using $n$ eigenvectors, this requires $n$ convolutions (which can be efficiently computed using an FFT) plus an additional local energy computation. The global minimum of this distance map is then selected as the best feature match.

## 4.2 Detection on a large database

The DFFS feature detector was also used for the automatic detection and coding of the facial features in our large database of 7,562 faces. The same representative sample of 100 individuals used in computing the eigenfaces was used to compute a set of corresponding eigenfeatures. Figure 4(a) shows examples of the training templates used for the facial features (left eye, right eye, nose and mouth). The entire database was processed by using independent detectors for each feature (with the DFFS computed based on projection on the first 10 eigenvectors). The matches were obtained by independently selecting the global minimum in each of the four distance maps. Typical detections are shown in Figure 4(b).

The DFFS metric associated with each detection can be used in conjunction with a threshold — *i.e.,* only the global minima with a DFFS value *less* than the threshold are declared to be a possible match. Consequently we can characterize the detection vs. false-alarm tradeoff by varying this threshold and generating a *receiver operating char-*



(a)

(b)

Figure 4: (a) Examples of facial feature training templates used and (b) the resulting typical detections.

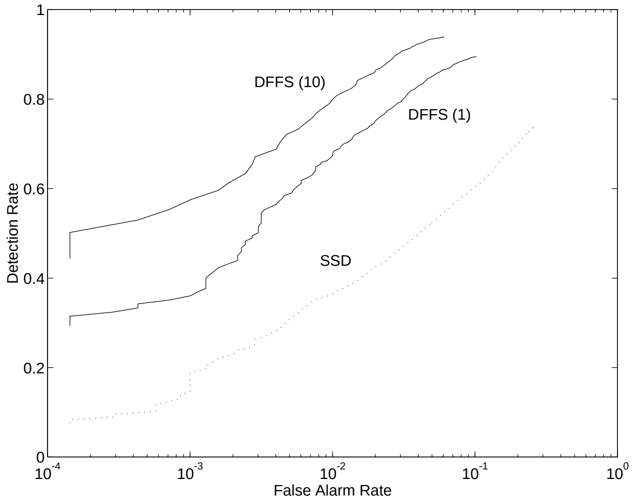



Figure 5: ROC curve for left eye using DFFS detectors with 1 and 10 eigenvectors. An SSD detector is shown for comparison.

*acteristics* (ROC) curve. Figure 5 shows the ROC curves for the left eye using the first and first 10 eigenvectors in the DFFS detector. A correct detection was defined as a below-threshold global minimum within 5 pixels of the mean left eye position. Similarly, a false alarm was defined as a below-threshold detection located *outside* the 5-pixel radius. Global minima *above* the threshold were undeclared. The peak performance of the DFFS detector using the first 10 eigenvectors corresponds to a 94% detection rate at a false alarm rate of 6%. Conversely, at a zero false-alarm rate, 52% of the eyes were correctly detected. To calibrate the performance of the DFFS detector, we have also shown the ROC curve corresponding to a standard sum-of-square-differences (SSD) template matching technique. The templates used in this case were the mean features in each case.

Note that the SSD can be considered a *degenerate* case of a DFFS detector, corresponding to a zero-th order encoding — *i.e.,* using only the mean vector for description. The addition of the principal components results in incremental improvements in detection performance, resulting in a gradation of ROC curves similar to those shown in Figure 5. Naturally, the incorporation of each additional eigenvector means an extra correlation. However, the increase in computational cost is linear with the number of eigenvectors and is typically offset by the subsequent gain in performance. In fact, as the ROC curves indicate, by using only the first eigenvector (at the cost of one additional convolution over SSD) we have substantially increased detection performance.

Finally, we note that the detection of facial features can be made more robust by incorporating constraints on the geometry of a face in terms of relative feature locations. These constraints can be used to guide the search for matches and thus restrict the regions over which a DFFS map is computed. This will not only reduce the number of false alarms but will also significantly reduce the computational cost. Preliminary experiments with such constraints indicate that the detection rate of mouths and noses can be greatly improved by "anchoring" the search with respect to more easily detected features, such as eyes.

# 5 MODULAR EIGENSPACES

With the ability to reliably detect facial features across a wide range of faces, we can automatically generate a modular representation of a face. The utility of this layered representation (eigenface plus eigenfeatures) was tested on a small subset of our face database. We selected a representative sample of 45 individuals with two views per person, corresponding to different facial expressions (neutral vs. smiling). These set of images was partitioned into a training set (neutral) and a testing set (smiling). Since the difference in the facial expressions is primarily articulated in the mouth, this particular feature was discarded for recog-

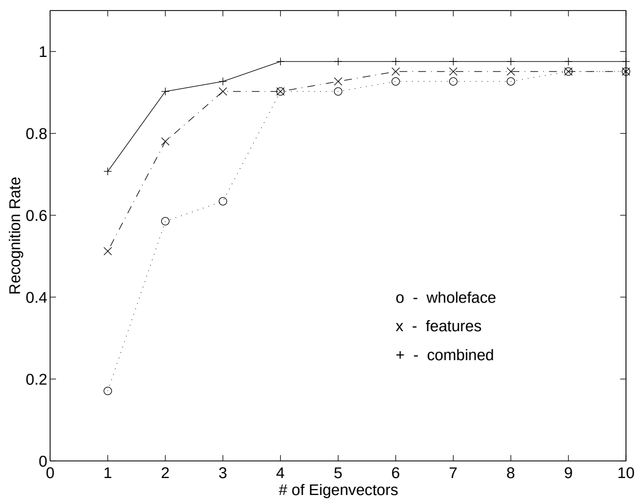



Figure 6: Recognition rates for eigenfaces, eigenfeatures and the combined modular representation.

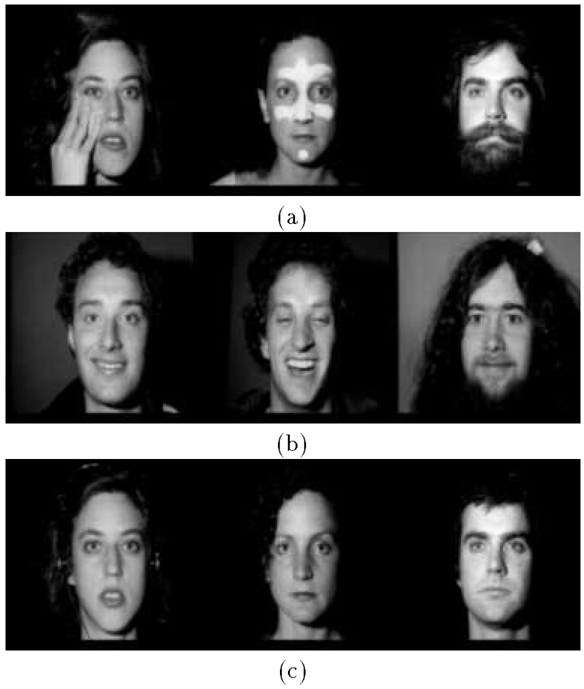



Figure 7: (a) Test views, (b) Eigenface matches, (c) Eigenfeature matches.

nition purposes. Figure 6 shows the recognition rates as a function of the number of eigenvectors for eigenface-only, eigenfeature-only and the combined representation. What is surprising is that (for this small dataset at least) the eigenfeatures alone were sufficient in achieving an (asymptotic) recognition rate of 95% (equal to that of the eigenfaces). More surprising, perhaps, is the observation that in the lower dimensions of eigenspace, eigenfeatures outperformed the eigenface recognition. Finally, by using the combined representation, we gain a slight improvement in the asymptotic recognition rate (98%). A similar effect has recently been reported by Brunelli and Poggio [2] where the cumulative normalized correlation scores of templates for the face, eyes, nose and mouth showed improved performance over the face-only templates.

A potential advantage of the eigenfeature layer is the ability to overcome the shortcomings of the standard eigenface method. A pure eigenface recognition system can be fooled by gross variations in the input image (hats, beards, etc.). Figure 7(a) shows additional testing views of 3 individuals in the above dataset of 45. These test images are indicative of the type of variations which can lead to false matches: a hand near the face, a painted face, and a beard. Figure 7(b) shows the nearest matches found based on a standard eigenface classification. Neither of the 3 matches correspond to the correct individual. On the other hand, Figure 7(c) shows the nearest matches based on the eyes and nose, and results in correct identification in each case. This simple example illustrates the advantage of a modular representation in disambiguating false eigenface matches.

We are currently exploring strategies for the optimal fusion of the available information in the modular representation. One simple approach is to form a cumulative score in terms of equal contributions by each of the components (head, eyes, nose and mouth). Alternatively, psychophysical data can be used in formulating a more elaborate weighting scheme for classification (*e.g.*, eyes tend to be the most salient features). A more ambitious scheme would be to modulate the contribution of each module in a task or state-dependent manner.

An attractive recognition strategy is to combine a sequential classifier with a coarse-to-fine matching procedure, whereby a pyramid sequence of (low-resolution) eigenface projections is used to limit the database search to a local region of facespace, and finally a (high-resolution) facial feature description is used to perform the final classification. By embedding this mechanism in the framework of our view-based eigenspace method, the overall system can perform robust face recognition under varying viewing geometries.

# 6 CONCLUSIONS

Our experimental results have demonstrated the success of eigenspace techniques for detection and recognition in a large face database. We have generalized this technique to handle a variable viewing geometry, using a *view-based* approach. We have described target objects in terms of their 2-D "aspects" (their appearance from a particular viewpoint). The key to the success of such a view-based approach is the ability to localize the object (or features on an object) and identify the correct aspect. We have also shown that the *distance-from-feature-space* computation in a view-based eigenspace formulation is an effective tool for robust detection and pose estimation.

Finally, we have extended the approach to a modular representation by incorporating information from different levels of description. Once again, the ability of the *distance-from-feature-space* computation to accurately and reliably detect features was critical for successfully incorporating a parts-based description. By using this modular approach we have been able to demonstrate robustness to localized variations in object appearance.


## Acknowledgements

The FERET face database as well as partial funding for this research was provided by the Army Research Lab, Ft. Belvoir, VA.


## References


[1] Bichsel, M., and Pentland, A., "Topological Matching for Human Face Recognition," M.I.T. Media Laboratory Vision and Modeling Group Technical Report No. 186, Jan. 1992 *to appear* CVGIP: Image Understanding

[2] Brunelli, R., and Poggio, T., "Face Recognition: Features vs. Templates,", *IEEE Trans. on Pattern Analysis and Machine Intelligence,*, vol. 15, no. 10, Oct. 1993.

[3] Darrell, T., and Pentland, A., "Space-Time Gestures," Proc. IEEE Conf. on Computer Vision and Pattern Recognition, NY NY, June 1993.

[4] Kanade, T., "Picture processing by computer complex and recognition of human faces," Tech. Report, Kyoto University, Dept. of Information Science, 1973.

[5] Kirby, M., and Sirovich, L., "Application of the Karhunen-Loeve procedure for the characterization of human faces," *IEEE Trans. Pattern Analysis and Machine Intelligence*, Vol. 12, No. 1, Jan. 1990.

[6] Kumar, B., Casasent, D., and Murakami, H., "Principal Component Imagery for Statistical Pattern Recognition Correlators," *Optical Engineering,* vol. 21, no. 1, Jan/Feb 1982.

[7] Murase, H., and Nayar, S. K., "Learning and Recognition of 3D Objects from Appearance" in *IEEE 2nd Qualitative Vision Workshop*, New York, June 1993.

[8] Pentland, A., Picard, R., and Sclaroff, S., "Photobook: Tools for Content-Based Manipulation of Image Databases," SPIE Storage and Retrieval Image and Video Databases II, No. 2185, San Jose, Feb 6-10, 1994.

[9] Pentland, A., Moghaddam, B., and Starner, T., "View-based and modular eigenspaces for face recognition," *Proc. IEEE Conf. on Computer Vision & Pattern Recognition*, Seattle, Washingtion, June 21-23, 1994.

[10] Reisfeld, D., Wolfson, H., and Yeshurun, Y., "Detection of Interest Points Using Symmetry,", *ICCV '90*, Osaka, Japan, Dec. 1990.

[11] Turk, M., and Pentland, A., "Face processing: models for recognition," *Intelligent Robots and Computer Vision VIII*, SPIE, Philadelphia, PA, 1989.

[12] Turk, M., and Pentland, A., "Eigenfaces for Recognition," *Journal of Cognitive Neuroscience,* Vol. 3, No. 1, pp. 71-86.

[13] Vincent, J. M., Waite, J. B., and Myers, D. J, "Automatic Location of Visual Features by a System of Multilayered Perceptrons," *IEE Proceedings,* vol. 139, no. 6, Dec. 1992.

[14] Welsh, J. W., and Shah, D., "Facial-Feature Image Coding Using Principal Components,", *Electronic Letters,* vol. 28, no. 22, October, 1992.