Face Recognition Using the Nearest Feature Line Method

Stan Z. Li and Juwei Lu

***Abstract*— In this paper, we propose a novel classification method, called the nearest feature line (NFL), for face recognition. Any two feature points of the same class (person) are generalized by the feature line (FL) passing through the two points. The derived FL can capture more variations of face images than the original points and thus expands the capacity of the available database. The classification is based on the nearest distance from the query feature point to each FL. With a combined face database, the NFL error rate is about 43.7–65.4% of that of the standard eigenface method. Moreover, the NFL achieves the lowest error rate reported to date for the ORL face database.**

***Index Terms*—Classification methods, eigenface, face recognition, nearest feature line, principal component analysis.**

I. INTRODUCTION

FACE recognition has a wide range of applications such as identity authentication, access control, and surveillance. Interest and research activities in face recognition have increased significantly over the past few years. A capable face recognition system should be able to deal with variations of face images in viewpoint, illumination, and expression. However, "the variations between the images of the same face due to illumination and viewing direction are almost always larger than image variations due to change in face identity" [4]. This makes face recognition a great challenge. Two issues are central: 1) what features to use to represent a face and 2) how to classify a new face image based on the chosen representation.

In geometric feature-based methods [1], [5], [6], facial features such as eyes, nose, mouth, and chin are detected. Properties and relations (e.g., areas, distances, angles) between the features are used as the descriptors of faces for recognition. Although economical and efficient in achieving data reduction and insensitive to variations in illumination and viewpoint, such features rely heavily on the extraction of facial features. Unfortunately, facial feature detection and measurement techniques developed to date have not been reliable enough to cater to this need [7].

In contrast, template matching and neural methods [2], [3] generally operate directly on an image-based representation (i.e., pixel intensity array). Because the detection and measurement of facial features are not required, this class of methods has been more practical and reliable as compared to geometric feature-based methods. Among various neural approaches, the convolutional neural network (CNN) [8] is a hybrid approach which combines local image sampling, a self-organizing map neural network, and a convolutional neural network. It has achieved the lowest error rate reported to date for the ORL database of Cambridge.

A successful example of face recognition using template matching is that based on the eigenface representation [9]. There, a face space is constructed or spanned by a number of eigenfaces [10] derived from a set of training face images by using Karhunen–Loeve transform or the principal component analysis (PCA) [11]. Every prototype face image in the database is represented as a feature point, i.e., a vector of weights, in the space and so is the query face image.

The nearest neighbor (NN) is a simple yet most popular method for classification. In the NN-based classification, the representational capacity of a face database and the error rate depends on how the prototypes are chosen to account for possible variations and also how many prototypes are available. However, it is impractical to exhaust all possibilities—there are an infinite number of them. In practice, only a small number of them are available for a face class, typically from one to about a dozen. It is desirable to have a sufficiently large number of feature points stored to account for as many variations as possible. We want to find a way to generalize the representational capacity of available prototype images.

A. Overview of Present Work

In this paper, we propose a novel method, called the nearest feature line (NFL), for face recognition. The basic assumption is that at least two distinct prototype feature points are available for each class, which is usually satisfied. In a feature space, which is an eigenface space in this study, the NFL method uses a linear model to interpolate and extrapolate each pair of prototype feature points belonging to the same class. More specifically, the two prototype feature points are generalized by the feature line (FL) which is the line passing through the two points. The FL approximates variants of the two prototypes under variations in pose, illumination, and expression, i.e., possible face images derived from the two. It virtually provides an infinite number of prototype feature points of the class. The capacity of the prototype set is thus expanded. The classification is done by using the minimum distance between the feature point of the query and the FL's. The classification result also provides a quantitative position number as a byproduct which can be used to indicate the relative change (in pose, illumination, and expression) between the query face and the two associated faces.

Two sets of experiments are presented to demonstrate advantages of the NFL. The first compares the NFL with the standard eigenface method of Turk and Pentland [9], the latter using the nearest center (NC) criterion. A compound data set from five databases is used: Cambridge, Bern, Yale, Harvard, and our own (see

Manuscript received March 19, 1998; revised November 10, 1998.

The authors are with the School of EEE, Nanyang Technological University, BLK S1, Singapore 639798.

Publisher Item Identifier S 1045-9227(99)02034-2.

http://markov.eee.ntu.ac.sg:8000/˜szli/demos.html for demos of our approach and comparisons). The NFL error rate is about 43.7–65.4% of that of the standard eigenface method. The second set compares the NFL with CNN [8], using the ORL database of Cambridge, the latter work having been reported as yielding the lowest error rate for that database. The result shows that the NFL error rate is about 81% of the CNN error rate.

B. Related Work

The NFL has a close relationship with the linear combination approach [12], the latter being a shape-based approach for recognizing three-dimensional (3-D) objects from two-dimensional (2-D) images. It makes use of a linear combination of two prototypes in a feature space, whereas in [12] a 3-D object is represented by a linear combination of 2-D boundary maps of the object and the knowledge of imaging parameters is not required. An object in the image is classified as belonging to a prototype model object if it can be expressed as a linear combination of the views of the object for some set of coefficients.

A theory of view-based object recognition is presented in [13]. It is based on the observation that the views of a shape-based 3-D rigid object undergoing transformation such as rotation reside in a smooth low-dimensional manifold embedded in the space of coordinates of points attached to the object; and for the object, there exists a smooth transformation function which can map any perspective view into another view of the object. Further, it is also demonstrated that this transformation function can be approximated from a small number of views of the object. The theory is further demonstrated in [14] on a variety of objects, and its application is extended from recognition to categorization. However, object recognition in those studies is based on shape information alone; variations in illumination and texture of objects and nonrigid shape changes, crucial issues for face recognition, are not dealt with.

In [15], a technique is presented to synthesize a new image of an object from a single 2-D view of the object using a linear combination of images of prototype objects of the same class, provided that the object belongs to *linear object classes*. This approach avoids the use of 3-D models for the view synthesis and is capable of generating a new view of a 3-D object from a single 2-D view of the object, using both shape and texture information. The technique requires correspondence between all feature points of prototype images and between the new image and one of the prototypes.

It is proven in [16] that with ideal point light sources, the brightness of a new image at a point can be expressed as a linear combination of the brightness of three prototype images at the corresponding point, when the viewpoint is fixed and the images are subject to variations in illumination only. This suggests that variations in illumination can be compensated for prior to recognition, by finding the underlying linear combination according to the brightness at the corresponding point in the images, expressing the new image as the linear combination of the three images, and then matching along all the nonshadowed corresponding points.

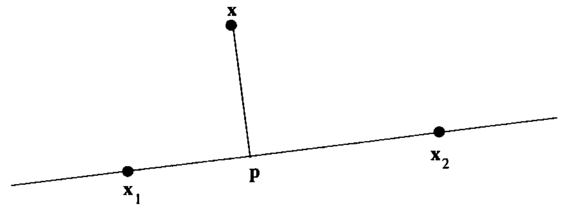

Fig. 1. Generalizing two prototype feature points $\boldsymbol{x}_1$ and $\boldsymbol{x}_2$ by the feature line $\overline{\boldsymbol{x}_1\boldsymbol{x}_2}$. The feature point $\boldsymbol{x}$ of a query face is projected onto the line as point $\boldsymbol{p}$.

The feature line can be considered as a simpler version of the spline type manifold of the parametric appearance representation [17]. There, the appearance manifold of an object is constructed from images of the object taken on a turnable (parameterized by a single parameter) under carefully controlled lighting (parameterized by another single parameter). However, such strictly controlled conditions are difficult to meet in acquiring face images. The NFL provides a simple yet useful solution.

II. The Nearest Feature Line Method

The NFL assumes that at least two prototype feature points are available for each class, which is usually satisfied. It attempts to generalize the representational capacity of available prototypes to cope with various changes by using linear interpolation and extrapolation between the feature points. In the following, we define a new distance measure that will be used in the NFL and describe NFL-based classification.

Eigenfaces [9] are used in the following as the start-point representation. We will not be addressing the issue of optimizing the selection of eigenfaces (in [18], it is pointed out that the best information for faces recognition is not contained in eigenvectors with relatively large eigenvalues but in those with relatively small eigenvalues). Rather, we focus on the issue of classification methods and apply the NFL classification on the conventional eigenface representation.

A. The Feature Line Distance

Consider a variation in the image space from point $\boldsymbol{z}_1$ to $\boldsymbol{z}_2$ and the incurred variation in the feature space (which is an eigenface space, cf. Appendix, in this work) from $\boldsymbol{x}_1$ to $\boldsymbol{x}_2$. The degree of the change may be measured by $\delta\boldsymbol{z} = \|\boldsymbol{z}_2 - \boldsymbol{z}_1\|$ or $\delta\boldsymbol{x} = \|\boldsymbol{x}_2 - \boldsymbol{x}_1\|$. When $\delta\boldsymbol{z} \to \boldsymbol{0}$ and thus $\delta\boldsymbol{x} \to \boldsymbol{0}$, the locus of $\boldsymbol{x}$ due to the change can be approximated well enough by a straight line segment between $\boldsymbol{x}_1$ and $\boldsymbol{x}_2$. Thus any change between the two can be interpolated by a point on the line. A further small change beyond $\boldsymbol{x}_2$ can be extrapolated using the linear model.

The straight line passing through $\boldsymbol{x}_1$ and $\boldsymbol{x}_2$ of the same class, denoted $\overline{\boldsymbol{x}_1\boldsymbol{x}_2}$, is called an FL of that class. The query feature point $\boldsymbol{x}$ is projected onto an FL as point $\boldsymbol{p}$ (Fig. 1). The *FL distance* between $\boldsymbol{x}$ to $\overline{\boldsymbol{x}_1\boldsymbol{x}_2}$ is defined as

$$d(\boldsymbol{x}, \overline{\boldsymbol{x}_1\boldsymbol{x}_2}) = \|\boldsymbol{x} - \boldsymbol{p}\| \tag{1}$$

where $\|\cdot\|$ is some norm.

The projection point can be computed as $\boldsymbol{p} = \boldsymbol{x}_1 + \mu(\boldsymbol{x}_2 - \boldsymbol{x}_1)$ where $\mu \in \mathcal{R}$, called the position parameter, can be calculated from $\boldsymbol{x}$, $\boldsymbol{x}_1$, and $\boldsymbol{x}_2$ as follows: Because $\overline{\boldsymbol{p}\boldsymbol{x}}$ is perpendicular to $\overline{\boldsymbol{x}_2\boldsymbol{x}_1}$, we have $(\boldsymbol{p} - \boldsymbol{x}) \cdot (\boldsymbol{x}_2 - \boldsymbol{x}_1) = [\boldsymbol{x}_2 + \mu(\boldsymbol{x}_2 - \boldsymbol{x}_1) - \boldsymbol{x}] \cdot (\boldsymbol{x}_2 - \boldsymbol{x}_1) = 0$ where "$\cdot$" stands for dot product, and thus $\mu = (\boldsymbol{x} - \boldsymbol{x}_1) \cdot (\boldsymbol{x}_2 - \boldsymbol{x}_1)/(\boldsymbol{x}_2 - \boldsymbol{x}_1) \cdot (\boldsymbol{x}_2 - \boldsymbol{x}_1)$. The parameter μ describes the position of $\boldsymbol{p}$ relative to $\boldsymbol{x}_1$ and $\boldsymbol{x}_2$. When $\mu = 0$, $\boldsymbol{p} = \boldsymbol{x}_1$. When $\mu = 1$, $\boldsymbol{p} = \boldsymbol{x}_2$. When $0 < \mu < 1$, $\boldsymbol{p}$ is an interpolating point between $\boldsymbol{x}_1$ and $\boldsymbol{x}_2$. When $\mu > 1$, $\boldsymbol{p}$ is a forward extrapolating point on the $\boldsymbol{x}_2$ side. When $\mu < 0$, $\boldsymbol{p}$ is a backward extrapolating point on the $\boldsymbol{x}_1$ side.

The FL provides information about linear variants of the two prototypes, i.e., possible face images derived from the two, and virtually provides an infinite number of prototype feature points of the class that the two prototypes belong to. The capacity of the prototype set is thus expanded. Assuming that there are $N_c > 1$ prototype feature points available for class c, a number of $K_c = N_c(N_c - 1)/2$ lines can be constructed to represent the class. For example, $N_c = 5$ feature points are expanded by their $K_c = 10$ feature lines. The total number of feature lines for a number of M classes is $N_{total} = \sum_{c=1}^{M} K_c$.

The locus of the feature point of a face image under a perceivable variation in viewpoint, illumination, or expression, which is highly nonconvex and complex [19], can hardly be precisely described by a straight line in the feature space. To obtain a more accurate description of the variations, one may suggest that a higher-order curve, such as splines [17], be used. This requires 1) that there should be at least three prototype points for every class and 2) that these points should be ordered to account for relative variations described by only one parameter. For the face recognition, requirement 2) is difficult to meet; this is because the parameters describing variations in viewpoint, illumination, and facial expression, if known, are not easily separable for face images taken live and hence the feature points cannot be ordered in terms of a single parameter as in [17]. However, the FL presentation turns out to be quite sufficient for the classification purpose when used with the NFL criterion to be described in the following.

B. NFL-Based Classification

Let $\boldsymbol{x}_i^c$ and $\boldsymbol{x}_j^c$ be two distinct prototype feature points belonging to class c. The FL distance between $\boldsymbol{x}$ of the query and each feature line $\overline{\boldsymbol{x}_i^c \boldsymbol{x}_j^c}$ is calculated for each class c, and each pair $i \neq j$. This yields a number of N_{total} distances. The distances are sorted in ascending order, each being associated with a class identifier, two prototypes, and the corresponding μ value. The *NFL distance* is the first rank distance

$$d(\boldsymbol{x}, \overline{\boldsymbol{x}_{i^*}^{c^*} \boldsymbol{x}_{j^*}^{c^*}}) = \min_{1 \leq c \leq M} \min_{1 \leq i < j \leq N_c} d(\boldsymbol{x}, \overline{\boldsymbol{x}_i^c \boldsymbol{x}_j^c}). \tag{2}$$

The first rank gives the NFL classification composed of the best matched class c^* and the two best matched prototypes i^* and j^* of the class.

The position parameter μ^* of the first rank match, which indicates the position of the projection $\boldsymbol{p}$ of the query relative to $\boldsymbol{x}_{i^*}^{c^*}$ and $\boldsymbol{x}_{j^*}^{c^*}$, can be used to infer the relative position of $\boldsymbol{x}$, as will be illustrated in experiments.

III. Experimental Results

Two sets of experiments are presented to compare our method with the standard eigenface method of Turk and Pentland [9] and with the CNN approach [8], both in terms of the error rate. Demonstrations of NFL and comparisons with various classification methods can be accessed at http://markov.eee.ntu.ac.sg:8000/~szli/demos.html.

A. Comparison with Standard Eigenface Method

A compound data set of 1079 face images of 137 persons is used in this experiment. It is composed of five databases:

1) The Cambridge (ORL) database contains 40 distinct persons, each person having ten different images, taken at different times, varying lighting slightly, facial expressions (open/closed eyes, smiling/nonsmiling), and facial details (glasses/no glasses). All the images are taken against a dark homogeneous background and the persons are in upright, frontal position (with tolerance for some side movement).
2) The Bern database contains frontal views of 30 persons, each person having ten images with slight variations in the head positions (one and two right into the camera, three and four looking to the right, five and six looking to the left, seven and eight downwards, nine and ten upwards).
3) The Yale database contains 15 persons. For each person, ten of its 11 frontal view images are randomly selected. The images are taken under ten different conditions: a normal image under ambient lighting, one with or without glasses, three images taken with different point light sources, and five different facial expressions.
4) Five persons are selected from the Harvard database, each person having ten images which are subject to heavy variations in lighting in which the longitudinal and latitudinal angles of light source direction reach up to 90°.
5) Since most images in the above databases are from Caucasians, we have constructed a database of our own, which is composed of 179 frontal views of 47 Chinese students, each person having three or four images taken at different facial expression, viewpoints, and facial details (glasses/no glasses).

A subset of the compound data set is used as the training set for computing the eigenfaces. It is composed of 544 images: five images per person are randomly chosen from the Cambridge, Bern, Yale, and Harvard databases, and two images per person are randomly chosen from our own database.

Two test schemes are designed to compare the error rate. In scheme 1, the query set is composed of the 535 images that are not used for the training (the compound data set minus the training set). Scheme 2 takes all the 1079 images as the query set; however, when an image is used as the query, it is *not* used as a prototype, i.e., it is removed from the prototype set, during the classification.

The error rates as functions of the number of eigenfaces are given in Fig. 2. Using test scheme 1, the error rate of the proposed method is between 55.6 and 65.4% of that of the

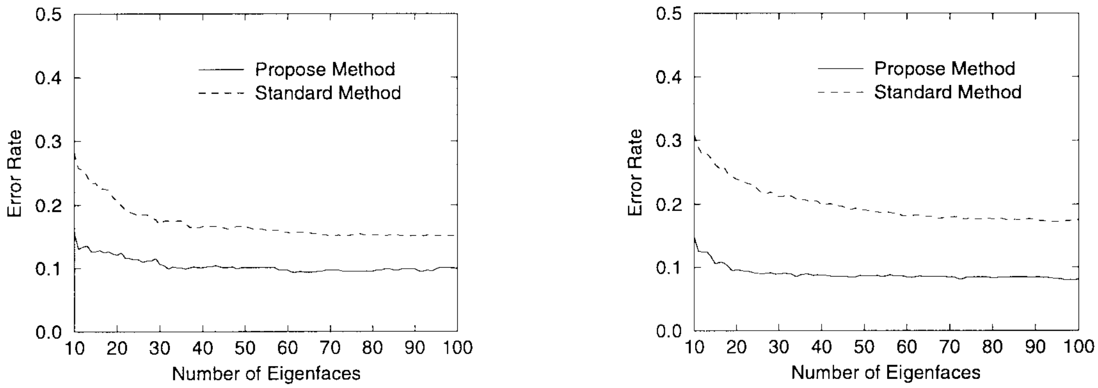

Fig. 2. Comparison of error rates obtained with test scheme 1 (left) and scheme 2.

Fig. 3. Faces under viewpoint variations. The query face (left) is at a center angle relative to the two best recognized faces which are at right and left angles, respectively. The position parameter is calculated as $\mu = 0.234$, suggesting that the query face is an interpolation of the two best recognized faces.

Fig. 4. Faces under illumination variations. The query face (left) is illuminated by a right light as compared to the two best recognized faces which are illuminated by left and center lights, respectively. The position parameter is calculated as $\mu = 1.138$, suggesting that the query face is a forward extrapolation of the two best recognized faces.

standard method. Using test scheme 2, it is between 43.7 and 48.3% of that.

The complexity of our method is $N_c(N_c - 1)$, which is 20 for $N_c = 5$, times that of [9]. Nonetheless, it takes only less than 0.1 s to recognize a face on an HP-9000/770 workstation when 40 eigenfaces are used.

B. Comparison with Convolutional Neural Network

This experiment compares the NFL with the CNN [8] using the ORL face database of Cambridge, CNN having been reported previously as yielding the lowest error rate for that database. The training set and query set are derived in the same way as in [8]: The ten images of each of the 40 persons are randomly partitioned into two sets, resulting in 200 training images and 200 test images, with no overlapping between the two. The NFL error rate is the average of the error rates obtained by four runs (the CNN error rate given in [8] is the average of three runs [20]), each run being performed on a random partition of the database into two sets. The NFL error rate with 40 eigenfaces is 3.125% whereas the CNN error rate is 3.83%. The former is about 81% of the latter and hence the proposed NFL approach updates the record of the lowest error rate.

C. Examples of Recognition Results

Some results of recognition under variations in viewpoint, illumination, and expression are shown in Figs. 3–5. On the left of each figure is the query face, with feature point $\boldsymbol{x}$, and the other two are the two best retrieved faces, with feature points $\boldsymbol{x}_1$ and $\boldsymbol{x}_2$, respectively. Every result is accompanied

Fig. 5. Faces under expression variations. The position parameter is calculated as $\mu = -0.519$, suggesting that the query face (left) is a backward extrapolation of the two best recognized faces.

by the value of the position parameter μ, which indicates how $\boldsymbol{x}$ is projected onto $\overline{\boldsymbol{x}_1\boldsymbol{x}_2}$ as $\boldsymbol{p} = \boldsymbol{x}_1 + \mu(\boldsymbol{x}_2 - \boldsymbol{x}_1)$. The caption illustrates how the parameter can be used to infer the position of $\boldsymbol{x}$ relative to $\boldsymbol{x}_1$ and $\boldsymbol{x}_2$, interpolating or extrapolating.

IV. Conclusions

We have proposed a new method called the NFL for face recognition. The NFL is applicable where there are at least two prototypes for each class. The error rate of the proposed method is 43.7–65.4% of that of the standard eigenface method [9]. The NFL with 40 eigenfaces has achieved the lowest error rate of 3.125% reported to date for the ORL database. The improvement is due to the feature lines' ability to expand the representational capacity of available feature points, and to account for new conditions not represented by original prototype face images.

The NFL turns out to be a general pattern recognition method, regardless of representations, applicable when there are at least two prototypes per class. Our recent research shows that the NFL outperforms the NN also in other applications such as image, texture, and audio classification and retrieval where the representations are totally different from one another (unpublished, see our demo page). We are beginning to develop a theory to justify the NFL concept.

APPENDIX
EIGENFACE FEATURES

The eigenfaces are a set of orthonormal basis vectors computed from a collection of training face images. They provide a basis of low dimensional representation of the face images and are optimal in the sense of minimum mean-square error [9], [10]. Denote the training set of N face images by $\{\boldsymbol{z}_1, \boldsymbol{z}_2, \cdots, \boldsymbol{z}_N\}$. The PCA is applied to the set of training images to find the N eigenvectors of the covariance matrix $1/N \sum_{n=1}^{N} (\boldsymbol{z}_n - \overline{\boldsymbol{z}})(\boldsymbol{z}_n - \overline{\boldsymbol{z}})^T$ where $\overline{\boldsymbol{z}} = 1/N \sum_{n=1}^{N} \boldsymbol{z}_n$ is the average of the ensemble. The eigenvalues of the covariance matrix are calculated.

Let ϕ_k be the eigenvector corresponding to the kth largest eigenvalue. The first N' $(\leq N)$ orthonormal vectors $\phi_1, \cdots, \phi_{N'}$ form a basis of an eigenface space. In [9], it was found that $N' = 40$ is sufficient for a very good description of a set of $N = 115$ face images.

Eigenface-based classification is performed in two stages:

1) eigen-feature extraction. Each training face image $\boldsymbol{z}_n$ is projected into the eigenface space as a point $\boldsymbol{x}_n = \phi^T(\boldsymbol{z}_n - \overline{\boldsymbol{z}})$ where $\phi = [\phi_1, \cdots, \phi_{N'}]$, and is used as a prototype feature point. Given a query face image $\boldsymbol{z}$ to be classified, its projection into the eigenface space is calculated as $\boldsymbol{x} = \phi^T(\boldsymbol{z} - \overline{\boldsymbol{z}})$.
2) Classification based on the eigenfeature vectors. The simplest classification method is based on the Euclidean distance $d(\boldsymbol{x}, \boldsymbol{x}_n)$ using the NN criterion. In [9], the nearest center (NC) criterion is used in which a class is represented by the center of the $\boldsymbol{x}_n$'s belonging to that class, and the classification is based on the distance from $\boldsymbol{x}$ to each class center.

ACKNOWLEDGMENT

The authors would like to thank S. Lawrence and C. Lee Giles for providing details of their experiments on their CNN approach.

REFERENCES

[1] A. Samal and P. A. Iyengar, "Automatic recognition and analysis of human faces and facial expressions: A survey," *Pattern Recognition,* vol. 25, pp. 65–77, 1992.
[2] D. Valentin, H. Abdi, A. J. O. Toole, and G. W. Cottrell, "Connectionist models of face processing: A survey," *Pattern Recognition,* vol. 27, no. 9, pp. 1209–1230, 1994.
[3] R. Chellappa, C. L. Wilson, and S. Sirohey, "Human and machine recognition of faces: A survey," *Proc. IEEE,* vol. 83, pp. 705–740, 1995.
[4] Y. Moses, Y. Adini, and S. Ullman, "Face recognition: The problem of compensating for changes in illumination direction," in *Proc. European Conf. Computer Vision,* 1994, vol. A, pp. 286–296.
[5] A. J. Goldstein, L. D. Harmon, and A. B. Lesk, "Identification of human faces," *Proc. IEEE,* vol. 59, no. 5, pp. 748–760, May 1971.
[6] R. Brunelli and T. Poggio, "Face recognition: Features versus templates," *IEEE Trans. Pattern Anal. Machine Intell.,* vol. 15, pp. 1042–1052, 1993.
[7] I. J. Cox, J. Ghosn, and P. N. Yianilos, "Feature-based face recognition using mixture-distance," in *Proc. IEEE Comput. Soc. Conf. Comput. Vision Pattern Recognition,* 1996, pp. 209–216.
[8] S. Lawrence, C. Lee Giles, A. C. Tsoi, and A. D. Back, "Face recognition: A convolutional neural network approach," *IEEE Trans. Neural Networks,* vol. 8, pp. 98–113, 1997.
[9] M. A. Turk and A. P. Pentland, "Eigenfaces for recognition," *J. Cognitive Neurosci.,* vol. 3, no. 1, pp. 71–86, Mar. 1991.
[10] L. Sirovich and M. Kirby, "Low-dimensional procedure for the characterization of human faces," *J. Opt. Soc. Amer. A,* vol. 4, no. 3, pp. 519–524, Mar. 1987.
[11] K. Fukunaga, *Introduction to Statistical Pattern Recognition,* 2nd ed. Boston, MA: Academic, 1990.
[12] S. Ullman and R. Basri, "Recognition by linear combinations of models," *IEEE Trans. Pattern Anal. Machine Intell.,* vol. 13, pp. 992–1006, 1991.
[13] T. Poggio and S. Edelman, "A network that learns to recognize three-dimensional objects," *Nature,* vol. 343, pp. 263–266, 1990.
[14] S. Edelman and S. Duvdevani-Bar, "A model of visual recognition and categorization," in *Proc. Roy. Soc., London,* 1997, vol. B-352, pp. 1191–1202.
[15] T. Vetter and T. Poggio, "Linear object classes and image synthesis from a single example image," *IEEE Trans. Pattern Anal. Machine Intell.,* vol. 19, pp. 733–742, 1997.
[16] A. Shashua, "Illumination and view position in 3-D visual recognition," in *Neural Inform. Processing Syst.,* J. Moody, S. J. Hanson, and R. L. Lippman, Eds. San Mateo, CA: Morgan Kaufmann, vol. 4, 1992, pp. 404–411.
[17] H. Murase and S. K. Nayar, "Visual learning and recognition of 3-D objects from appearance," *Int. J. Comput. Vision,* vol. 14, pp. 5–24, 1995.
[18] A. J. O'Toole, H. Abdi, K. A. Deffenbacher, and D. Valentin, "Low-dimensional representation of faces in higher dimensions of the face space," *J. Opt. Soc. Amer. A,* vol. 10, pp. 405–410, 1993.
[19] M. Bichsel and A. P. Pentland, "Human face recognition and the face image set's topology," *CVGIP: Image Understanding,* vol. 59, pp. 254–261, 1994.
[20] S. Lawrence and C. Lee Giles, "Convolutional neural network," personal communication, 1998.