Generalizing Capacity of Face Database For Face Recognition

Stan Z. Li Juwei Lu
School of Electrical and Electronic Engineering
Nanyang Technological University, Singapore 639798

szli@szli.eee.ntu.ac.sg

Abstract

A face image can be represented by a point in a feature space such as spanned by a number of eigenfaces. In methods based on nearest neighbor classification, the representational capacity of face database depends on how prototypical face images are chosen to account for possible image variations and also how many prototypical images or their feature points are available.

In this paper, we propose a novel method for generalizing the representational capacity of available face database. Any two feature points of the same class (individual) are generalized by the feature line passing through the points. The feature line covers more of the face space than the feature points and thus expands the capacity of the available database. In the feature line representation, the classification is based on the distance between the feature point of the query image and each of the feature lines of the prototypical images.

Experiments are presented using a data set from five databases: the MIT, Cambridge, Bern, Yale and our own. There are 620 images of 124 individuals subject to varying viewpoint, illumination, and expression. The results show that the error rate of the proposed method is about 55%-60% of that of the standard Eigenface method of Turk and Pentland. They also demonstrate that the recognition result can be used for inferring how the position of the input face relative to the two retrieved faces.

1 Introduction

Face recognition has a wide range of applications in which personal identification is required [1]. Two issues are central: The first is what kind of features to use to represent a face. A face image is subject to changes in viewpoint, illumination, and expression. An effective representation should be able to deal with possible changes. The second issue is how to classify a new face image using the selected representation.

Approaches using the Karhunen-Loeve transform (KLT) or the principle component analysis (PCA) for face representation and recognition have achieved impressive success. In [5], a set of eigenvectors (eigenpictures) are calculated from the covariance matrix of the ensemble of training data. The eigenpictures corresponding to the largest eigenvalues are selected to form the basis of a face space. Any face image can be represented as a point in the eigenpicture space, and reconstructed by a linear combination of the eigenpictures. As the number of eigenpictures increases, the reconstruction quality improves. This work is extended in [2] by adding the mirror images of the original training faces to the ensemble. In this way, the inherent symmetry of faces in the eigenpicture representation is accounted for. The number of prototypical feature points and thus the capacity of the face database are doubled.

The eigenpicture representation has been used as eigenfaces for face recognition [6]. Given the eigenfaces, every face image in the database is represented as a feature point in the face space which is a vector of weights. A query image is also represented in the same way. The nearest distance criterion is used for face recognition.

In distance based classification, the representational capacity of a face database depends on how the prototypes are chosen to account for various changes and also how many prototypes are available. The selection of prototypes and the number of prototypes affect the error rate. In practice, only a small number of them are available, typically from one to about a dozens. It is desirable to generalize the representational capacity of the database using available prototypical images.

In this paper, we propose a novel method, called the feature line method, for generalizing the representational Capacity of face database consisting of prototypical feature points. The eigenface approach is used for feature extraction. A linear model is used to interpolate and extrapolate face images in the face space. Any two feature points of the same class (individual) are generalized by the feature line passing through the points. A feature line covers more of the space than the feature points. It virtually provide an infinite number of feature points, which accounts for more of image changes than the two points. The capacity of the face database is thus expanded. The classification is done by using the minimum distance between the feature point of the query image and that of the prototypical feature line. The result of classification also provide a quantitative position number which indicating the relative change (in pose, illumination and expression) between the query face and the retrieved faces.

Experiments are presented using a data set from five databases: the MIT, Cambridge, Bern, Yale and our own. There are 620 images of 124 individuals subject to varying viewpoint, illumination, and expression. The results show that the error rate of the proposed method is about 55%-60% of that of the standard Eigenface approach of Turk and Pentland [6]. They also demonstrate that the use of the proposed approach for inferring the relative change of the input face with respect to the two retrieved faces.

2 Eigenface Features

The eigenfaces are a set of orthonormal basis vectors for face representation which provide an optimum approximation for a collection of the face images in the sense of minimum mean-square error [5, 6]. Denote the training set of N face images by $\{\mathbf{z}_1, \mathbf{z}_2, \ldots, \mathbf{z}_N\}$. The PCA is applied to the set of training images to find the N eigenvectors of the covariance matrix $\frac{1}{N}\sum_{n=1}^{N}(\mathbf{z}_n - \bar{\mathbf{z}})(\mathbf{z}_n - \bar{\mathbf{z}})^T$ where $\bar{\mathbf{z}} = \frac{1}{N}\sum_{n=1}^{N}\mathbf{z}_n$ is the average of the ensemble. The principal values of the covariance matrix are calculated. Let ϕ_k be the eigenvector corresponding to the k-th largest eigenvalue. A set of N' ($\leq N$) orthonormal vectors $\phi_1, \ldots, \phi_{N'}$ form a basis for the face images. In [6], it was found that $N' = 40$ is sufficient for a very good description of a set of $N = 115$ face images.

In recognition [6], a training face image $\mathbf{z}_n$ is projected into the face space as a point $\mathbf{x}_n = \phi^T(\mathbf{z}_n - \bar{\mathbf{z}})$ where $\phi = [\phi_1, \ldots, \phi_{N'}]$. Given a new face image $\mathbf{z}$ to be classified, its projection into the face space is calculated as $\mathbf{x} = \phi^T(\mathbf{z} - \bar{\mathbf{z}})$. The Euclidean distance $\epsilon_n = \|\mathbf{x} - \mathbf{x}_n\|$ is calculated for each $\mathbf{x}_n$. The simplest classification method is based on the nearest distance criterion. If $\epsilon_c = \min_n \epsilon_n$, the face in $\mathbf{z}$ is classified as belonging to the class that $\mathbf{x}_c$ represents. In [6], a face class is represented by the center of the $\mathbf{x}_n$'s (as few as one) belonging to that class, and the classification is based on comparing $\mathbf{x}$ to each of the class center.

3 The Feature Line Method

In the nearest neighbor classification, the representational capacity and the error rate depends on how the prototypical images are chosen to account for possible image variations and also the number of prototypical images. It is desirable to have a sufficient number of feature points stored to account for as many changes as possible. However, it is impractical to exhaust all possibilities – there are an infinite number of them. In practice, only a few, typically one to a couple of dozens are available. Our method is aimed to generalize the representational capacity of available prototypes to cope with various changes.

3.1 Representation

Consider a change from $\mathbf{z}_1$ to $\mathbf{z}_2$ and the incurred change from the eigenface features $\mathbf{x}_1$ to $\mathbf{x}_2$. The size of the change may be measured by the variations $\delta\mathbf{z} = \|\mathbf{z}_2 - \mathbf{z}_1\|$ or $\delta\mathbf{x} = \|\mathbf{x}_2 - \mathbf{x}_1\|$. When $\delta\mathbf{z} \to \mathbf{0}$ and thus $\delta\mathbf{x} \to \mathbf{0}$, the locus of $\mathbf{x}$ due to the change can be approximated well enough by a straight line segment between $\mathbf{x}_1$ and $\mathbf{x}_2$. Thus any change between the two can be interpolated by a point on the line. A further small change beyond $\mathbf{x}_2$ can be extrapolated using the linear model. The straight line passing through $\mathbf{x}_1$ and $\mathbf{x}_2$ of the same class, denoted $\overline{\mathbf{x}_1\mathbf{x}_2}$, is called a feature line of that class.

The feature point $\mathbf{y}$ of a query face is projected onto the line as point $\mathbf{p}$ as illustrated in Fig.1. The projection point can be represented as $\mathbf{p} = \mathbf{x}_1 + \mu(\mathbf{x}_2 - \mathbf{x}_1)$ where $\mu \in \mathcal{R}$, called the position parameter (to be explained later) can be calculated from $\mathbf{y}, \mathbf{x}_1$ and $\mathbf{x}_2$ as follows: Because $\mathbf{p} - \mathbf{y}$ is perpendicular to $\mathbf{x}_2 - \mathbf{x}_1$, we have $(\mathbf{p} - \mathbf{y}) \cdot (\mathbf{x}_2 - \mathbf{x}_1) = [\mathbf{x}_2 + \mu(\mathbf{x}_2 - \mathbf{x}_1) - \mathbf{y}] \cdot (\mathbf{x}_2 - \mathbf{x}_1) = 0$ where "$\cdot$" stands for dot product, and thus $\mu = \frac{(\mathbf{y}-\mathbf{x}_1)\cdot(\mathbf{x}_2-\mathbf{x}_1)}{(\mathbf{x}_2-\mathbf{x}_1)\cdot(\mathbf{x}_2-\mathbf{x}_1)}$.

The position parameter μ describes the position of $\mathbf{p}$ relative to $\mathbf{x}_1$ and $\mathbf{x}_2$. When $\mu = 0$, $\mathbf{p} = \mathbf{x}_1$. When $\mu = 1$,

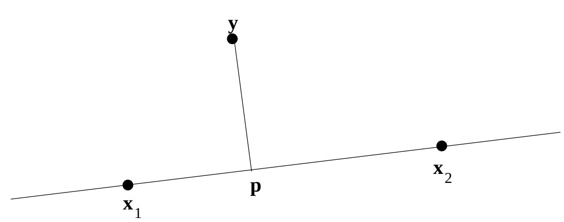

Figure 1. Generalizing two prototypical feature points $\mathbf{x}_1$ and $\mathbf{x}_2$ by the feature line $\overline{\mathbf{x}_1\mathbf{x}_2}$. The feature point $\mathbf{y}$ of a query face is projected onto the line as point $\mathbf{p}$.

$\mathbf{p} = \mathbf{x}_2$. When $0 < \mu < 1$, $\mathbf{p}$ is an interpolating point between $\mathbf{x}_1$ and $\mathbf{x}_2$. When $\mu > 1$, $\mathbf{p}$ is a forward extrapolating point on the $\mathbf{x}_2$ side. When $\mu < 0$, $\mathbf{p}$ is a backward extrapolating point on the $\mathbf{x}_1$ side.

It is proven [4] 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 prototypical images at the corresponding point, when the viewpoint is fixed and the images are subject to changes in illumination only. This suggests that changes 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 non-shadowed corresponding points.

The locus of the feature point of a face image under a perceivable change in viewpoint, illumination or expression can hardly be precisely described by a straight line in the feature space. From face reconstruction experiments, we found that the proposed feature line representation well interpolated and extrapolated face images over variations in illumination. However, the results were not as satisfactory for expression and especially viewpoint changes. This is because that such a variation violates more the assumption that the variation should be small. Therefore, the linear interpolation or extrapolation of two face images are not recommended for the reconstruction purpose.

To overcome this problem, one may suggest that a higher order curve such as based on splines be used. This requires (i) that there are at least three prototypical points for every class, and (ii) that these points are ordered to account for relative changes described by only one parameter. This suggested method has been successfully used in industry inspection [3] where rigid object models are imaged on a turnable (parameterized by a single parameter) under carefully controlled lighting (parameterized by another single parameter). For the face recognition, requirement (ii) may be difficult to meet; this is because the parameters describing changes 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.

3.2 Recognition

The feature line method is well suited for face recognition using the following recognition strategy. Assume that there are N_c feature points available for class c, a number of $K_c = \frac{N_c(N_c-1)}{2}$ lines can be constructed to represent the class. For example, $N_c = 5$ feature points are expanded to $K_c = 10$ feature lines. The total number of feature lines for a number of M face classes is $N_{total} = \sum_{c=1}^{M} K_c$.

Denote by $\mathbf{x}_i^c$ and $\mathbf{x}_j^c$ the prototypical feature points of two face images belonging to the same class c. The distance between the point $\mathbf{x}$ of the query face and the feature line $\overline{\mathbf{x}_i^c\mathbf{x}_j^c}$ is calculated for each class c. This yields a number of N_{total} distances denoted as $\{d(\mathbf{x}, \overline{\mathbf{x}_i^c\mathbf{x}_j^c}) \mid 1 \leq i, j \leq N_c,\ i \neq j,\ 1 \leq c \leq M\}$. The face represented by $\mathbf{x}$ is classified as belonging to class c^* if $d(\mathbf{x}, \overline{\mathbf{x}_i^{c^*}\mathbf{x}_j^{c^*}}) = \min_{1 \leq c \leq M} \min_{1 \leq i,j \leq N_c} d(\mathbf{x}, \overline{\mathbf{x}_i^c\mathbf{x}_j^c})$.

The recognized class c^*, the distance $d^* = d(\mathbf{x}, \overline{\mathbf{x}_i^{c^*}\mathbf{x}_j^{c^*}})$, the associated position parameter μ, and the two face images corresponding to $\mathbf{x}_i^{c^*}$ and $\mathbf{x}_j^{c^*}$ are given as the recognition result. Because d^* is usually relatively small, the position parameter, which indicates the position of the projection $\mathbf{p}$ relative to $\mathbf{x}_i^{c^*}$ and $\mathbf{x}_j^{c^*}$, can be used to infer the relative position of $\mathbf{x}$, as will be illustrated in experiments.

The recognition results are ranked by the distance values. Currently, the value of μ is not used for recognition and there is no limit on it for the recognition purpose.

A feature line covers more of the space than the feature points constructing the line. It virtually provide an infinite number of feature points, which accounts for more of image changes than the original two points and thus expands the representational capacity of the feature points. When the number of training data is doubled by adding the mirror images as in [2], the number of feature lines is quadrupled, further extending the capacity.

4 Experimental Results

Experimental results are presented as follows. The training set contains a total number of 700 images of 140 subjects (5 images each, *i.e.* $N_c = 5$ for c) from five databases: 40 subjects from the Cambridge database, 30 subject from the Bern database, 15 subjects from the Yale database, 13 subjects from the MIT database, 5 subjects from the Harvard database, and 37 subjects from our own database. The query set contains 250 images of 50 subjects (5 images for each subject) from the Cambridge, Bern, Harvard and Yale

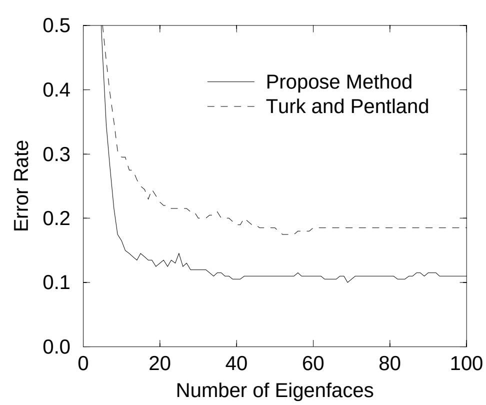

Figure 2. Comparison of error rates.

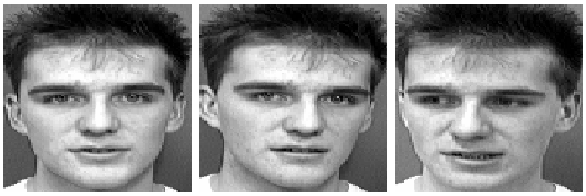

Figure 3. Data set 1: faces under viewpoint changes. The query face is at a center angle relative to the two retrieved faces which are at right and left angles, respectively. The position parameter is calculated as $\mu = 0.234$, indicating that the query face is an interpolation of the two retrieved faces.

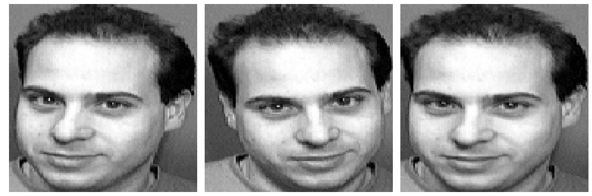

Figure 4. Data set 2: faces under viewpoint changes. The query face is at a left angle relative to the two retrieved faces which are at right and center angles, respectively. The position parameter is calculated as $\mu = 1.233$, indicating that the query face is a forward extrapolation of the two retrieved faces.

databases. Both training and query sets includes variations in viewpoint, illumination, and expression, race and gender.

The performance of the proposed feature line method is compared relative to that the feature point based method of [6]. For both, eigenfaces are computed using the same set of face databases and the same segmentation results. The error rates as functions of the number of eigenfaces are shown in Fig.2. Using 40 eigenfaces, the error rate of the proposed method is 10.5% (23 out of 220) while that of [6] is 19.1% (42 out of 220). Overall, the former is between 55% to 60% of the latter.

Some results of recognition under changes in viewpoint, illumination, and expression are shown in Figs.3–8. On the left of each figure is the query face, with feature point $\mathbf{x}$, and the other two are the two best retrieved faces, with feature points $\mathbf{x}_1$ and $\mathbf{x}_2$ respectively. Every result is accompanied by the value of the position parameter μ, which indicates how $\mathbf{x}$ is projected onto $\overline{\mathbf{x}_1\mathbf{x}_2}$ as $\mathbf{p} = \mathbf{x}_1 + \mu(\mathbf{x}_2 - \mathbf{x}_1)$. The caption illustrates how the parameter can be used to infer the position of $\mathbf{x}$ relative to $\mathbf{x}_1$ and $\mathbf{x}_2$, interpolating or extrapolating.

The complexity of our method is $N_c(N_c - 1)$, which is 20 for $N_c = 5$, times that of [6]. Nonetheless, it takes only less than 0.1 second to recognize a face on an HP-9000/770 workstation when 40 eigenfaces is used.

5 Conclusions

The experimental results show that the use of the feature line representation as opposed to the feature point representation improves the error rate for face recognition to a significant extent. This is due to that the feature lines expand the representational capacity of available feature points in face database. The obtained position parameter provides for inferring the relative position of the query face with respect to the two best retrieved faces.

References

[1] R. Chellappa, C. Wilson, and S. Sirohey. "Human and machine recognition of faces: A survey". *P-IEEE*, 83:705–740, 1995.

[2] M. Kirby and L. Sirovich. "Application of the Karhunen-Loeve procedure for the characterization of human faces". *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 12(1):103–108, January 1990.

[3] H. Murase and S. Nayar. "Visual learning and recognition of 3-D objects from appearance". *International Journal of Computer Vision*, 14:5–24, 1995.

[4] A. Shashua. "Illumination and view position in 3D visual recognition". In J. Moody, S. J. Hanson, and R. L. Lippman, editors, *Neural Information Processing Systems*, volume 4, pages 404–411, San Mateo, CA, 1992. Morgan Kaufmann.

[5] L. Sirovich and M. Kirby. "Low-dimensional procedure for the characterization of human faces". *Journal of the Optical Society of America A*, 4(3):519–524, March 1987.

[6] M. A. Turk and A. P. Pentland. "Eigenfaces for recognition". *Journal of Cognitive Neuroscience*, 3(1):71–86, March 1991.

Figure 5. Data set 3: faces under illumination changes. The query face is illuminated by a center light as compared to the two retrieved faces which are illuminated by left and right lights, respectively. The position parameter is calculated as $\mu = 0.452$, indicating that the query face is an interpolation of the two retrieved faces.

Figure 6. Data set 4: faces under illumination changes. The query face is illuminated by a right light as compared to the two retrieved faces which are illuminated by left and center lights, respectively. The position parameter is calculated as $\mu = 1.138$, indicating that the query face is a forward extrapolation of the two retrieved faces.

Figure 7. Data set 5: faces under expression changes. The position parameter is calculated as $\mu = 0.648$, indicating that the query face is an interpolation of the two retrieved faces.

Figure 8. Data set 6: faces under expression changes. The position parameter is calculated as $\mu = -0.519$, indicating that the query face is a backward extrapolation of the two retrieved faces.