Hierarchical Linear Combinations for Face Recognition

Stan Z Li Juwei Lu Kap Luk Chan Jun Liu Lei Wang
School of EEE, NTU, Singapore 639798
szli@szli.eee.ntu.ac.sg

Abstract

A hierarchical representation consisting of two levels linear combinations (LC) is proposed for face recognition. At the first level, a face image is represented as a linear combination (LC) of a set of basis vectors, i.e. *eigenfaces. Thereby a face image corresponds to a feature vector (prototype) in the eigenface space. Normally several such prototypes are available for a face class, each representing the face under a particular condition such as in viewpoint, illumination, and so on. We propose to use the second level LC, that of the prototypes belonging to the same face class, to treat the prototypes coherently. The purpose is to improve face recognition under a new condition not captured by the prototypes by using a linear combination of them. A new distance measure called nearest LC (NLC) is proposed, as opposed to the NN. Experiments show that our method yields significantly better results than the one level eigenface methods.*

1 Introduction

There are two broad types of representations: geometric, feature-based and templet-based [1]. When measured reliably and accurately, geometrical facial features provide powerful constraints for face recognition to produce good results. However, automated detection of facial feature landmarks and measurement of the positions remains a challenge that has yet been well solved, and manual interaction have been needed. Templet-based representation is computationally more stable. Usually one or more prototypical templets are available for a class, and the matching is performed by comparing the query image with all or a subset of the templets. A crucial assumption made in templet matching is that the prototypes are representative of query images under various conditions. The location and scale may be achieved by location and scale normalization. Multiple templets per person are normally used to represent changes in illumination and rotation because such changes are difficult to compensate by normalization operations.

To achieve data reduction, a raw image is represented by a feature vector in a feature space. The Karhunen-Loeve (K-L) transform or principal component analysis (PCA) [2] provides the optimal reduction in the least squares sense. In [3], a set of eigenvectors are calculated from a set of prototypical (training) face images to form a basis, and any face image in the prototypical set is represented and reconstructed by a linear combination (LC) of the basis vectors. In the eigenface approach [4], which has been successfully used for face recognition, a face image, either query or prototypical, is represented by a point in the eigenface space, and a distance-based criterion is used for face recognition.

The nearest neighbor (NN) used in many classification algorithms relies crucially on the assumption that the prototypes are representative of query images. The prototypes are treated individually. The results depend on how prototypes are chosen to account for possible image variations and also how many prototypes are available. No matter how representative the prototypes may be, there are always un-prototyped viewing, expression and illumination conditions, because only a finite, often small, number of prototypes are available as compared to all possibilities. An approach is needed to cope with the missing conditions.

We introduce the concept of hierarchical LC representation for face recognition. In this representation, the first level LC if of eigenfaces, which has been used in most eigenface based methods, and the second level of feature vectors (prototypes) in the eigenface space. The rational of the second level is that while a prototype accounts for the face imaged under a certain condition, the face under a new condition may be approximated by linear interpolation or extrapolation of several prototypes of the same face. Variations in lighting, viewing angle and expression among the prototypical face images are accounted for by variations in the weights that determine the second level LC. This leads to a new patter recognition approach called the nearest linear combination (NLC), as opposed to the conventional NN. Experimental results show that the error rate of the proposed method is about 40%-45% of that of the standard Eigenface method. See http://markov.eee.ntu.ac.sg:8000/ szli/demos.html .

2 Hierarchical Linear Combinations

In the linear combination approach for recognizing 3D objects from 2D images [5], a 3D object is represented by a linear combination of 2D boundary maps of the object and the knowledge of imaging parameters is not required. An object in the image is considered as a prototype of the model object if it can be expressed as a linear combination of the model views for some set of coefficients. In another work [6], a linear combination of 3D prototypical views are used to synthesize new views of an object. Here in this paper, the LCs are of eigenfaces and feature vectors in the eigenface space.

Let a space be spanned by N basis vectors $\{\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_N\}$. Any vector in the space can be represented as a *linear combination* (LC) of the basis vectors

$$\mathbf{x} = \mathbf{x}(\mathbf{A}) = \sum_{k=1}^{m} a_k \mathbf{x}_k = \mathbf{A}^T \mathbf{X} \tag{1}$$

where $\mathbf{A}^T = (a_1, \ldots, a_m)$ is the vector of weights. A *constrained linear combination* (CLC) is defined as a linear combination subject to a constraint $C(\mathbf{A}) = 1$. Under the linear constraint $C(\mathbf{A}) = \sum_k a_k = 1$, only $m-1$ of $\{\mathbf{x}_1, \ldots, \mathbf{x}_m\}$ are free. Let the first $m-1$ be free. Then the last can be derived as $a_m = 1 - \sum_{k=1}^{m-1} a_k$ and $\mathbf{x} = \mathbf{x}_m + \sum_{k=1}^{m-1} a_k \mathbf{x}'_k = \mathbf{x}_m + \mathbf{A}'^T \mathbf{X}'$ where $\mathbf{x}'_k = \mathbf{x}_k - \mathbf{x}_m$, $\mathbf{X}' = (\mathbf{x}'_1, \ldots, \mathbf{x}'_{m-1})$, and $\mathbf{A}'^T = (a_1, \ldots, a_{m-1})$.

Denote a training set of N face images by $\{\mathbf{z}_1, \mathbf{z}_2, \ldots, \mathbf{z}_N\}$. The *first level LC* is aimed to approximate a training image by using a LC of $D < N$ basis vectors, or eigenfaces:

$$\mathbf{z}_n \approx \sum_{k=1}^{D} a_k^n \mathbf{x}_k \tag{2}$$

such that the total squared error $\sum_{n=1}^{N} (\mathbf{z}_n - \sum_{k=1}^{D} a_k^n \mathbf{x}_k)^2$ is minimized. Let $\{\mathbf{x}_1, \ldots, \mathbf{x}_N\}$ be the N eigenvectors of the covariance matrix $\mathbf{C}$, ordered in the descending eigenvalues, where $\mathbf{C} = \frac{1}{N} \sum_{n=1}^{N} (\mathbf{z}_n - \bar{\mathbf{z}})(\mathbf{z}_n - \bar{\mathbf{z}})^T$ and $\bar{\mathbf{z}} = \frac{1}{N} \sum_{n=1}^{N} \mathbf{z}_n$ is the mean. According to the Karhunen-Loeve (K-L) transform or principal component analysis (PCA) [2], such a basis can be constructed by using the set of first D eigenvectors.

In the eigenface approach [4], face recognition is done based on the feature vectors. Given a new face image $\mathbf{y}$ to be classified, the projection is calculated as $\mathbf{p} = \mathbf{X}^T(\mathbf{z} - \bar{\mathbf{z}})$. The classification is carried out by comparing $\mathbf{p}$ to the prototypes $\mathbf{A}$'s. Two classification criteria are the nearest center (NC) and the nearest neighbor (NN).

While the purpose of the first level LC is to achieve data reduction of face images, the *second level LC* is motivated as follows. In the NN classification, the nearest prototype is chosen to represent the class that the query belongs to. During the process, each prototype is treated individually and independently. Our second level LC linearly combines the prototypes to reach a new prototype for classifying the query. The intuition the use of the second level LC to infer the query is illustrated below.

Denote a face image by $\mathbf{x}$. Consider a change from $\mathbf{x}_1$ to $\mathbf{x}_2$ in the image space. When the size of the change, which may be measured by the variations $\delta\mathbf{x} = \|\mathbf{x}_2 - \mathbf{x}_1\|$, is small, *i.e.* $\delta\mathbf{x} \to \mathbf{0}$, the change can be approximated well enough by a linear interpolation or extrapolation, and the locus of the changing $\mathbf{x}$ is a straight line passing through the two images corresponding two points in the image space. This can be expressed as $\mathbf{x} = a_1\mathbf{x}_1 + a_2\mathbf{x}_2$ with $a_1 + a_2 = 1$. When the change is not small enough, it may be more sensible to approximate the image by $\mathbf{x} \approx= \mathbf{p} = a_1\mathbf{x}_1 + a_2\mathbf{x}_2$ where $\mathbf{p}$ is the projection of $\mathbf{x}$ onto the line. The projection point, referred to below as the nearest LC, is the new prototype generated for representing the class in classifying $\mathbf{x}$.

Assume that there are C face classes and a set of N_c prototypes are available for class c, denoted $\{\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_{N_c}\}$ (corresponding to the $\mathbf{A}_n$'s at level one for the same class). The prototypes are treated *dependently* during the classification by representing a class as a LC described by Eq.1, where $m = N_c$ for class c, $\mathbf{X} = (\mathbf{x}_1, \ldots, \mathbf{x}_m)$ is formed by the m prototypical vectors and $\mathbf{A}^T = (a_1, \ldots, a_m)$ is the vector of weights. Only the within-class LCs are used to produce samples of that class; cross-class combinations are not considered. Such an LC interpolates or extrapolates the prototypes $\{\mathbf{x}_k\}$, cases for $m = 2$ having been illustrated previously. The following uses least squares method to determine the second level LC's weights $\mathbf{A}$.

Let the eigenface feature vector of the query image be $\mathbf{y}$. Its Euclidean distance to a linear combination is $e(\mathbf{A}) = \|\mathbf{y} - \mathbf{x}(\mathbf{A})\|$ It depends on the weights $\mathbf{A}$ when $\mathbf{y}$ and $\mathbf{X}$ are given. We define the *nearest linear combination* (NLC) of the m points $(\mathbf{x}_1, \ldots, \mathbf{x}_m)$ for $\mathbf{y}$, without the constraint $C(\mathbf{A}) = 1$, as the linear combination that minimizes $e(\mathbf{A})$. This is a least squares problem and the NLC weights can be calculated by using $\mathbf{y}$ and the pseudo-inverse $\mathbf{X}^+$ $\mathbf{A}^* = \arg\min_{\mathbf{A}} e(\mathbf{A}) = \mathbf{y}\mathbf{X}^+$ The NLC, $\mathbf{x}(\mathbf{A}^*)$, is the projection of $\mathbf{y}$ onto the (unconstrained) linear combination space. The projection point is

$$\mathbf{p} = \mathbf{p}(\mathbf{y}; \mathbf{x}_1, \ldots, \mathbf{x}_m) = \mathbf{x}(\mathbf{A}^*) = \sum_{k=1}^{m} a_k^* \mathbf{x}_k \tag{3}$$

When the linear constraint $C(\mathbf{A}) = 1$ is imposed, we have the nearest constrained LC (NCLC). The constrained optimization can be converted to an unconstrained one determined by $m-1$ of the m weights. The minimal weight vector for NCLC is $\mathbf{A}^* = \mathbf{y}'\mathbf{X}'^+$ where $\mathbf{y}' = \mathbf{y} - \mathbf{x}_m$ and $\mathbf{X}'^+$ is the pseudo-inverse of $\mathbf{X}'$. The NCLC can be calculated by using the same formula (3) with $a_m^* = 1 - \sum_{k=1}^{m-1} a_k^*$.

The NLC is the projection of $\mathbf{y}$ onto the subspace spanned by $\mathbf{x}_1, \ldots, \mathbf{x}_m$ whereas the NCLC is the projection onto the subspace spanned by $\mathbf{x}_1, \ldots, \mathbf{x}_{m-1}$. The recognition is done by choosing the smallest NLC or NCLC distance.

3 Experiments

Two sets of experiments were done. The first set compared four classification methods and the second compares the NCLC itself with varying parameter $m = 2, 3, 4, 5$. The eigenfaces are used as the bottom level feature representation. Four methods are compared: (i) the nearest center used in the standard eigenface method [4], as the baseline performance; (ii) the nearest neighborhood (NN) method; (iii) the nearest linear combination (NLC), with $m = N_c = 5$; and (iv) the nearest constrained linear combination (NCLC), also with $m = N_c = 5$.

The data contains the images which are subject to variations in viewpoint, illumination, and expression, race and gender. The faces are located by using a simple algorithm with manual interaction, which does not align the faces accurately but provide a basis for relative comparisons. The prototype set contains a total number of 600 images of 120 individuals (5 images each, *i.e.* $N_c = 5$) from four databases: 40 individuals from the Cambridge database, 30 individual from the Bern database, 13 individuals from the MIT database, and 37 individuals from our own database. The query set contains a total number of 200 images of 40 individuals from the Cambridge database.

For the first set of experiments, the error-rates of the four compared methods are plotted on the left of Fig.1 as functions of the number of eigenfaces. We see that the methods can be ordered in descending error-rates as NC, NN, NLC, NCLC. NLC and NCLC are significantly better than NC and NN. When D is between 30 and 40, the error rate of NCLC is about 45% of that of NC for the first data set and; it is about 60% of that of NN. The results suggeststhat NLC and NCLC are better classification algorithms than NC and NN, and between NLC and NCLC, NCLC is more preferable. Therefore the use of NCLC is recommended. We would also like to mention that k-NN with $k = m$ yields results even not as good as NN.

The second set of experiments compares the error-rates within NCLC as m increases from 2 to 5. The result is shown one the right of Fig.1. We see that the differences in error-rates between different m settings are not significant when D is between 30 and 40. As commonly reported, the best classification is achieved with D around 40, suggesting the use of $D = 40$. With $D = 40$, the use of $m = N_c$ is recommended for the computational convenience because

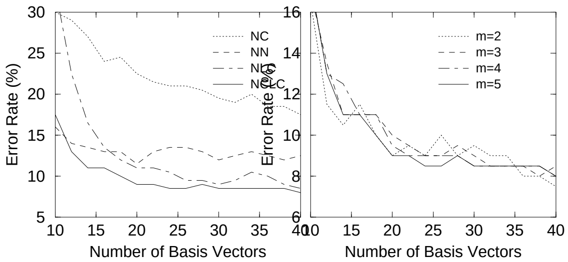

Figure 1. Comparison of error rates of the four methods (left), and comparison of error rates of NCLC with varying combination numbers m.

in this case, only one combination need to be computed.

4 Conclusions

A hierarchical linear combination representation is proposed for face recognition. It is shown to significantly reduce the error rates of the standard and NN classification approaches in eigenface based face recognition. The improvement is due to that the second level LC expands the representational capacity of available prototypes in face database: Variations in lighting, viewing angle and expression among the prototypical face images are accounted for by variations in the weights that determine linear combinations.

Currently, the feature vectors used in the paper are those in an eigenface space. The approach can be extended naturally to incorporate any other types of feature vectors, as long as there are more than one prototypes for some classes and where a distance based criterion is suitable for classification.

Acknowledgments This work was supported by NTU AcRF projects RG 43/95 and RG 51/97.

References

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

[2] K. Fukunaga. *Introduction to statistical pattern recognition*. Academic Press, Boston, 2 edition, 1990.

[3] 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.

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

[5] S. Ullman and R. Basri. "Recognition by linear combinations of models". *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 13:992–1006, 1991.

[6] T. Vetter and T. Poggio. "Linear object classes and image synthesis from a single example image". *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 19(7):733–742, 1997.