Face Recognition with One Sample Image per Class

Shaokang Chen
Intelligent Real-Time Imaging and Sensing (IRIS) Group
The University of Queensland
Brisbane, Queensland, Australia
shaokang@itee.uq.edu.au

Brian C. Lovell
Intelligent Real-Time Imaging and Sensing (IRIS) Group
The University of Queensland
Brisbane, Queensland, Australia
lovell@itee.uq.edu.au

Abstract

There are two main approaches for face recognition with variations in lighting conditions. One is to represent images with features that are insensitive to illumination in the first place. The other main approach is to construct a linear subspace for every class under the different lighting conditions. Both of these techniques are successfully applied to some extent in face recognition, but it is hard to extend them for recognition with variant facial expressions. It is observed that features insensitive to illumination are highly sensitive to expression variations, which result in face recognition with changes in both lighting conditions and expressions a difficult task. We propose a new method called Affine Principle Components Analysis in an attempt to solve both of these problems. This method extract features to construct a subspace for face representation and warps this space to achieve better class separation. The proposed technique is evaluated using face databases with both variable lighting and facial expressions. We achieve more than 90% accuracy for face recognition by using only one sample image per class.

1. Introduction

One of the difficulties in face recognition (FR) is the numerous variations between images of the same face due to changes in lighting conditions, view points or facial expressions. A good face recognition system should recognize faces and be immune to these variations as mush as possible. Yet, it is been reported in [19] that differences between images of the same face due to these variations are normally greater than those between different faces. Therefore, most of the systems designed to date can only deal with face images taken under constrained conditions. So these major problems must be overcome in the quest to produce robust face recognition systems.

In the past few years, different approaches have been proposed for face recognition to reduce the impact of these nuisance factors. Two main approaches are used for illumination invariant recognition. One is to represent images with features that are less sensitive to illumination changes such as the edge maps of the image. But edges generated from shadows are related to illumination changes and may have an impact on recognition. Experiments in [19] show that even with the best image representations using illumination insensitive features and distance measurement, the misclassification rate is more then 20%. The second approach presented in [21] and [22], is to prove that images of convex Lambertian objects under different lighting conditions can be approximated by a low dimensional linear subspaces. Kreigman, Belhumeur and Georghiades proposed an appearance-based method in [7] for recognizing faces under variations in lighting and view point based on this concept. Nevertheless, these methods all suppose the surface reflectance of human faces is Lambertian reflectance and it is hard for these systems to deal with cast shadows. Furthermore, these systems need several images of the same face taken under different lighting source directions to construct a model of a given face. However, sometimes it is hard to obtain different images of a given face under specific conditions.

As for expression invariant recognition, it is still unsolved for machine recognition and is even a difficult task for humans. In [23] and [24], images are morphed to be the same shape as the one used for training. But it is not guaranteed that all images can be morphed correctly, for example an image with closed eyes cannot be morphed to a neutral image because of the lack of texture inside the eyes. It is also hard to learn the local motions within the feature space to determine the expression changes of each face, since the way one person express a certain emotion is normally somewhat different from others. Martinez proposed a method to deal with variations in facial expressions in [20]. An image is divided into several local areas and those that are less sensitive to expressional changes are chosen and weighed independently. But features that are insensitive to expression changes may be sensitive to illumination variation. This is discussed in [19] which says that "when a given representation is sufficient to overcome a single image variation, it may still be affected by other processing stages that control other imaging parameters".

It is known that performance of face recognition systems is acutely dependent on the choice of features [3], which is thus the key step in the recognition methodology. Principal Component Analysis (PCA) and Fisher Linear Discriminant (FLD) [1] are two well-known statistical feature extraction techniques for face recognition. PCA, a standard decorrelation technique, derives an orthogonal projection basis, which allows representation of faces in a vastly reduced feature space — this dimensionality reduction increases generalisation ability. PCA finds a set of orthogonal features, which provide a maximally compact representation of the majority of the variation of the facial data. But PCA might extract some noise features that degenerate performance of the system. For this reason, Swets and Weng [8] argue in favor of methods such as FLD which seek to determine the most discriminatory features by taking into account both within-class and between-class variation to derive the Most Discriminating Features (MDF). However, compared to PCA, it has been shown that FLD overfits to the training data resulting in a lack of generalization ability [2].

We propose a new method Affine Principle Component Analysis (APCA) that can deal with variations both in illumination and facial expression. This paper discusses APCA and presents results, which show that the recognition performance of APCA greatly exceeds that of both PCA and FLD when recognizing known faces with unknown changes in illumination and expression.

2. Review of PCA & FLD

PCA and FLD are two popular techniques for face recognition. They abstract features from training face images to generate orthogonal sets of feature vectors, which span a subspace of the face images. Recognition is then performed within this space based on some distance metric (possibly Euclidean).

2.1. PCA (Principal Component Analysis)

PCA is a second-order method for finding the linear representation of faces using only the covariance of data and determines the set of orthogonal components (feature vectors) which minimise the reconstruction error for a given number of feature vectors. Consider the face image set $I = [I_1, I_2, \dots I_n]$, where I_i is a p×q image, $i \in [1..n]$, $p,q,n \in Z^+$, the average face Ψ of the image set is defined by:

$$\Psi = \frac{1}{n}\sum_{i=1}^{n} I_i \, . \quad (1)$$

Normalizing each image by subtracting the average face, we have the normalized difference image:

$$\tilde{D}_i = I_i - \Psi \, . \quad (2)$$

Unpacking $\tilde{D}_i$ row-wise, we form the N ($N = p \times q$) dimensional column vector d_i. We define the covariance matrix C of the normalized image set $D = [d_1, d_2, \dots d_n]$ by:

$$C = \sum_{i=1}^{n} d_i d_i^T = DD^T \quad (3)$$

An eigendecomposition of C yields eigenvalues λ_i and eigenvectors u_i which satisfy:

$$Cu_i = \lambda_i u_i \, , \quad (4)$$

$$DD^T = C = \sum_{i=1}^{N} \lambda_i u_i u_i^T \, , \quad (5)$$

where $i \in [1..N]$. Since those eigenvectors obtained looks like human faces physically, they are also called eigenfaces. Generally, we select a small subset of $m < n$ eigenvectors, to define a reduced dimensionality facespace that yields higher recognition performance on unseen examples of faces. Choosing $m = 10$ or thereabouts seems to yield good performance in practice. Although PCA defines a face subspace that contains the greatest covariance, it is not necessarily the best choice for classification since it may retain principle components with large noise and nuisance factors [2].

2.2 FLD (Fisher Linear Discriminant)

FLD finds the optimum projection for classification of the training data by simultaneously diagonalizing the within-class and between-class scatter matrices [2]. The FLD procedure consists of two operations: whitening and diagonalization [2]. Given M classes S_j, $j \in [1...M]$, we denote the exemplars of each class by $s_{j,k} = [s_{j,1}, s_{j,2}, \cdots, s_{j,K_j}]$ where K_j is the number of exemplars in class j. Let μ_j denote the mean of class j

and $\bar{\mu}$ denote the grand mean for all the exemplars, then the between class scatter matrix is defined by:

$$B = \sum_{j=1}^{M} K_j (\mu_j - \bar{\mu})(\mu_j - \bar{\mu})^T , \qquad (6)$$

and the within class scatter matrix is defined by:

$$W = \sum_{j=1}^{M} \sum_{k=1}^{K_j} (s_{j,k} - \mu_j)(s_{j,k} - \mu_j)^T \qquad (7)$$

$$W_{FLD} = \arg\max_A \frac{|A^T BA|}{|A^T WA|} \qquad (8)$$

In other words, FLD extracts features that are strong between classes but weak within class. While FLD often yields higher recognition performance than PCA, it tends to overfit to the training data, since it relies heavily on how the within-class scatter captures reliable variations for a specific class [2]. In addition, it is optimised for specific classes, so it needs several samples in every class and thus can determine only a maximum of *M*-1 features.

3. PROPOSED METHOD

An Affine PCA method is introduced in this section in an attempt to overcome some of the limitations of both PCA and FLD. First of all, we apply PCA for dimensionality reduction and to obtain the eigenfaces U. Every face image can be projected into this subspace to form an *m*-dimensional feature vector $s_{j,k}$, where $m < n$, denotes the number of principal eigenfaces chosen for the projection, and $k = 1,2,...K_j$, denotes the k[th] sample of the class S_j, where $j = 1,2,...M$. We often use the nearest neighbor method for classification, where the distance between two face vectors represents the energy difference between them. In the case of variable illumination, lighting changes dominate over the characteristic differences between faces. It has also been proved in [19] that the distance between face vectors with facial expression variations are generally greater than that with face identity. This is the main reason why PCA does not work well under variable lighting and expression. In fact, not all the features have the same importance in recognition. Features that are strong between classes and weak within class are much more useful for the recognition task. Therefore, we propose an affine model (Affine PCA) to resolve this problem. The affine procedure involves three steps: eigenspace rotation, whitening transformation and eigenface filtering.

3.1. Eigenspace Rotation

The eigenfaces extracted from PCA are Most Expressive Features (MEF) and these are not necessarily optimal for face recognition performance as stated in [8]. Applying FLD we can obtain the Most Discriminating Features but overfits to only training data lacking of generalization capacity. Therefore, in order not to lose generalization ability while still keep the discrimination, we prefer to rotate the space and find the most variant features that can represent changes due to lighting or expression variation. That is to extract the within class covariance and apply PCA to find the best eigen features that maximally represent within class variations. The within class difference is defined as:

$$D_{Within} = \sum_{j=1}^{M} \sum_{k=1}^{K_j} s_{j,k} - \mu_j , \qquad (9)$$

and the within class covariance become:

$$Cov_{Within} = D_{Within} D_{Within}^T , \qquad (10)$$

which is a *m* by *m* matrix. Applying singular value decomposition (SVD) to within class covariance matrix, we have,

$$Cov_{Within} = USV^T = \sum_{i=1}^{m} \sigma_i v_i v_i^T .$$

Then the rotation matrix M is the set of eigen vectors of covariance matrix, $M = [v_1, v_2, ...v_m]$. Then all the vectors represented in the original subspace are transformed into new space by multiply by *M*.

3.2. Whitening Transformation

The purpose for whitening is to normalize the scatter matrix for uniform gain control. Since as stated in [3] "mean square error underlying PCA preferentially weights low frequencies", we would need to compensate for that. The whitening parameter Γ is related to the eigenvalues λ_i. Conventionally, we would use the standard deviation for whitening, that is: $\Gamma i = \sqrt{\lambda_i}$, $i = [1...m]$. But this value appears to compress the eigenspace so much that class separability is diminished. We therefore use $\Gamma_i = \lambda_i^p$, where the exponent *p* is determined empirically.

3.3. Filtering the Eigenfaces

The aim of filtering is to diminish the contribution of eigenfaces that are strongly affected by variations. We want to be able to enhance features that capture the main differences between classes (faces) while diminishing the contribution of those that are largely due to lighting or expression variation (within class differences). We thus define a filtering parameter Λ which is related to identity-to-variation (ITV) ratio. The ITV is a ratio measuring the correlation with a change in person versus a change in variations for each of the eigenfaces. For an *M* class problem, assume that for each of the M classes (persons) we have examples under *K* standardized different variations in illumination or expression. In case of illumination changes, the lighting source is positioned in front, above, below, left and right as illustrated in Figure 2. The facial expression changes are normal, surprised and unpleasant as shown in Figure 3. Let us denote the ith eigenface of the kth sample for class (person) S_j by $s_{i,j,k}$. Then

$$ITV_i = \frac{\text{Between Class Scatter}}{\text{Within Class Scatter}} = \frac{\frac{1}{M}\sum_{j=1}^{M}\frac{1}{K}\sum_{k=1}^{K}\left|s_{i,j,k} - \varpi_{i,k}\right|}{\frac{1}{M}\sum_{j=1}^{M}\frac{1}{K}\sum_{k=1}^{K}\left|s_{i,j,k} - \mu_{i,j}\right|}, \tag{11}$$

$$\varpi_{i,k} = \frac{1}{M}\sum_{j=1}^{M} s_{i,j,k},$$

and $$\mu_{i,j} = \frac{1}{K}\sum_{k=1}^{K} s_{i,j,k}, \quad i = [1 \cdots m].$$

Here $\varpi_{i,k}$ represents the ith element of the mean face vector for variation k for all persons and $\mu_{i,j}$ represents the ith element of the mean face vector for person j under all different variations. We then define the scaling parameter Λ by:

$$\Lambda_i = ITV_i^{\,q} \tag{12}$$

where q is an exponential scaling factor determined empirically as before. Instead of this exponential scaling factor, other non-linear functions such as thresholding suggest themselves. These possibilities have been explored, but so far the exponential scaling perform best. After the affine transformation, the distance d between two face vectors $s_{j,k}$ and $s_{j',k'}$ is:

$$d_{jj',kk'} = \sqrt{\sum_{i=1}^{m}\left[\omega_i(s_{i,j,k} - s_{i,j',k'})\right]^2}, \tag{13}$$

$$\omega_i = \Gamma_i \Lambda_i / \left|\Gamma \Lambda^T\right|.$$

The weights ω_i scale the corresponding eigenfaces. To determine the two exponents *p* and *q* for Γ and Λ, we introduce a cost function and optimise them empirically. It is defined by:

$$OPT = \sum_{j=1}^{M}\sum_{k=1}^{K}\sum_{m}\left(\frac{d_{jj,k0}}{d_{jm,k0}}\right), \forall m \in d_{jm,k0} < d_{jj,k0} \tag{14}$$

where $d_{jj,k0}$ is the distance between the sample $s_{j,k}$ and $s_{j,0}$ which is the standard image reference for class S_j (typically the normally illuminated image). Note that the condition $d_{jm,k0} < d_{jj,k0}$ is only true when there is a misclassification error. Thus *OPT* is a combination of error rate and the ratio of between-class distance to within-class distance. By minimizing *OPT*, we can determine the best choices for *p* and *q*. Figure 1, shows the relationship between *OPT* and *p*, *q*. For one of our training database, a minimum was obtained at $p = -0.2, q = -0.4$.

From the above, our final set of transformed eigenfaces would be:

$$u_i' = \omega_i u_i M = \frac{1}{\sigma_i}\omega_i D v_i M \tag{15}$$

where $i = [1 \cdots m]$. After transformation, we can apply PCA again on the compressed subspace to further reduce dimensionality (two-stage PCA).

4. EXPERIMENTAL RESULTS

The method is tested on an Asian Face Image Database PF01 [6] for both changes in lighting source positions and facial expressions. The size of each image is 171×171 pixels with 256 grey levels per pixel. Figure 2 and 3 show some examples from the database. To evaluate the performance of our methods, we performed a 3-fold cross validation on the database as follows. We choose one-third of the 107 subjects to construct our APCA model, one-third for training. Then we just add the normally faces (pictures in the first column in Figure 1 and 2) of the remaining one-thirds of the data into our recognition database. We then attempt to recognize these faces under all the other conditions. This process is repeated three-fold using different partitions and the performance is averaged. All the results listed in this paper are obtained from experiments only on testing data. Table 1 is the comparison of recognition rate between APCA and PCA. It is clear from the results that Affine PCA performs much better than PCA in face recognition under variable lighting conditions. The proposed APCA outperforms PCA remarkably in recognition rate with 99.3% for training data and 95.6% for testing data with negligible reduction in performance for normally lit faces. Figure 3 displays the recognition rates against numbers of eigenfaces used (m). It can be seen that selecting the principal 40 to 50 eigenfaces is sufficient for invariant luminance face recognition. This number is somewhat higher than is required for standard PCA, where selecting m in the range 10 to 20 is sufficient — this is possibly a necessary consequence of the greater complexity of the APCA face subspace compared to standard PCA.

Figure 1. Examples of illumination changes in Asian Face Database PF01.

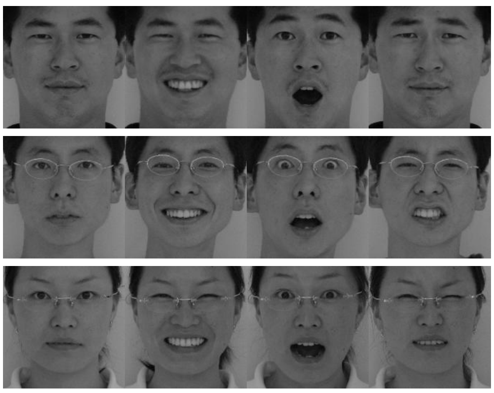

Figure 2. Examples of expression changes in Asian Face Database PF01.

As for variations in facial expression, APCA achieves higher recognition rate than PCA with an increase of 10%. For changes in both lighting condition and expression, APCA always performs better than PCA despite of the change in number of eigenfaces. The gain is almost stable with high dimension of subspace. It can also be seen from Figure 3, that recognition rate of expression changes does not decrease dramatically with the reduce of number of eigen features compared to illumination variations. Therefore, only as low as 20 features is enough to recognition faces with facial expression variations.

We also test the performance of APCA on variations on illumination and expression simultaneously. The recognition rate of APCA is less than 5% lower than that of illumination changes and expression changes, but it is obviously higher than the recognition rate of PCA. Thus it shows that performance of APCA is stable in spite of the complexity of variations. However, PCA is not as robust as APCA with different variations. For illumination changes, PCA only achieve less than 60% accuracy while the accuracy increase to more than 80% for expression variations. It drops back to 60% with changes combining illumination and expression. This phenomenon has also been reported in [19] as any given representation is not sufficient to overcome variations in both illumination and expression.

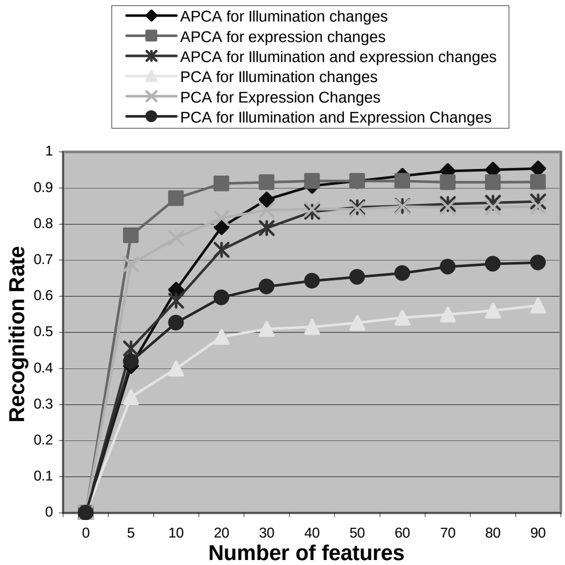

Figure 3. Recognition Rate Vs. Number of features.

Method	Recognition rate		
	Illumination Variation	Expression Variation	Illumination and Expression Variations
PCA	57.3%	84.6%	70.6%
Affine PCA	95.6%	92.2%	86.8%

Table 1. Comparison of recognition rate between APCA and PCA.

Conclusion

We have described an easy to calculate and efficient face recognition algorithm by warping the face subspace constructed from PCA. The affine procedure contains three steps: rotating the eigen space, whitening Transformation, and then filtering the eigenfaces. After affine transformation, features are assigned with different weights for recognition which in fact enlarge the between class covariance while minimizing within class covariance. There only have as few as two variable parameters during the optimization compared to other methods for high dimensionality problems. This method can not only deal with variations in illumination and expression separately but also perform well for the combination of both changes with only one sample image per class. Experiments show that APCA is more robust to change in illumination and expression and have better generalization capacity compared to the FLD method.

A shortcoming of the algorithm is that we can not guarantee that the weights achieved are the best for recognition since we only rotate the eigen space to the direction that best represent the within class covariance. Future work will be to search the eigen space and find the best eigen features suitable for face recognition.

References

[1] P.Belhumeur, J. Hespanha, and D. Kriegman, “Eigenfaces vs. fisherfaces: Recognition using class specific linear projection”, IEEE Trans. Pattern Analysis and Machine Intelligence, Vol.19, No.7, 711-720, 1997.

[2] Chengjun Liu and Harry Wechsler, “Enhanced Fisher Linear Discriminant Models for Face Recognition”, 14th International Conference on Pattern Recognition, ICPR’98, Queensland, Australia, August 17-20, 1998.

[3] Chengjun Liu and Harry Wechsler, “Evolution of Optimal Projection Axes (OPA) for Face Recognition”, Third IEEE International Conference on Automatic face and Gesture Recognition, FG’98, Nara, Japan, April 14-16,1998.

[4] Dao-Qing Dai, Guo-Can Feng, Jian-Huang Lai and P.C. Yuen, “Face Recognition Based on Local Fisher Features”, 2nd Int. Conf. on Multimodal Interface, Beijing, 2000.

[5] Hua Yu and Jie Yang, “A Direct LDA Algorithm for High-Dimensional Data-with Application to Face Recognition”, Pattern Recognition 34(10), 2001, pp. 2067-2070.

[6] Intelligent Multimedia Lab., “Asian Face Image Database PF01”, http://nova.postech.ac.kr/.

[7] Georghiades, A.S. and Belhumeur, P.N. and Kriegman, D.J., "From Few to Many: Illumination Cone Models for Face Recognition under Variable Lighting and Pose", IEEE Trans. Pattern Anal. Mach. Intelligence, vol.23, No. 6, 2001, pp. 643-660.

[8] Daniel L. Swets and John Weng, “Using discriminant eigenfeatures for image retrieval”, IEEE Trans. on PAMI, vol. 18, No. 8, 1996, pp. 831-836.

[9] X.W. Hou, S.Z. Li, H.J. Zhang. "Direct Appearance Models". In Proceedings of IEEE International Conference on Computer Vision and Pattern Recognition. Hawaii. December, 2001.

[10] Z. Xue, S.Z. Li, and E.K. Teoh. "Facial Feature Extraction and Image Warping Using PCA Based Statistic Model". In Proceedings of 2001 International Conference on Image Processing. Thessaloniki, Greece. October 7-10, 2001.

[11] S.Z. Li, K.L. Chan and C.L. Wang. ``Performance Evaluation of the Nearest Feature Line Method in Image Classification and Retrieval". IEEE Transactions on Pattern Analysis and Machine Intelligence, 22(11):1335-1339. November, 2000.

[12] G.D. Guo, H.J. Zhang, S.Z. Li. "Pairwise Face Recognition". In Proceedings of 8th IEEE International Conference on Computer Vision. Vancouver, Canada. July 9-12, 2001.

[13] S. Mika, G. Ratsch, J.Weston, and K. R. M. B. Scholkopf, “Fisher discriminant analysis with kernels”, Neural networks for Signal Processing IX, 1999, pp.41-48.

[14] M. A. Turk and A. P. Pentland, “Eigenfaces for recognition”, Journal of Cognitive Neuroscience, vol. 3, No. 1, 1991, pp.71-86.

[15] Jie Zhou and David Zhang “Face Recognition by Combining Several Algorithms”, ICPR 2002.

[16] Alexandre Lemieux and Marc Parizeau, “Experiments on Eigenfaces Robustness”, ICPR 2002.

[17] A. M. Martinez and A. C. Kak, “PCA versus LDA”, IEEE TPAMI , 23(2):228–233, 2001.

[18] A. Yilmaz and M. Gokmen, “Eigenhill vs. eigenface and eigenedge”, In Proceedings of International Conference Pattern Recognition, Barcelona, Spain, 2000, pp.827-830.

[19] Yael Adin, Yael Moses, and Shimon Ullman, “Face Recognition: The problem of Compensating for Changes in Illumination Direction”, IEEE PAMI, Vol. 19, No. 7, 1997.

[20] Aleix M. Martinez, “Recognizing Impercisely Localized, Partially Occluded and Expression Variant Faces from a Single Sample per Class”, IEEE TPAMI, Vol. 24, No. 6, 2002.

[21] Ronen Basri and David W. Jacobs, “Lambertian Reflectance and Linear Subspaces”, IEEE TPAMI, Vol. 25, No.2 2003.

[22] Peter W. Hallinan, “A Low-Dimensional representation of Human faces for Arbitrary Lighting Conditioins”, Proc. IEEE Conf. Computer Vision and Pattern recognition, 1994.

[23] D. Beymer and T. Poggio, “Face Recognition from One Example View”, Science, Vol. 272, No. 5250, 1996.

[24] M. J. Black, D. J. Fleet and Y. Yacoob, “Robustly estimating Changes in Image Appearance”, Computer Vision and Image Understanding, Vol. 78, No. 1, 2000.

[25] Shaokang Chen, Brian C. Lovell and Sai Sun, “Face Recognition with APCA in Variant Illuminations”, Workshop on Signal Processing and Applications, Australia, December, 2002.