Representative Feature Chain for Single Gallery Image Face Recognition

Shaokang Chen, Conrad Sanderson, Sai Sun and Brian C. Lovell
NICTA, Queensland, Australia
ITEE, The University of Queensland, Australia

Abstract

*Under the constraint of using only a single gallery image per person, this paper proposes a fast multi-class pattern classification approach to 2D face recognition robust to changes in pose, illumination, and expression (PIE). This work has three main contributions: **(1)** we propose a representative face space method to extract robust features, **(2)** we apply a learning method to weight features in pairs, **(3)** we combine the feature pairs into a feature chain in order to find the weights for all features. The approach is evaluated for face recognition under PIE changes on three public databases. Results show that the method performs considerably better than several other appearance-based methods and can reliably recognise faces at large pose angles without the need for fragile pose estimation pre-processing. Moreover, computational load is low (comparable to standard eigenface methods), which is a critical factor in wide-area surveillance applications.*

1 Introduction

Most recent research on face recognition has been focused on diminishing the impact of nuisance factors such as changes in pose, illumination, and expression (PIE). Various methods have been proposed to handle certain kinds of face image variation successfully, but drawbacks still restrict their application.

Approaches in [2, 6] attempt to construct face specific models to describe changes in lighting or pose for certain faces. However, to construct a specific model (subspace) for each person, these methods need to take multiple images per person under controlled conditions. This leads to expensive image capture processes, poor scalability of the face model, and restrictions on applications where only one gallery image is available per person. Other researchers divide the range of variation into several subranges (e.g., low, medium, and high pose angles) and construct multiple face spaces to describe face variations lying in the corresponding subrange [13]. These approaches require us to register several images representing different variations per person into the corresponding variation models so that matching can be done in each interval individually. Again, acquiring multiple images per person under specific conditions is often very difficult or even practically impossible.

In this paper we propose an appearance based approach for reliable face recognition under PIE variations using a single gallery image. In order to extract meaningful features, we develop a representative face space method to enhance feature representativeness by space rotation. This is combined with a technique for finding the optimal weights within feature pairs. We then construct a feature chain to combine all of the feature pairs. We call this approach Representative Feature Chain (RFC).

It is worth noting that complexity of the proposed method is similar to standard Principal Component Analysis (PCA), while achieving considerably better performance as demonstrated on several public databases. Compared to other recent approaches for dealing with pose variations [4, 5], RFC is by design a relatively low-complexity technique and is hence more suitable for real-time surveillance applications.

We continue the paper as follows. The representative face space is described in Section 2. Feature weighting and feature chain construction are described in Section 3. In Section 4 we first discuss the effect of space rotation on discrimination power, followed by performance comparisons against several recent face recognition methods. The main findings and areas for future work are given in Section 5.

2 Representative face space

For appearance based face representation and recognition, many decomposition approaches have been proposed to reduce face space dimensionality, such as PCA, Linear Discriminant Analysis (LDA), Evolutionary Pursuit (EP), Independent Component Analysis (ICA), and their kernel counterparts. PCA minimises the mean square error for face representation, making the spectrum maximally compacted with greatest concentration of energy in the leading eigenvectors. LDA finds the optimum projection that maximises between-class differences and simultaneously minimises the within-class variation. Because LDA depends significantly on how the within-class scatter captures reliable variations

[*] **Acknowledgements:** NICTA is funded by the Australian Government as represented by the *Department of Broadband, Communications and the Digital Economy* as well as the Australian Research Council through the *ICT Centre of Excellence* program.

for a specific class, it tends to overfit to training data [9]. EP and ICA overcome the drawbacks of PCA and LDA by pairwise axis rotation transformations to find the best projection axis according to certain criteria. Experiments on face recognition show that recognition performance is significantly improved after such face space rotation is used in conjunction with processing such as whitening [10]. However, due to the huge number of combinations of axis pairs and rotation angles, computationally intensive methods such as genetic algorithms are required to search for the solution.

Adopting the idea of rotation, we propose the following method. Let us assume we have an offline training set (separate from the gallery set), which is used solely for the construction of the representative subspace. Furthermore, every face image from the training set is represented in the eigenface subspace by an m-dimensional feature vector $s_{j,k}$ with $k = 0, 1, ..., K_j$ denoting the k-th sample of the class S_j. We then choose one sample (generally the normally lit frontal neutral expression image) in each class as a reference denoted by $s_{j,0}$. A referenced within-class covariance matrix is then constructed by:

$$\begin{aligned} D_w &= \sum\nolimits_{j=1}^{N} \sum\nolimits_{k=1}^{K_j} (s_{j,k} - s_{j,0}) \\ C_w &= D_w D_w^T, \end{aligned} \tag{1}$$

where N is the number of classes. Applying eigendecomposition on C_w we have:

$$C_w u_i = \lambda_i u_i. \tag{2}$$

where $i \in [1, m]$. All the eigenvectors u_i are combined into one matrix to obtain the square rotation matrix U. The within-class rotation captures the within-class variations corresponding to a certain reference class sample. Given the original vector $s_{j,k}$, a rotated version is then obtained with:

$$\tilde{s}_{j,k} = U^T s_{j,k} \tag{3}$$

3 Pairwise weighting and feature chain

After extracting features, we need to assign optimal weights to all the features due to the observation that not all features have the same importance in recognition. While various methods have been developed to improve PCA and LDA by whitening [9, 10] to compensate for the overweighting of the leading features, normal whitening may excessively enhance minor features, which leads to over-fitting to the training data. It is difficult to assign appropriate weights to all features in a high dimensional space at the same time. We thus propose a learning method to weight features pairwise.

For any feature pair $[a, r]$ from the m dimensional space, we assign weight $\eta_{[a,r]} \in [0, 1]$ to feature a, and weight $\sqrt{1 - \eta^2_{[a,r]}}$ to feature r. We define the difference of two face images $I_{j,k}$ and $I_{j'k'}$ lying in the subspace (plane) defined by features a and r as the projection of the Euclidean distance of their transformed vectors onto this plane:

$$D^{[a,r]}_{jk,j'k'} = \sqrt{\begin{array}{l} \eta^2_{[a,r]}(\tilde{s}_{j,k,a} - \tilde{s}_{j',k',a})^2 \\ + (1 - \eta^2_{[a,r]})(\tilde{s}_{j,k,r} - \tilde{s}_{j',k',r})^2 \end{array}} \tag{4}$$

where $\tilde{s}_{j,k,a}$ and $\tilde{s}_{j,k,r}$ represents the a-th and the r-th element of the vector $\tilde{s}_{j,k}$. We then define a continuous cost function Λ to search the one dimensional space to determine the optimal value for $\eta_{[a,r]}$ as follows:

$$\begin{aligned} \Lambda\left(\eta_{[a,r]}\right) &= \sum\nolimits_{j=1}^{N} \sum\nolimits_{k=1}^{K_j} \sum\nolimits_{n} \left(\frac{D^{[a,r]}_{jk,j0}}{D^{[a,r]}_{jk,n0}} \right) \\ & \qquad \forall n \in D^{[a,r]}_{jk,n0} < D^{[a,r]}_{jk,j0}, n \in [1, \cdots, N] \end{aligned} \tag{5}$$

where $D^{[a,r]}_{jk,j0}$ is the within-class difference between the sample $I_{j,k}$ and its corresponding reference image $I_{j,0}$ in class S_j. Note that the condition $D^{[a,r]}_{jk,n0} < D^{[a,r]}_{jk,j0}$ is only true when there is a misclassification error. The optimal weight for feature pair $[a, r]$ is found with:

$$\eta_{[a,r]} = \arg\min_{\widehat{\eta}_{[a,r]}} \Lambda\left(\widehat{\eta}_{[a,r]}\right). \tag{6}$$

We assign the weight $\sqrt{1 - \eta^2_{[a,r]}}$ to feature r so that $\eta^2_{[a,r]} + (\sqrt{1 - \eta^2_{[a,r]}})^2 = 1$. This ensures that Eqn. (4) is comparable across different feature pairs. We empirically found that the shape of most Λ curves tends to be approximately concave, and hence elected to use a golden section search [8] for the minimisation.

We now construct a feature chain to combine all of the features by assigning appropriate weights for each pair, as follows. First, a reference feature r is chosen from the m available features. Second, feature pair weighting is found for feature r paired with each of the remaining $m - 1$ features. Consequently, we have a set of η values: $(\eta_{[1,r]}, \eta_{[2,r]}, \cdots, \eta_{[r-1,r]}, \eta_{[r+1,r]}, \cdots, \eta_{[m,r]})$.

The weights w_i associated to each feature pair $[i, r]$ must satisfy the following constraints to ensure that the weight f_r finally assigned to the reference feature r by each pair is consistent inside the chain, that is:

$$f_r = w_1\sqrt{1\text{-}\eta^2_{[1,r]}} = w_2\sqrt{1\text{-}\eta^2_{[2,r]}} = \cdots = w_m\sqrt{1\text{-}\eta^2_{[m,r]}} \tag{7}$$

In addition, for the whole feature chain, the weight f_i assigned to each feature must follow:

$$\sum\nolimits_{i=1}^{m} f_i^2 = 1, \tag{8}$$

$$w_i \cdot \eta_{[i,r]} = f_i, \quad (i \neq r). \tag{9}$$

From the above three equations, we have:

$$f_r = \sqrt{1/\left(1 + \frac{\eta^2_{[1,r]}}{1\text{-}\eta^2_{[1,r]}} + \frac{\eta^2_{[2,r]}}{1\text{-}\eta^2_{[2,r]}} + \cdots + \frac{\eta^2_{[m,r]}}{1\text{-}\eta^2_{[m,r]}}\right)}, \tag{10}$$

$$w_i = \frac{1}{\sqrt{1 - \eta^2_{[i,r]}}} \cdot f_r, \quad (i \neq r). \tag{11}$$

4 Empirical evaluation

In this section we first study some of the effects of the Representative Face Space (Section 4.1), followed by a performance comparison of the proposed RFC method against five other techniques (Section 4.2).

4.1 Effects of representative face space

Figure 1 shows the spectral analysis of the covariance of features [9] derived by PCA, LDA, and our RFC on the pose subset of the FERET database [11]. The face space is constructed by applying PCA on 603 images of 67 people with each person having 9 different poses. The top 120 eigen features are selected which accounts for 95% of the overall covariance. The x-axis is the feature number (index) from 1 to 120 and the y-axis is the log scale of the relative magnitude of the covariance of the feature over the total covariance of face images.

The features are sorted in descending order of relative magnitude. Generally, the higher the feature number is, the higher the frequency it spans in the spectrum. We can see from this figure that features extracted by PCA capture most of the variation in face space in the first 10 features, with relative magnitudes larger than 0.01. But the high frequency features of PCA with feature numbers ranging from 62 to 120, have relative magnitude much smaller than 0.001. Because noise in the training samples used to construct the face space normally appears in the higher frequencies, these features with small relative magnitude will have low signal to noise ratio (SNR).

LDA abstracts features that capture minor covariance having only 2 features with value larger than 0.01 and 41 features with value less than 0.001. Most of the features of LDA have small SNR which may help explain why LDA tends to overfit to the noise in the training data. Thus, many researchers improve LDA by applying a cut off frequency to the features [9] to increase the average feature SNR.

118 out of 120 features in our proposed RFC have a relative magnitude value greater than 0.001 with better SNR. The spectral magnitude of representative space is always the largest of the three spectra except for the first two features. That is because the first two features of the RFC space mainly models within-class covariance, whilst the top two leading features of PCA capture both within- and between-class covariance. In summary, discriminative features extracted by representative space have a higher SNR than PCA and LDA, we thus conjecture that classifiers built on these features may have better generalization ability.

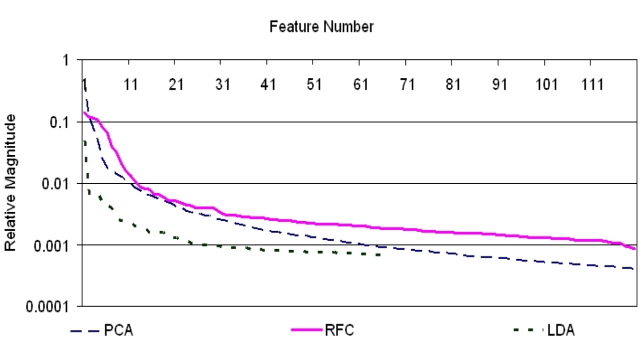

Figure 1. **Relative magnitude of feature covariance.**

4.2 Performance Comparison

In this section we evaluate the performance of the RFC method on three publicly available databases, Asian Face Database [1], a subset of PIE [14] and FERET [11], against the following five techniques: PCA, LDA, Synthesis+PCA [12], Pose-Robust Features [12] and Eigen Light-Fields [7]. For all trials, we divide the corresponding data set into three equal-sized disjoint partitions with different subjects. We then choose images from one of the partitions for training and the remaining two partitions for testing. In each case the training set is used to construct the representative space and weight feature pairs. The test set contains images of unseen subjects. For testing, we only register one neutral normally lit frontal image per subject as gallery and use the remainder of the images as probe. All of the results are the average of three-fold cross validation using three different partitions of the datasets.

The Asian Face Database contains 103 persons with each person 17 images including 1 normal face, and 16 images with PIE variations as listed in Table 1. All images are aligned according to their eye positions.

As can be seen from Table 1, PCA and LDA are considerably outperformed by RFC — an average of 60% and 80.9% correct recognition vs 93%, respectively. RFC always performs the best among three methods on all variations. The largest difference occurs on illumination changes, where RFC is more than three times better than PCA, due to PCA's sensitivity to within-class changes. All three methods are relatively more sensitive to expression changes due to the fact that different people express the same expression somewhat differently to others, which makes the expression changes harder to model. However, RFC is still far better than PCA and LDA on expression changes with more than 21 percentage points higher in accuracy.

The lowest recognition rate of 80.3% for RFC is for expression changes with closed eyes, which also affects PCA and LDA substantially as they achieve only 50.7% and 54.9%, respectively. The reason is that the alignment of face images relies heavily on the eyes – with the eyes closed, the alignment is less accurate, leading to differences in scale. Overall, RFC is far more robust to PIE variations than PCA and LDA with relatively little change in accuracy across all three variations.

For comparison with the Synthesis+PCA and Pose-Robust Features methods, the pose variation subsets of the PIE (204 images) and FERET (1800 images) databases were used. For each person, the frontal face image was the gallery image and the remaining images were the probe images. All the images were horizontally scaled and aligned according to their eye positions. This normalization is an approxima-

Table 1. **Accuracy on the Asian Face database**

Variation Type	Method	Database subset 1	2	3	4	Average
Illumination	PCA	16.9	39.4	26.8	32.4	28.9
	LDA	73.2	94.4	91.5	95.8	88.7
	RFC	**95.8**	**98.6**	**95.8**	**99.1**	**97.3**
Pose 1	PCA	84.5	77.5	62.0	74.6	74.7
	LDA	93.0	87.3	74.6	88.7	85.9
	RFC	**95.8**	**93.1**	**87.3**	**91.6**	**92.0**
Pose 2	PCA	83.1	70.4	60.6	69.0	70.8
	LDA	88.7	77.5	73.2	81.7	80.3
	RFC	**94.4**	**95.8**	**88.7**	**95.3**	**93.6**
Expression	PCA	80.3	67.6	64.8	50.7	65.9
	LDA	88.7	69.0	62.0	54.9	68.7
	RFC	**95.8**	**90.1**	**93.2**	**80.3**	**89.9**

Table 2. **Accuracy on the PIE database**

Pose	PCA	LDA	Synthesis + PCA	Pose-Robust Features	RFC
-22.5°	30.2	62.3	60.0	83.3	**92.5**
+22.5°	13.2	37.7	56.0	80.6	**88.7**

Table 3. **Accuracy on the FERET database.**

Pose	PCA	LDA	Synthesis + PCA	Pose-Robust Features	RFC
$-60°$	23.3	62.4	-	-	**75.2**
$-40°$	36.8	71.4	-	-	**84.2**
$-25°$	53.4	78.2	50.0	85.6	**90.2**
$-15°$	79.7	84.2	71.0	88.2	**94.0**
$+15°$	66.1	85.7	67.4	88.1	**93.2**
$+25°$	46.6	81.2	42.0	66.8	**92.5**
$+40°$	35.3	75.9	-	-	**89.5**
$+60°$	28.6	69.2	-	-	**75.9**

Table 4. **Average accuracy on the FERET database**

Method	PCA	LDA	Eigen Light-Fields	RFC
Average Accuracy	40.6	76.0	75.0	**88.5**

tion of the 3-point normalization used in [7].

Tables 2 and 3 show the results obtained on the PIE and FERET databases, respectively. The RFC method remarkably outperforms the other four methods on both databases under all pose angles. This effect is more significant for poses with angles greater than $\pm 40°$.

Table 4 shows the comparison with the Eigen Light-Fields [7] method. For consistency with the results presented in [7], we report the average recognition accuracy across all poses, using each pose angle individually for gallery images. The result of standard PCA method in our test is 40.6%, comparable to 39.4% in [7], which implies that our image normalization is a close approximation of the 3 point normalization. It can be seen that RFC performs considerably better than Eigen Light-Fields. We note that in RFC we do not need to determine the pose angles of the images, while in the Eigen Light-Fields method camera intrinsics and relative orientation of the camera to the object should be acquired beforehand. This is often difficult or impossible in some situations.

5 Conclusions and Future Work

In this paper we have proposed a novel appearance-based method, dubbed Representative Feature Chain (RFC), for face recognition robust to changes in pose, illumination, and expression (PIE). RFC consists of three main steps: (1) referenced within-class space rotation to enhance feature representativeness; (2) assignment of optimal weights to features in pairs; and (3) feature chaining to find the weights of all features. Comparisons on three public databases with PIE variations show that RFC outperforms three recent appearance-based recognition methods: Synthesis+PCA, Pose-Robust Features, and Eigen Light-Fields. However, being a holistic method, RFC is still sensitive to geometric transformations such as scale change and translation; we note that the technique presented in [3] can be adopted to overcome these drawbacks.

References

[1] Asian Face Image Database PF01. Intelligent Multimedia Lab, Pohang University of Science and Technology. http://nova.postech.ac.kr/.

[2] R. Basri and D. W. Jacobs. Lambertian reflectance and linear subspaces. *IEEE TPAMI*, 25(2):218–233, 2003.

[3] H. Bischof and A. Leonardis. Robust recognition of scaled eigenimages through a hierarchical approach. In *CVPR*, 1998.

[4] V. Blanz, K. Scherbaum, and H.-P. Seidel. Fitting a morphable model to 3D scans of faces. In *ICCV*, 2007.

[5] C. D. Castillo and D. W. Jacobs. Using stereo matching for 2D face recognition across pose. In *CVPR*, 2007.

[6] A. S. Georghiades, P. N. Belhumeur, and D. J. Kriegman. From few to many: Illumination cone models for face recognition under variable lighting and pose. *IEEE TPAMI*, 23(6):643–660, 2001.

[7] R. Gross, I. Matthews, and S. Baker. Appearance-based face recognition and light-fields. *IEEE TPAMI*, 26(4):449– 465, 2004.

[8] J. Kiefer. Sequential minimax search for a maximum. *Proceedings of the American Mathematical Society*, 4:502 – 506, 1953.

[9] C. Liu and H. Wechsler. Enhanced fisher linear discriminant models for face recognition. In *ICPR*, 1998.

[10] C. Liu and H. Wechsler. Evolutionary pursuit and its application to face recognition. *IEEE TPAMI*, 22(6):570–582, 2000.

[11] P. J. Phillips, H. Moon, P. J. Rauss, and S. Rizvi. The FERET evaluation methodology for face recognition algorithms. *IEEE TPAMI*, 20(10):1090–1104, 2000.

[12] C. Sanderson, T. Shan, and B. C. Lovell. Towards pose-invariant 2D face classification for surveillance. In *Analysis and Modeling of Faces and Gestures, Lecture Notees in Computer Science (LNCS) Vol. 4778*, pages 276–289, 2007.

[13] P. Sankaran and V. Asari. A multi-view approach on modular PCA for illumination and pose invariant face recognition. In *Proc. of Applied Imagery Pattern Recognition Workshop*, 2004.

[14] T. Sim, S. Baker, and M. Bsat. The CMU pose, illumination, and expression database. *IEEE TPAMI*, 25(12):1615 – 1618, 2003.