Modular PCA and Probabilistic Similarity Measure for Robust Face Recognition

Kwanyong Lee[1] and Hyeyoung Park[2]
[1]Korea National Open University, Seoul, Korea
[2]School of Electrical Engineering and Computer Science
Kyungpook National University, Daegu, Korea
kylee@knou.ac.kr, hypark@knu.ac.kr

Abstract

This paper addresses a probabilistic approach to develop a robust face recognition system to partial variations such as occlusions. Based on the statistical feature extraction methods, we take the modular PCA method which finds eigenspace not for the set of whole images but for the sets of local image patches. Through the local feature extraction approach, we try to overcome the drawback of wholistic appearance-based conventional PCA, and consequently expect to improve robustness to local variations. The obtained local features are then applied to define two probabilistic models for facial images: one for modeling distribution of features observed in usual facial images, and the other for modeling distribution of environmental variations observed in face image from one subject. The probabilistic model for general facial images are used to evaluate the importance of each local patch. The probabilistic model for the environmental variations is used to evaluate the similarity between two local features. By combining two probabilistic models, we finally define a distance measure between two face images, which can be applied for face recognition. Computational experiments on benchmark face database show that the proposed face recognition method can achieve remarkable robustness to local variations.
Keywords: *face recognition, statistical feature extraction, probabilistic model, modular PCA, similarity measure, local variations*

1: Introduction

In spite of active research works on face recognition in the field of computer vision and pattern recognition[4, 10], it is still a challenging problem due to the diverse variations of facial images. While some variations such as illumination and translation cause the changes of whole images, there are also local variations such as facial expressions and partial occlusions that cause distortions of local area of an image. In order to deal with these variations efficiently, it is important to develop a robust feature extraction method that can keep the essential information and also can exclude the unnecessary variational information.

Statistical feature extraction methods such as PCA, LDA, and their nonlinear extensions[6, 7, 8, 9] can give efficient low dimensional features through learning the variational proper-

[2]Corresponding Author

ties of data set. However, since the statistical approaches consider a sample image as a data point in the input space, it is difficult to handle local variations in image data. Especially, in the case of facial images, there are many types of face-specific occlusions by sun-glasses, scarfs, and so on. Therefore, for the facial data with occlusions, it is hard to expect the statistical approaches to give good performances.

To solve this problem, some modular statistical approaches such as modular PCA has been developed[1, 2, 3]. By dividing an image into a number of local image patches and by applying statistical feature extraction to each local patches, it is possible to obtain a sort of local features that can represent local variations. These previous works on modular PCA have mainly focused on the feature extraction stage, and the obtained local features are merged in the classification stage by using simple average. Therefore, strong local distortion due to partial occlusions are not excluded and can still make bad influence in the recognition.

To overcome the drawback, this paper propose a new similarity measure between two images represented by local features. The similarity between two local features are measured by using two independent probability density functions. The first pdf is obtained by modeling the distribution of local features observed in usual faces. The probability value implies the normality of the local features in the sense of general face images. When a given local path includes occluded part, the corresponding probability value becomes low. The second pdf is obtained by modeling the distribution of the difference between two local features obtained from single subject. When two given local features are given from the same subject, the probability of their difference becomes high. Thus, the probability value implies the similarity of two local features. By combining two probability density, we can define a novel probabilistic distance measure for two facial images. Detail description on the probabilistic models are given in next sections.

2: Probabilistic model for general facial images

As described in the above section, an image I can be represented by a fixed number (M) of local features $\mathbf{x}_m$ ($m = 1, \ldots, M$). When the training set of facial images are given as $\{\mathrm{I}^i\}_{i=1,\ldots,N}$, we can compose M sets of local patches, which can be written as

$$X_m = \{\mathbf{x}_m^i | \mathbf{x}_m^i \in \mathrm{I}^i,\ i = 1, \ldots, N\},\ m = 1, \ldots, M. \tag{1}$$

The set X_m has local patches at a specific location (i.e. mth location) of facial images obtained from all training data. Then the whole set of local patches for training set is given as

$$\mathbf{X} = \bigcup_{m=1}^{M} X_m. \tag{2}$$

To obtain low-dimensional statistical features, we apply PCA for the set $\mathbf{X}$ and find a projection matrix W and low-dimensional features $\mathbf{y}_m$. For each local features $\mathbf{y}_m^i$, their corresponding feature set is given as

$$\mathbf{Y}_m = \left\{\mathbf{y}_m^i |\ \mathbf{y}_m^i = W^T \mathbf{x}_m^i,\ \mathbf{x}_m^i \in \mathrm{I}^i\right\}. \tag{3}$$

Using the set $\mathbf{Y}_m$, we try to estimate the probability density of mth feature $\mathbf{y}_m$. As a simple model, we use the Laplacian distribution which can be written by

$$p_m(\mathbf{y}) = \prod_{i=1}^{d} f(y_i|\mu_i, b_i) \quad (4)$$

$$= \prod_{i=1}^{d} \frac{1}{2b_i} \exp\left\{-\frac{|y_i - \mu_i|}{b_i}\right\} \quad (5)$$

where y_i is the ith element of d-dimensional feature vector $\mathbf{y}$. The two model parameters, μ_i and b_i, can be estimated by sample median and sample mean of the absolute deviation of y_i from μ_i respectively.

Using the estimated probability density function, we can calculate the probability that each local feature is observed at a specific position of an average facial image without variations. When a test image given, its local features can have corresponding probability values, and we can use them as an importance value of the feature in the distance measure, as will be described in next section.

3: Probabilistic similarity measure for face recognition

In order to get a similarity measure between two images I^i and I^j, we need to define the similarity between two local features $\mathbf{y}_m^i$ and $\mathbf{y}_m^j$. To obtain the similarity function, we define a difference vector between two local features as

$$\delta_m^{ij} = \mathbf{y}_m^i - \mathbf{y}_m^j. \quad (6)$$

Using the difference vector, we can compose an intraclass difference set such as

$$\mathbf{\Delta}_m = \left\{\delta_m^{ij} | \delta_m^{ij} = \mathbf{y}_m^i - \mathbf{y}_m^j,\ \mathrm{I}^i \text{ and } \mathrm{I}^j \text{ are from the same subject}\right\}. \quad (7)$$

We then define a similarity between two features $\mathbf{y}_m^i$ and $\mathbf{y}_m^j$ using the probability that δ^{ij} belongs to the set $\mathbf{\Delta}_m$, which can be written as

$$S(\mathbf{y}_m^i, \mathbf{y}_m^j) = q_m(\delta^{ij}|\mathbf{\Delta}_m). \quad (8)$$

To obtain the explicit pdf $q_m(\delta^{ij}|\mathbf{\Delta}_m)$, we also use the Laplace distribution as defined in (4) and estimate the parameters using the set $\mathbf{\Delta}_m$.

When a test image $\mathrm{I}^{tst}=\{\mathbf{y}_m^{tst}\}_{m=1\ldots M}$ is given, the similarity between I^{tst} and a training image $\mathrm{I}^i=\{\mathbf{y}_m^i\}_{m=1\ldots M}$ can be defined by combining two probabilities as follows:

$$S(\mathrm{I}^{tst}, \mathrm{I}^i) = \sum_{i=1}^{M} p_m(\mathbf{y}_m^{tst}) q_m(\delta^{tst,i}|\mathbf{\Delta}_m). \quad (9)$$

Using the similarity measure and K-NN classifier, a test image is assigned to a class in which the image with maximum similarity is included.

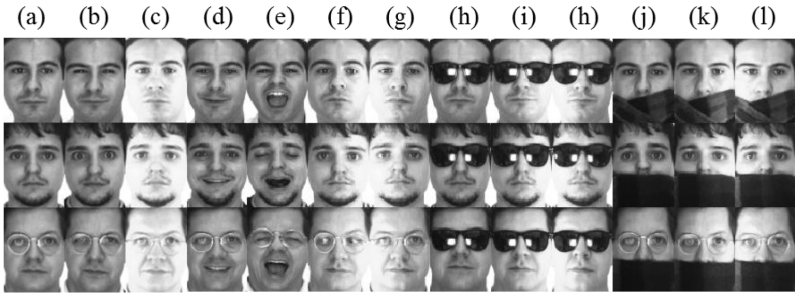


Figure 1. Sample images of AR database.

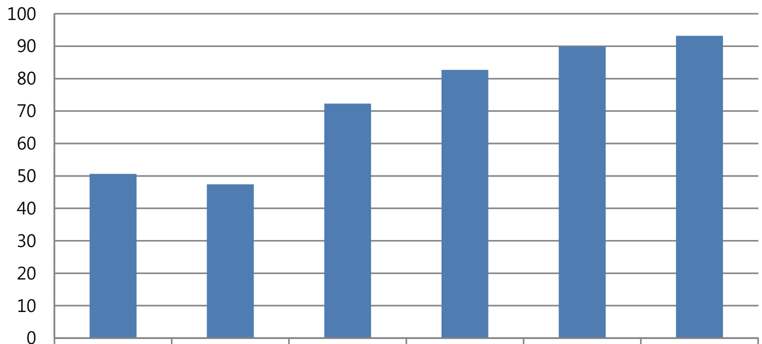
4: Experimental Comparisons

4.1: Experimental data

In order to confirm the robustness of the proposed method, we conducted computational experiments on AR database [5] with local variations including partial occlusions. We compare the proposed method with the conventional statistical feature extraction methods. The AR database consists of over 3,200 color images of frontal faces from 126 individuals: 70 men and 56 women. There are 26 different images for each person. For each subject, these were recorded in two different sessions separated by two weeks delay. Each session consists of 13 images which has differences in facial expression, illumination and partial occlusion. In this experiment, we selected 100 individuals and used 13 images taken in the first session for each individuals. Through preprocessing, we obtained manually aligned images with the location of eyes. After localization, faces were morphed and resized to 88 by 64 pixels. Sample images from three subjects are shown in Fig. 1. As shown in the figure, the AR database has several examples with occlusions. In the experiments, three non-occluded images (i.e., Fig. 1. (a), (b), and (c)) from each person were used for training, and other ten images for each person were used for testing.

4.2: Experimental results

Considering two main advantages of the proposed method, which are the use of pdf for evaluating importance of local feature and the use of pdf for evaluating the similarity between two local features, we conducted recognition tasks using six different methods. The naive K-NN classification methods using Manhattan distance for raw images are first applied to check the baseline of the performance. We also conducted experiments using the original PCA, and the conventional modular PCA with naive K-NN classification. For modular PCA, we divided an image with 88×64 size into 11×8 local patches so as to obtain 88 low-dimensional local features. The dimensionality of local features are determined based on the 98% reconstruction rates, so as to obtain 10-dimensional vector. To the obtained local features, we applied the proposed probabilistic models. To see the effect of the probabilistic model of usual facial images, we only used the probability (importance) $p_m(\mathbf{y}^{tst})$ with Manhattan distance for similarity between two local features. Also, we conducted experiments with only probability (similarity) $q_m(\delta^{tst,i})$, so as to see the effect of the probabilistic model of facial variations. Finally, the proposed method with $p_m(\mathbf{y}^{tst})$ and $q_m(\delta^{tst,i})$ is applied to the same data.



	raw image	PCA	modular PCA	modular PCA + pdf 1	modular PCA + pdf 2	modular PCA + pdf 1, 2
classification rate (%)	50.6	47.4	72.3	82.7	89.9	93.2

Figure 2. Result of face recognition on AR database with occlusion

The result of the experiments for the six methods are shown in Fig. 2. As we can expect, the methods using raw images and original PCA show low classification rates. By using local features obtained by modular PCA, the classification rate was significantly improved compared to original PCA. In addition, the use of pdf 1(the importance of local features) and pdf 2 (the similarity of two local features) makes improvement in the recognition performance respectively, and their combination further improved the performance.

5: Conclusions

In this paper, we proposed a robust face recognition method by using two probabilistic models for facial data. Through estimating the probability density of local features observed in training images, we can measure the importance of each local features of test images. Through estimating the probability density of difference between local features from the same subject, we can measure the similarity of two local features. Though we use modular PCA to obtain local features, other types of local feature descriptors can also be used. In addition, the proposed method may be applied to other visual recognition problems such as object recognition by choosing appropriate training set and probability density model of local features.

Acknowledgements

This research was partially supported by the MKE(The Ministry of Knowledge Economy), Korea, under the ITRC(Information Technology Research Center) support program (NIPA-2012- H0301-12-2004) supervised by the NIPA(National IT Industry Promotion Agency)

This research was partially supported by the Converging Research Center Program funded by the Ministry of Education, Science and Technology (2011K000659)

References

[1] T. Deb, D. Bhattacherjee, M. Nasipuri, D. K. Basu and M. Kundu, An enhanced face recognition technique based on Overlapped Modular PCA approach for cropped Log-polar images, International Journal of Recent Trends in Engineering, 2(1), 2009.

[2] An improved face recognition technique based on modular PCA approach, Pattern Recognition Letters 25, 429–436, 2004.

[3] C. Han. Modular PCA Face Recognition Based on Weighted Average Modern Applied Science, 3(11), 64–70, 2009.

[4] R. Jafry and H.R. Arubnia, A Survey of Face Recognition Techniques, Journal of Information Processing Systems, 5(2), 41–68, 2009.

[5] A. Martinez and B. Benavente, R. The AR face database. CVC Technical Report #24, June, 1998

[6] A. M. Martinez and A. C. Kak. PCA versus LDA. IEEE Trans. Pattern Anal. Mach. Intell. 23(2) 228–233, 2001.

[7] M. Turk, A., Pentland, Eigenfaces for recognition. Journal of cognitive neuroscience **3**(1) (1991) 71–86

[8] M. Yang, N. Ahuja, and D. Kriegman, Face Recognition Using Kernel Eigenfaces, in Proceedings of the 2000 IEEE International Conference on Image Processing (ICIP 2000), 37–40, Vancouver, Canada, September, 2000.

[9] J. Yang, A. F. Frangi, J. Yang, and D. Zhang, KPCA Plus LDA: A Complete Kernel Fisher Discriminant Framework for Feature Extraction and Recognition, IEEE Trans. on Pattern Analysis and Machine Intelligence, 27(2), 230-244, 2005.

[10] W. Zhao, R. Chellappa, , Phillips, P.J., Rosenfeld, A.: Face recognition: A literature survey. ACM Comput. Surv. 35(4) 399–458, 2003.