Similarity of Macaque Face Images based on PCA

Jiaqi Zhang, Xiaoyuan Li, Yuxia H
School of Electrical Engineering,
Zhengzhou University
Zhengzhou 450001
E-mall: jiaqi52361@hotmail.com

Zhenlong Wang
Department of Biology
Zhengzhou University
Zhengzhou 450001
E-mall:˙,lixiaoyuan@zzu.edu.cn

Abstract—**In this work, Block PCA based method is used to build several subspaces for each macaque face image and a similarity criterion is proposed to distinguish family properties of macaques. A proper threshold value is determined through statistical analysis. Then, a series of experiments are carried out to verify the feasibility of this approach. The results demonstrate that it is proper to utilize similarity value as judgment criterion of macaque family properties and higher accurate rate is obtained. It can provide an effective way for population recognition of other species.**

Keywords-principle component analysis(PCA), similarity, Block PCA, t-test

I. INTRODUCTION

It was known that almost all kinds of ammal animals were bred in the matrilineal inheritance way, especially for macaques, which are in maternal gregarious way. Maybe, we can conclude that the similarity of macaques come from the same family should be higher than that from a different family. However, how to identify two likeness animals belonging to a same group is a problem. Unfortunately, nobody should run the risk of being regarded as a madman to ask a macaque," Who is your mother?" Then, some other knowledge is necessary to tell us if the two macaques have blood relationship and help the managers to judge which ones are come from the same family among a large group in animal protection area.

Face recognition has attracted numerous researchers' significant interest and attention in the pattern recognition field [1]. At the same time, numerous methods could be used to deal with face recognition. Some of these methods work perfectly well. Macaque faces are similar to human and some mature theories can help us study the similarity of animal facial feature.

Principal component analysis(PCA) is a well-known feature extraction method which is widely used in areas of feature pattern recognition, computer vision and the signal processing[2,6].It reduce datadimensionality by performing a covariance analysis among factors. When applied on to some conditions, PCA will explore correlations between samples or conditions [3,5]. One of the most useful aspects is human face recognition [4]. Because macaque also belongs to primates, the same as human beings, PCA is chosen to identify the similarity of macaque face images in this work. Meanwhile, it is necessary to realize that macaque faces are more complicated than human faces. Compared with human face recognition, there are many aspects influencing analysis effects (e.g. environment, frontal face photo acquisition and too many curled hair hiding their facial feature).

II. MEHTODS AND MATERIALS

A. Facial Image Acquisition

The experimental data is collected from Wangwushan fieldwork bases. It contains three macaque families. For each family, there are nine images coming from nine different macaques. All the photos are frontal images taken in similar environment. In addition, the images are manually cropped and re-scaled to a resolution of 320*480, RGB format.

B. Facial Image Pre-processing

Prior to the experiments, the RGB images need to be transferred to Grayscale format. Some attempts must be done to reduce the impact of the environment, such as light, distance, etc. Besides, when an image is processed by computer, it perhaps lacks fidelity because of various noises recorded by transform devices. In this work, normalization processing measures are utilized to make all face images standard according to the location of macaque eyes. Secondly, histogram equalization is employed to eliminate the impact of light to some extent. Finally, we will get a normalized 256 gray level image. Some images having been processed are presented in Fig.1.

Figure 1. Sample pre-processing images

C. Principal Component Analysis

It's clear that classification techniques can hardly be applied to learn the underlying classification rules due to high dimensionality of the features. Principal component analysis (PCA) is a generally utilized dimensionality reduction technique employing correlated variables into smaller number of uncorrelated principal components [1,5]. Considering an image as a point in a very high-dimensional space, these principal components are essentially the

eigenvectors of the covariance matrix of a set of facial images. Each individual face can be represented exactly by a linear combination of eigenfaces, or approximately, by a subset of "best" eigenfaces—those that account for the most variance [3].The main process of PCA can be stated as follows:

Firstly, each image is transformed into one dimension vector (the number of elements is L=320*480). The average of the row vector of Y images, $\overline{X_i}$, is calculated. Y is the numbers of images. Consequently, each face differs from the average by Eq.1.

$$A = X - \left[\overline{X_i} \quad \ldots \quad \overline{X_i} \right]^T \qquad (1)$$

Then eigenvector V matrix and eigenvalues vector D of covariance matrix U are calculated respectively by Eq.2.

$$[VD] = eig(U) = eig(A * A^T) \qquad (2)$$

PCA generates a set of orthonormal basis vector, known as principal component that maximize the scatter of all projected samples. The eigenvalues and corresponding eigenvector are arranged in descending order. The number of principal components k is chosen to satisfy the threshold above0.95.The eigenvector of subspace is obtained using Eq.3.

$$T = A^T V D^{-\frac{1}{2}} \qquad (3)$$

Finally, a new matrix $P = [T_1 T_2 \cdots T_k]$ is formed, called subspace matrix. For any facial image of the macaque, we can map it into the subspace through matrix operations by the following equation. Some samples can be seen in Fig.2.

$$Y = P^T * X \qquad (4)$$

Figure 2.   Reconstructed images from subtracted eigenvectors

Figure 3.   The block of one image

D.   Block PCA

There are some difficulties in explaining the results by the traditional PCA method when doing similarity analysis, especially when we have to give out the concrete differences between two samples. Block PCA is a reasonable resolution in order to overcome these difficulties by giving us the differences of diverse blocks [7,10]. The PCA coefficients of image blocks reflect some important and robust features of an image. Block PCA [8] is used to in this paper. The main steps of Block PCA are described as follows.

Block: Let $\left[X_1 X_2 \cdots X_N \right]$ be a training set including N face images, and the size of each image X_i is p*q. If we divide X_i into n blocks with size of $p'*q'$, we can use $X_{ij}, j = 1,2...n, i = 1,2,...N$ to represent the jth block of the ith image.

Combination: To compose the blocks from the same parts of all the n*N blocks, there are n groups of vectors for every image.

Block PCA: In this step, PCA is applied to analyze those n groups of vectors. After PCA process, we get n groups of principal component, whose eigenvalues are ranked in descending orders.

At last, n subspaces are built and then map each image to these subspaces.

A   Similarity criterion

Now a criterion to determine the degree of similarity between two macaques is necessary. However, it is difficult to deal with the high-dimensional vector directly. Cosine operation seems a good idea and the similarity criterion is defined as Eq.5.

$$b_1 = \cos(\partial_1, \partial_1) \qquad (5)$$

If the two vectors are identical, the cosine of the angle between them will be 1, and completely different, will be 0. Cosine can perfectly describe the similarity between two different vectors. The higher the cosine is, the more similar they are. For all the images, six vectors are the same type. Cosine of the angle between two vectors calculated from the same block of different images is convenient. Then, we can get six different values between two images. The final similarity criterion is proposed by selecting proper weights analysis of each eigenvalue [9] to improve the judgment accuracy and it is described as follow:

$$Z = \sum_{i=1}^{6} a_i b_i \qquad (6)$$

III.   EXPERIMENTS AND RESULTS

The proposed Block PCA approach is used to study the similarity of macaque face images among all the 27 images from three different families mentioned in section 2.1. All of the experiments are carried out on a PC machine with Core 2 Duo 2.00GHz CPU and based on Matlab version R2009b platform.

In addition, the appropriate weights selection mainly depends on a large number of calculation results. In this paper, three kinds of weights values are determined, including the value of eyes-block(Fig.3 top) which is 0.5, nose-block(Fig.3 middle) which is 0.3 and the remain(mouth-block,Fig.3 bottom) which is 0.2.

A. Similarity Distribution

At first ,the similarity is calculated in pairs in the same family to get similarity distribution shown in Fig.4. M1,M2···M9 respectively represent 9 macaques images for the first family. The same analysis has been done for other two families and 108 paired similarity values are obtained in three families, signed set A. The statistical results are presented in Table1. Then, the similarity coming from two families are calculated in order to investigate the difference among different families. There are 243 paired similarity values need to be calculated according to permutation and combination principle, signed set B. Finally, two groups of paired similarity values respectively from same and different families are represented with a line chart in Fig.5. It is obvious that the similarity values of macaques with consanguinity are higher than that from different families. The key problem is to find a proper threshold of similarity to help us distinguish the origin of macaques.
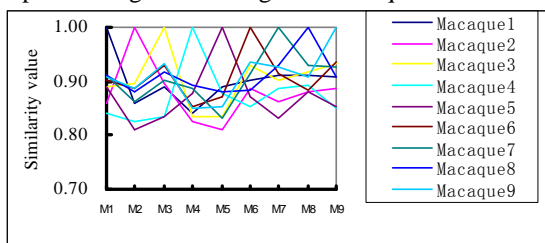

Figure 4.   The similarity distribution from family1

TABLE I.   STATISTICAL RESULTS IN THE SAME FAMILY

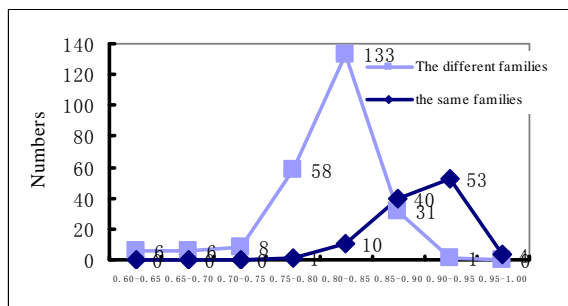|	Similarity of the maximum	Similarity of the minimum	Similarity of the average	Similarity variance
Fmaily1	0.9353	0.8247	0.8907	0.0011
Family2	0.9651	0.8450	0.9144	0.0008
Family3	0.9461	0.7665	0.8817	0.0015

Figure 5.   Similarity distribution line chart

B. The Determination of the Threshold Value

It's apparent that a threshold value should be set to judge if two macaques belong to a same family. Accuracy rate of judgment will be influenced remarkably when different

threshold value Z are selected. If the similarity value is higher than Z, the two macaques are considered to come from the same family. Conversely, when the value is lower than Z, they will be regarded as coming from different families. As shown in Fig.7, the similarity between 0.80-0.90 is a range of threshold choosing. Then Z is started with 0.80 and increases up to Z =0.90 with a step of 0.25. The effects of accurate rate of recognition are demonstrated in Table 2.

In this study, Z=0.85 is set for its superior performance both the same family and different families. However, things will not be as simple as it looks. The T-test analysis must be made to verify whether the difference between same and different families is enough significant for the researchers to do correct judgment. T-test analysis is utilized to calculate the significance between set A and B. The result is as follow, t=2.0017, df=349, p<0.01, which means that they have significant difference. It is proper to utilize similar value as judgment criterion of family properties and higher accurate rate is obtained in Table 2(Z=0.85).

TABLE II.   ACCURACY OF RECOGNITION CORRESPONDING TO THRESHOLD VALUE Z

	Z=0.825	Z=0.850	Z=0.875	Z=0.900
Family1	97.22%	83.33%	69.44%	36.11%
Family2	100.00%	97.22%	94.44%	66.67%
Family3	97.22%	88.89%	69.44%	55.56%
Different family	56.48%	89.81%	99.53%	99.53%

IV. CONCLUSIONS

In this wok, an efficient method for judging the origin of two likeness macaques has been proposed. An optimal threshold value was obtained by combining with the theory of Block PCA and weighted analysis. The experiment proved that this method is effective.

In addition, some similarity is out of the normal error range due to the following reasons.

1. The maternally inherited macaques are also subjected to the impact of paternal;
2. The impact of the original maternal gene is diminishing after biological continuation of many generations.
3. The photos perhaps lack fidelity because of the limitation of outdoor environment.
4. The number of high quality images is less.

The supplement of large numbers of photos and the improved Block PCA training method will be studied further in order to improve the classified accuracy.

REFERENCES

[1] Guan-Chun luh, Face recognition using pca based immune networks with single training sample per person, International Conference on Machine Learning and Cybernetics, Guibn,(2011) 1773-1779.

[2] Ying wen,Pengfei Shi, Image PCA : a new approach for face recognition, International Conference on Acoustics, Speech and Signal Processing I1 (2007) 241-244.

[3] Tahia Fahrin Karim ,Molla Shahadat Hossain Lipu, Face Recognition Using PCA-Based Method, International Conference on Advanced Management Science (2010) 158-162.

[4] Bruce Pooni,M.Ashraful Amin, Hong Yan, PCA based face recognition and testing criteria, International Conference on Machine Learxting and Cybernetics (2009) 2945-2949.

[5] T.Phiasai,S.Arunrzmngrusmi,and K.Chamnong, Face recognition system with PCA and moment invariant method, International Conference on Circuits and Systems (2001) 165-168.

[6] Chunming Li, Yanhua Diao, Hongtao Ma, Yushan Li, A statistical PCA method for face recognition, Second International Symposium on Intelligent Information Technology Application (2009) 376-380.

[7] Wei Zheng, Yan Zhang, A novel improvement to PCA for image classification, Computer Science and Service System(2011) 1964-1967.

[8] Shumin Ding, ChunLei Li,ZhouFeng Lium, A novel PCA-based authentication watermarking, International Conference on Information, Communications and Signal Processing(2009)1-5.

[9] Zhihua xie, Guodong Liu,Zhijun Liang, A Weighted Infrared Face Recognition Method Based on Block-DCT, International Conference on Image and Signal Processing(2010) 60-63

[10] B. Qiu, V. Prinet, E. Perrier, O. Monga, Multi-Block PCA Method for Image Change detection, International Conference on Image Analysis and Processing (2003) 385-390.