

## Research on face recognition method based on PCA

Tingting Chen<sup>1, a</sup>, Riuan Liu<sup>1, b</sup> and Lirong Diao<sup>1, c</sup>

<sup>1</sup>School of electronic and communication engineering, Tianjin Normal University, Tianjin300387, China.

<sup>a</sup>13002219230@163.com, <sup>b</sup>ruianliu@sina.com, <sup>c</sup>diaolirong0716@126.com

**Keywords:**Principal component analysis (PCA),K-L transform.

**Abstract.** Principal component analysis (PCA) is one of the most widely used face feature extraction methods, and has evolved a lot of new algorithms, which has become a hot research topic. It<sup>[1]</sup> is a multivariate statistical method, can effectively reduce the dimension of the face image, and can keep the original data of most of the major information, has been widely used in the field of pattern recognition and computer vision. This paper introduces the basic principles of PCA, as well as the improved PCA algorithm, and finally the simulation experiments are carried out.

### 1. Introduction

The principal component analysis (PCA) method was first introduced by Pentlant<sup>[2]</sup> in 1991 and applied in the field of face recognition. In essence, it is a kind of network recursive implementation of K-L transform<sup>[3]</sup>, its central idea is dimensionality reduction and the main principle is extracting the main features of a low dimensional subspace on the original data in high dimensional space constitutes, it contains as much information which can better represent the original data. All the data can be processed to the low dimensional feature subspace projection, which can reduce the data redundancy to achieve the goal of dimension reduction, and solve the problem of high dimensionality of data space<sup>[4]</sup>. Different from other orthogonal transforms (such as Fourier transform, DCT transform, discrete K-L transform), the K-L transform is a kind of change based on the statistical features of the image, the energy is most concentrated in the transform domain, its data compression has good effect, high encoding efficiency and small errors, widely used<sup>[5]</sup>.

The basic idea of K-L transform is as follows:

Assumption  $x$  is a random vector of  $N \times 1$ , is that each element  $x_i$  of  $x$  is a random variable. The mean values ( $m_x$ ) of  $x$  can be estimated using  $k$  which is a sample vector:

$$m_x \approx \frac{1}{k} \sum_{i=1}^k x_i \quad (1-1)$$

Covariance matrix can be estimated by the formula (1-2)

$$S_x = E[(x - m_x)(x - m_x)^T] \approx \frac{1}{k} \sum_{i=1}^k x_i x_i^T - m_x m_x^T \quad (1-2)$$

According to the formula (1-2), the covariance matrix is a real symmetric matrix of  $N \times N$ , the diagonal elements are the variance of each random variable, the non diagonal elements are the covariance of the random variables. K-L transform using the matrix  $A$  to define a linear transformation, it can be arbitrary vector by the following linear transformation to get a vector  $y$

$$y = A^T (x - m_x) \quad (1-3)$$

The lines of  $A$  is the feature vector of  $Sx$ . First of all, the feature values of  $Sx$  are arranged in order of size, each feature value corresponds to a feature vector, and the feature vectors are arranged according to their corresponding feature value. Because the vector  $y$  from K-L transform has zero mean characteristic and the covariance matrix of  $Sx$  can be got by  $y$ .

$$S_y = A^T S_x A = \Lambda = \begin{pmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_N \end{pmatrix} \quad (1)$$

-4)

$S_y$  is a diagonal matrix whose diagonal elements are feature values of  $Sx$ . According to the formula (1-4), it is known that each element of the vector  $y$  is not related to each other, so that the linear transformation  $A$  to remove the correlation between the variables. Because the characteristic matrix of the real symmetric matrix is unitary matrix, the K-L inverse transform is

$$x = Ay + m_x \quad (1-5)$$

Using the K-L transform, can select just for larger feature value corresponding feature vectors a subspace and for those small feature values and the corresponding eigenvectors because it contains a face image information is less, so we can give up it. For a smaller set of feature values, it depends on the proportion of all the feature values. This can not only reduce the dimension of the transformed vector  $y$ , but also maximize the retention of the main information, which can be approximated and reconstructed vector  $x$ . The front  $M$  lines ( $M < N$ ) of a matrix  $A$  is composed of the new matrix  $A$  of  $M \times N$ , the vector  $y$  is reduced to  $m \times 1$  dimension from vector by matrix  $A$

$$\hat{y} = \hat{A}^T (x - m_x) \quad (1-6)$$

The approximate reconstruction of  $x$  is

$$\hat{x} = \hat{A} \hat{y} + m_x \quad (1-7)$$

The mean square deviation of approximate reconstruction is

$$\sigma = \sum_{i=1}^N \lambda_i - \sum_{i=1}^M \lambda_i = \sum_{i=M+1}^N \lambda_i \quad (1-8)$$

## 2. Improvement of PCA algorithm

PCA method makes full use of the two order statistics in the image data, and does not consider the higher order statistics, and get the best description of the feature, rather than the best classification feature. So when the image within class differences, and the difference is relatively small, part of the face images are from other classes, and deviate from their classes, lead to erroneous recognition. In order to solve this problem, the linear discriminant analysis ideas into the PCA algorithm, the PCA algorithm is improved, firstly according to the subspace PCA algorithm, linear discriminant analysis in this subspace. The characteristics of the face abandon mainly reflects the within class differences, differences between class feature vector face selection mainly reflect the composition a new subspace. The algorithm is described in detail as follows:

Set face training samples is  $\{T_i | i = 1, \dots, M\}$ , and  $M$  is the total number of samples in the training sample set, use the PCA method to obtain subspace  $U = [u_1, u_2, \dots, u_k]$ . In order to improve the classification accuracy, it is necessary to select the feature vectors which reflect the difference between the classes.

Let  $c$  be the number of training samples,  $c_k$  is the number of samples belonging to the  $k$  class,  $\bar{m}_k$  is the average vector of class  $k$  sample,  $\bar{m}$  is the average vector of all samples,  $S_W$  is the class scatter matrix corresponding to within class scatter matrix of the  $i$  feature vector,  $S_B$  is the class scatter matrix corresponding to between class scatter matrix of the  $i$  feature vector.

$$Var_{inter}(U_i) = \det(S_B) = \det\left[\sum_{k=1}^c (U_i^T \bar{m}_k - \bar{m})(U_i^T \bar{m}_k - \bar{m})^T\right] \quad (2-1)$$

$$Var_{intra}(U_i) = \det(S_W) = \det\left[\sum_{i=1}^c \sum_{t \in c_k} (U_i^T T_t - U_i^T \bar{m}_k)(U_i^T T_t - U_i^T \bar{m}_k)^T\right] \quad (2-2)$$

$$Q = \max \left( \frac{Var_{inter}(U_i)}{Var_{intra}(U_i)} \right) \quad (2-3)$$

$$a = \frac{\sum_{i=1}^q \lambda_i}{\sum_{i=1}^M \lambda_i} \quad (2-4)$$

In the formula (2-1),  $Var_{inter}(U_i)$  is the inter class difference represented by the first  $i$  feature vector  $W_i$ , and the formula (2-2),  $Var_{intra}(U_i)$  is the intra class difference represented by the first  $i$  feature vector  $U_i$ . According to the value of  $Q$  in the formula (2-3), the characteristics of PCA value obtained in ascending order, Select  $Q$  feature vectors which mainly reflects the inter class difference constructed a new subspace by formula (2-4). Because the method is a linear analysis in one dimensional cast shadow space, it avoids the singularity of the class scatter matrix and solves the small sample problem in LDA.

According to the above, the basic steps of face recognition using the improved PCA method are as follows:

- (1)  $p$  feature vectors are obtained from the sample set using PCA;
- (2) Calculation  $Var_{inter}(U_i)$  and  $Var_{intra}(U_i)$ ;
- (3) Select  $Q$  feature vectors which mainly reflects the inter class difference constructed a new subspace  $U$ ;
- (4) The test image to the subspace  $U$  projection, according to the SVM identification.

### 3. Experimental results and analysis

The effect of improved PCA method on image recognition rate is studied in this experiment. The image preprocessing using wavelet to take a low frequency information to identify, classification using SVM, the recognition rate is shown in figure 3.1:

Table 1 Effect of improved PCA method on image recognition rate

a	0.001	0.005	0.01	0.02	0.1	0.2	0.8	0.9
Number of feature vectors	3	15	22	39	97	126	190	192
Recognition rate %	26.5	74.5	79	91.5	93	93	94	94

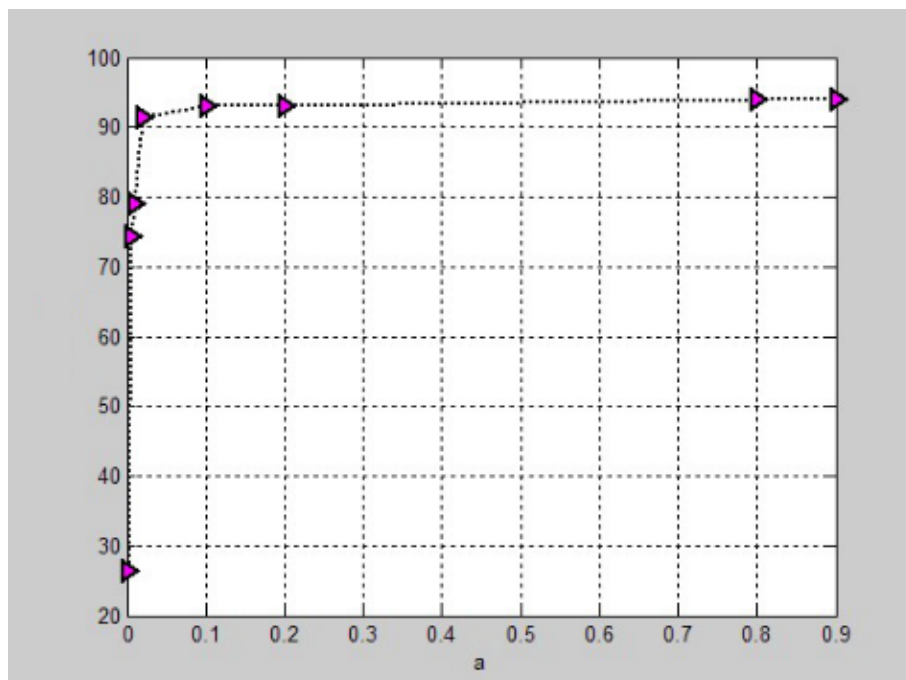


Fig. 3.1 Effect of improved PCA method on image recognition rate

The experimental results show that the recognition rate changes with  $a$ , when the  $a$  value to a certain extent, the recognition rate is no longer significant changes in a value, only small change is especially obvious, this is because the most characteristic vector difference between classes in the front row. It can be seen that the improved PCA method can achieve very good recognition results.

#### **4. Summary**

This paper mainly introduces the PCA algorithm from the aspects of the principle and the basic idea of the K-L transform. In order to solve the difference between classes is small, recognition error problem, the linear discriminant analysis ideas into the PCA algorithm, the PCA algorithm is improved, and do the simulation experiment, the experimental results show that the algorithm recognition effect is very good.

#### **Acknowledgments**

This research was supported by the Tianjin natural science fund(13JCYBJC15800).

#### **References**

- [1]Turk M, Pentland A P,Eigenfaces for Recognition, Journal of Cognitive Neuroscience,1991, 3(1):71-86
- [2]Turk M A,Pentland A P.Eigenfaces for recognition[J].Cognitive Neuroscience,1991,3(1):71-86.
- [3]K. Fukunaga, Introduction to Statistical Pattern Recognition. New York: Academic Press,Inc. 1990
- [4]M.Pantie, LJ.M. Rothkrantz.Automatic analysis of facial expressions:The state of the art.IEEE Trans.Patten Anai.Mach.Intell., 2000, 22(12):142-1445
- [5]G.H.Golub,CF.Van Loan.Matrix Computations[M],2nd Edition,The Johns Hopkins University Press,Baltimore,1989