

## An application of KPCA and SVM in the human face recognition

Feng Yue, Meng Qing Song and Yuan Hai Bo

*School of Automation, Harbin University of Science and Technology HUST  
Harbin, China  
mq50530@163.com*

### **Abstract**

*Face Recognition technology is a very important biological feature Recognition technology. Face Recognition is more and more researchers' attention, especially the principal Component Analysis method (Principle Component Analysis, PCA) after the application of Face Recognition, Face Recognition application domain expands unceasingly in daily life, such as immigration, entrance guard system, the Olympic security, airport security checks, etc. Although the face recognition system has better recognition effect has been achieved, but still by illumination, posture, facial expression change, hairstyle, with or without glasses, and the influence of various factors such as aging. Therefore, in this paper, the study of face recognition technology, has important theory significance and practical application value. A face recognition method that based on KPCA and SVM is proposed in this paper.*

**Keywords:** *human face recognition; Support Vector Machine; Kernel Principal Component Analysis*

### **1. Introduction and Background**

The face recognition is a very active and hot research direction in the field of biometrics, and plays an important role in the field of modern security and anti-terrorism. Face features are the recognizing medias for a face recognition system, which collect the local detail features and facial contour to apply the identification and matching method. At present, it has widely been used in the aspects of identification, access control and other fields. Face Recognition is a challenging interdisciplinary frontier, at present many of the methods are still in research stage, the recognition efficiency is still far from the practical. The biggest problem we are facing right now is to develop a automatically recognize faces in unconstrained environment system faces to deal with the influence of illumination change and posture change. Face recognition system has many problems to be solved:

(1) The face detection and location. Due to the various face model and the uncertainty in the process of image capture, face in the image space distribution is very complex, establish precise distribution model of human face in the high-dimensional image space is a very difficult thing. To establish a reliable statistical estimation not only needs a lot of positive cases samples, but also need enough number of valid counterexamples Samples. The present trend of study is a production of Counterexample Sample and its use problem. This is the inevitable road to improve the detection accuracy ultimately. According to the needs of practical application, carry on the reasonable assumption for testing environment, such as movement, color, etc.), so as to simplify the problem. Improve the system performance is also a feasible way. It is not difficult to foresee, the comprehensive application of knowledge and statistical method are necessary to solve practical problems.

(2) Recognize faces major feature selection and extraction based on person face characteristics, that is to say, according to the large differences between different

individuals to do relatively stable measure for a person. Due to the complexity of face changing, so the feature representation and feature extraction is very difficult. Different from rigid body, face is plastic variable, and more suitable to be described by elastic model. Therefore any feature extraction method based on rigid body is difficult to achieve satisfactory results. Second, face recognition is considered to be a unique process in human visual, so the combination of physiology and psychology is very helpful. Predictably, the two aspects research results in the human visual and non rigid body of will be helpful to find the end solution for face feature extraction and description.

(3) When facial recognition is doing the identify matching, not only should consider various factors caused by the small deformation and also can't hurt the effectiveness of face recognition at the same time of the tolerance of the deformation. In addition, the practical recognition system must also consider the computational complexity. Because each of recognition method has its advantages and disadvantages. To make the variety methods effectively synthetical will be a trend of research in the future. How to combine other biometric identification system based on in order to improve the recognition efficiency is also a meaningful things. How to integrate all kinds of information which we are able to get and use them to the greatest extent is a universal research topic, and also effective means of improving the efficiency of face recognition system.

## 2. Basic Methods

The basic methods of face recognition can be divided into two categories: the overall characteristics analysis method and the characteristics analysis method. A very classic face method is put forward by Turk and *etc.*, and good results have been achieved in the application by using PCA to extract face features. However, only the second-order statistics of the image is considered in the PCA method, which cannot be extended for the data of the high order statistics when the nonlinear correlation between pixels is not ignored. It has been shown that a higher-order statistics information of one image usually includes image edges or curve's nonlinear relationship between multiple pixels. Fortunately, Kernel principal component analysis (KPCA) is suitable for the input data of the high order statistics, and can well describe the correlation between multiple pixels. So KPCA can get better effect than PCA method in image feature extraction.

### (1) statistical methods

Turk and Pentland characteristics to detect and recognize faces, and everyone can be said with a projection weight vector, the test image of weight vector and the training of the weight vector image comparison, determine which picture the closest training images and testing images. They further found that face image with the face image projection vector is obvious different, therefore, determine whether a human face in an image is presented.

### (2) neural network method

Due to the inherent mechanism of parallel computing and artificial neural network to model the distribution of global storage, so it can be used in pattern recognition, and is not affected by the mode of deformation. Due to the face recognition method of neural network can be trained stronger noise and partial defect image, the nonlinear method is sometimes more effective than the linear method. Lin, positioning is proposed for face detection, glasses and facial recognition of neural network based on probabilistic decision, such as Lee of fuzzy BP network used for face recognition was proposed, Lawrence put forward such as convolution neural network used for face recognition.

(3) multiple classifier combination method

Multiple classifiers combination has become a frontier research topic in the field of pattern recognition, and in many ways, such as character recognition, target recognition and other fields, and have achieved good application effect. There are some literature using multiple classifiers combination method to recognize faces. Classifier combination method is composed of many issues worthy of further research, for facial recognition to this category

Number is more, the training sample number less, how to choose the right portfolio guidelines to ensure a better structure is still a difficult problem to be solved.

Support vector machine (SVM) which is based on VC (Vapnik - Chervonenkis) theory, uses structural risk (SRM) principle, and can consider both the training sample errors and generalization ability. SVM has many advantages in solving nonlinear, small sample size and high dimensional pattern recognition problems. The concept of kernel function is used in SVM to simplify the classifier kernel principal component analysis, map a group of the first data input to a high-dimensional feature space, and to classify the data which cannot be linearly classified.

In this paper, the kernel principal component analysis is used to extract face feature images, and face recognition method that based on KPCA and SVM is proposed in this paper. A classifier is designed through the SVM.

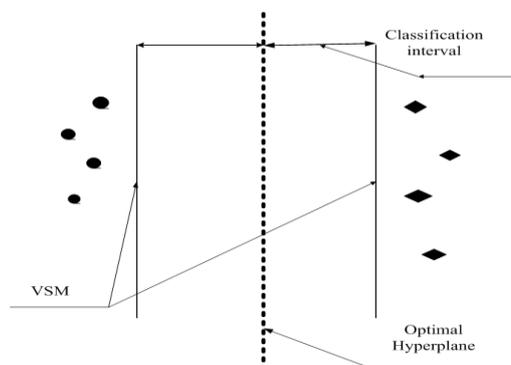
**3. Support Vector Machine**

Support vector machine, put forward in the 1990s, is a learning algorithm of statistical learning theory and has been popular for a classifier in recent years. Its basic principle is to map the original data into a high-dimensional feature space through nonlinear mapping with the structural risk minimum principle. The basic idea of VSM is to map the input feature vector into a high dimensional feature space through a sort of nonlinear mappings, and then reformulate the nonlinear problem into a linear separable problem with high dimensional feature space. From Figure 1 for several data, we can classify them into two classes with a so-called "distance" between the two classes maximized. When it is extended to a high dimension space, it is an optimal classification plane problem.

The Sample set  $(x_i, y_i)$  and  $x_i \in R^N, y_i \in \{-1,1\}$  is a class code, dividing  $x_i$  into two hyperplanes, one is  $P: y_i = +1$ , the other is  $M: y_i = -1$ , the general form of linear discriminate function in N dimensional space is:  $g(x) = \omega x + b$

So the equation of the classification plane is :

$$\omega x + b = 0 \tag{1-1}$$



**Figure 1. Optimal Hyperplane Schematic**

When the sample set is linearly separable, the discriminant function can be normalized such that the two types of samples satisfy  $|g(x)| \geq 1$ , and the nearest classification surface samples satisfy  $|g(x)| = 1$  and the classification interval equals to  $2/\|\omega\|$ .

So the classification interval maximization problem is converted to minimize ( $\|\omega\|$  or  $\|\omega\|^2$ ), and if all samples of classification plane are required to be classified correctly, the following should hold.

$$y_i [(\omega \cdot x_i) + b] - 1 \geq 0, i = 1, \dots, l \quad (1-2)$$

where  $\omega$  is the weight coefficient,  $b$  is the threshold. The optimization function is to maximize

$$\begin{aligned} \omega(a) &= \sum_{i=1}^l a_i - \frac{1}{2} \sum_{i=1}^l \sum_{j=1}^l a_i a_j y_i y_j x_i x_j \\ 0 &\leq a_i \leq C, i = 1, \dots, l \quad \sum_{i=1}^l a_i y_i = 0 \end{aligned} \quad (1-3)$$

The solution satisfying (1-3) makes  $\|\omega\|^2$  minimum and is the optimal classification plane. The optimal classification function is:

$$f(x) = \sum_{i=1}^l y_i a_i x \cdot x_i + b \quad (1-4)$$

The points in the two types of samples which are most closed to the classification plane and parallel to optimal classification plane are called the Support vector Machine.

The optimal classification plane (generalized optimal classification) under the condition of nonlinear classification can increase a flabby items in the inequality (1-2), such as

$$y_i [(\omega \cdot x_i) + b] - 1 + \zeta_i \geq 0, i = 1, \dots, l \quad (1-5)$$

to minimize 
$$\frac{1}{2} (\omega \cdot \omega) + C \left[ \sum_{i=1}^n \zeta_i \right] \quad (1-6)$$

where  $C$  is an arbitrary specified constant, which play a regulation role in the algorithm between the complexity and the proportion of error classification sample.

On the other hand, with the appropriate inner product function  $K(x_i, x_j)$ , a linear classification can be completed for a nonlinear transformation, in which the amount of computation and complexity will not increase.

With the kernel function  $K : K(x, z) \equiv \phi(x) \cdot \phi(z)$  introduced, the maximum function is as follows for the no error cases and linear separable cases:

$$W(a) = \sum_{i=1}^l a_i - \frac{1}{2} \sum_{i=1}^l \sum_{j=1}^l a_i a_j y_i y_j K(x_i, x_j) \quad (1-7)$$

from which we can obtain coefficient  $a_i$ , ( $x_i$  which is support vector machine), and

$$\omega = \sum_{i=1}^l a_i y_i x_i, b = y_i - \omega \cdot x_i \quad (1-8)$$

Then the classification of the corresponding function is:

$$f(x) = \sum_{i=1}^l y_i a_i K(x, x_i) + b \quad (1-9)$$

#### 4. KPCA Application in Human Face Feature Extraction

The main principle of KPCA is to firstly map the raw data to a feature space F, and then to do statistical principal component analysis in F, such that it can obtain the nonlinear characteristics. For any given Mercer kernel K, a mapping of the feature space can be defined as follows:

$$\phi: R^N \rightarrow F, x \mapsto K(., x), x \mapsto X \quad (2-1)$$

The centralized situation in feature space of the sample set is:

$$\sum_{k=1}^M \phi(x_k) = 0 \quad (2-2)$$

A covariance matrix can be defined in the feature space:

$$\bar{C} = \frac{1}{M} \sum_{j=1}^M \phi(x_j) \phi(x_j)^T \quad (2-3)$$

Principal components can be calculated through the eigenvalue problem of covariance matrix

$$\lambda V = \bar{C} V \quad (2-4)$$

where the eigenvalue  $\lambda \geq 0$ , and the feature vector  $V \in F \setminus \{0\}$ , and the coefficient  $a_i, i=1, \dots, M$ , satisfy

$$V = \sum_{i=1}^M a_i \phi(x_i) \quad (2-5)$$

So the equation can be written as:

$$\lambda (\phi(x_k) \cdot V) = (\phi(x_k) \cdot \bar{C} V), k=1, \dots, M \quad (2-6)$$

with (2-1) and the kernel function  $K_{ij} = (\phi(x_i) \cdot \phi(x_j))$ , the above equation can be written as

$$M \lambda K a = K^2 a \quad (2-7)$$

Due to the positive semidefinite and Symmetry of K, the equation  $M \lambda a = K a$  is required to be solved, where all the eigenvalues are non-negative. Since the principal component of the first k of the sample x (corresponding to the mapping  $\phi$ ) is:

$$(V^k \cdot \phi(x)) = \sum_{i=1}^M a_i^k (\phi(x_i) \cdot \phi(x)) \quad (2-8)$$

it can be concluded that the standardization of feature vector  $V^k$  is proceed through the  $\lambda^k (a^k \square a^k) = 1$ . Face test sample vector makes eigenvalues sorted from big to small by solving the eigenvalue equation. Choose eigenvectors of the corresponding q largest eigenvalues, and then direct map the human face testing samples to the best direction. We can get the characteristics of the testing samples.

$$(V^k \cdot \phi(x)) = \sum_{i=1}^M a_i^k (\phi(x_i) \cdot \phi(x)) = \sum_{i=1}^M a_i^k K(x_i, x) \quad (2-9)$$

show the positions of human face detection sample vector in feature space F, and it can be used as the basis for face recognition.

## 5. Experiments

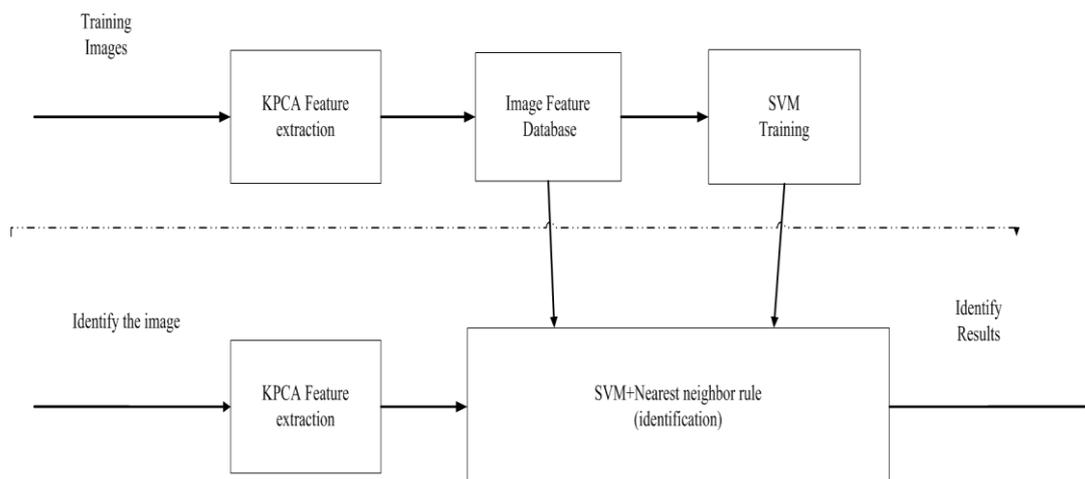
The SVM can be a Polynomial kernel function, a Gauss kernel function, a Sigmoid kernel function and others. Here a three-order polynomial kernel function is employed.

In this paper we select the ORL face database, which consists of 40 different people and each has 10 pictures (totally of 400 images). The images of these people are obtained in different times, different perspectives, different expressions (close eyes /open eyes, smiling/surprised/angry/ furious/happy) and different facial details (wear glasses/not wear glasses, beard/no bear, different hair style). Each original image has 256 grayscale, and the size is 92 x 112.



**Figure 2. Face Image**

The overall process is as shown in the following Figure 3



**Figure 3. Overall Process**

First, the kernel principal component analysis is used to extract the face images, and then the classification and recognition is done through the kernel principal component based on support vector machine (SVM) and the nearest neighbor rule.

The description for the three kinds of function in the SVM Train function is:

- 1) polynomial:  $(\text{Gamma} * (X(:, I), X(:, j)) > + \text{Coefficient}) ^ \text{Degree}$
- 2) RBF:  $(\exp(\text{Gamma} - * (: I) | X - X(:, j) | ^ 2))$ , and
- 3) sigmoid:  $\tanh(\text{Gamma} * < X(:, I), X(:, j) > + \text{Coefficient})$

The test results of three kinds of kernel function are as follows

- ① polynomial kernel function

The influence of the selection of the Gamma and C are not very clear. In the test. One set of parameters are Gamma = 50, C = 0.01, and the correct recognition rate is 93.0%.

②Gauss kernel function results

The selection values of the Gamma and C have a great influence on experimental results. In the test, when Gamma=5\*10<sup>-3</sup>, C>=1, the correct recognition rate is 92.5%.

③Sigmoid kernel function

The selection value of the Gamma has a great influence on experimental results and the value of C has little influence. In the test, when Gamma=9\*10<sup>-6</sup>, 0<C<10, the correct recognition rate is 2.5%, but when Gamma=5\*10<sup>-9</sup>, 0<C<10, the correct recognition rate is 83.5%.

By comparison, a polynomial kernel function is employed to achieve the recognition. The polynomial kernel function is in the form of  $K(x, y) = (1 + x \cdot y)^d$ . In the experiment, the order for numbers of kernel functions of d = 3, and then a comparison between the experimental results and the traditional PCA + SVM are shown in TABLE1.

**Table I. A Comparison between the Experimental Results and the Traditional PCA + SVM**

Recognition Method	Training Rate				
	1	2	3	4	5
PCA+SVM	0.56	0.75	0.76	0.80	0.88
KPCA+SVM	0.66	0.80	0.88	0.88	0.93

From the experimental data we can see an obviously improvement on the recognition accuracy than the traditional PCA. This is because of KPCA method had considered the nonlinear relation between image pixels.

**6. Conclusion**

Considering the high order correlation between image pixels, and converting the inseparable problem in low dimensional space into the linear separable problem in high dimension space, face recognition method that based on KPCA and SVM is proposed in this paper. The experimental results show that the proposed method can improve the correct recognition rate. There are two shortcomings in this paper, (1) didn't take Image illumination change into account sufficiently that could affect the recognition rate. (2) In order to improve the recognition rate, it can be started from the aspects of classification decision. Because of that the classification method based on minimum distance classification which the author used is a linear classifier. But it may get better effect in high dimensional face recognition if use this kind of nonlinear classifier of neural network. For these two types of problems, we will improve it in the follow-up work.

## References

- [1] Y. Liu, L. Zhang, S. Zhang and J. Liu, "Comparison Research of Single Kernel and Multi-kernel Relevance Vector Machine", *Compture Engineering Press*, (2010), pp. 195-197.
- [2] C. W. Hsu and C. J. Lin, "A comprehensive of methods for multiclass support vector machines", *IEEE Transactions on Neural Networks*, vol. 13, no. 2, (2002), pp. 415-425.
- [3] S. K. Shevade, S. S. Keerthi and C. Bhattacharyya, "Improvements to the SMO algorithm for SVM regression", *IEEE Trans. Neural Net-work*, vol. 11, no. 5, (2000), pp. 1188-1193.
- [4] S. S. Keerthi, S. K. Shevade and C. Bhattacharyya, "Improvements to Platt's SMO algorithm for SVM classifier design", *Neural Comput.*, vol. 13, (2001), pp. 637-649.
- [5] W. Wang, X.-Y. Sun, S. Karungru and K. Terada, "Face recognition algorithm using wavelet decomposition and Support Vector Machines", *IEEE Optomechatronic Technologies (ISOT)*, (2012), pp. 1-6.
- [6] C. N. Shawe Taylor, "An Introsuction to Support Vector machines", Cambridge, UK: Cambridge University Press, (2000).
- [7] G. Dai and C. L. Zhou, "Face recognition using support vector machines with the robust feature Robot and Human Interactive", *Communication 2003, Proceeding. ROMAN 2003. The 12th IEEE International Workshop on 31th Oct-2th Nov [C]. 2003*, pp. 49-53.
- [8] M. Jia and R. Fuji, "Detect and track the dynamic deformation human body with the active shape model modified by motion vectors", *IEEE International Conference on Beijing*, (2011) September, pp. 587-591, 15-17.
- [9] J. Jianming, H. Xionghu and W. Qingren, "Mathematical Formulas Extraction", *Proceedings of the 7th ICDAR [C]*, (2003), pp. 1138-1141.
- [10] M. H. Yang, "Kernel Eigenfaces vs. Kernel Fisherfaces: Face Recognition Using Kernel Methods", *Proc. Fifth IEEE Int'l Conf. Automatic Face and Gesture Recognition*, (2002) May, pp. 215-220.
- [11] M. Turk and A. Pentland, "Eigenfaces for Recognition", *Cognitive Neuro science*, vol. 3, no. 1, (1991), pp. 1-86.
- [12] X. Chen, P. J. Fynn and K. W. Bwyer, "CA-Based Face Recognition in Infrared Imagery: Baseline and Comparative Studies", *Proc of the IEEE International Workshop on Analysis and Modeling of Faces and Gstures [C]*, Nice, France, (2003), pp. 127-134.
- [13] P. Viola and M. J. Jones, "Robust Real-time Object Detection", Cambridge Research Laboratory, Technical Report Series, CRL 2001/01.
- [14] P. Viola and M. J. Jones, "Rapid Object Detection using a Boosted Cascade of Simple Features", *Computer Vision and Pattern Recognition*, vol. 1, (2001), pp. 8-14.
- [15] Y Ming-Hsuan, J. K. David and A. Narendra, "Detecting Faces in Images: A Survey", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 24, no. 1, (2002) January, pp. 34-35.