Face Recognition Based on Kernel Principal Components Analysis and Proximal Support Vector Machines

Li Yunfeng, Sun Lihua

College of Mechatronics Engineering, Henan University of Science & Technology, Luoyang, 471003 China

Summary
High order statistics of the original data can be handled by Kernel Principal Components Analysis, which can describe multiple correlations between pixels in the image recognition, at the same time the nonlinear features of the image can be preferably extracted. Support Vector Machine has a better ability of nonlinear mapping and a stronger generalization capability, while the Proximal Support Vector Machine is an improvement for Support Vector Machine. In this paper, the computer simulation was progressed based on ORL face database, the effectiveness of the Proximal Support Vector Machine algorithm and Kernel Principal Components Analysis algorithm were shown by experimental results.

Key words:
Kernel Principal Components Analysis, Support Vector Machine, Proximal Support Vector Machine, face recognition

1. Introduction

Face recognition with other biological characteristics is an important part of the biological characteristics, and has been the focus of many research fields. In the 1990s, Turk advanced a classic method of features face, the algorithm of which has always become the benchmark of face recognition algorithms running to now. Nowadays, good results can be obtained by using Principal Components Analysis to distill face feature extraction. What was known is not only the second order statistics information but also higher order statistics information is included in an image. However Principal Components Analysis only considers the image data of the second order statistics information, ignoring nonlinear pertinence between multiple pixels. Some research shows that sometimes nonlinear pertinence of multiple pixels of edges or curves of an image often indwells in higher order statistics. While Kernel Principal Components Analysis, a higher order statistics based on the input data, can describe the relationships between multiple pixels, so these important information can be caught by KPCA, which has achieved better performance. Meanwhile, the problem that cannot be classified by the linearity in the input space can be converted into the feature space to realize the linear classification by KPCA, which can make classifier design simpler. Support Vector Machine is based on VC dimension theory of the statistical learning theory and structural risk minimization principle, which can better solve many problems such as small sample, nonlinear, high dimension, local point tiny practical and so forth, and has nicer generalization ability, which already become a new type of machine learning method. Proximal Support Vector Machine is a new method based on support vector machine, keeping the high recognition rate of the standard support vector machine and the characteristics of small sample training, but the pace is greatly enhanced.

In this paper, Kernel Principal Components Analysis and traditional Principal Components Analysis were mainly adopted to take orders with face image in ORL face database, then the classification of faces was respectively realized imposing Support Vector Machines, the Nearest and Proximal Support Vector Machine. At last, the superiority of Kernel Principal Components Analysis and Support Vector Machine was proved through the experimental data.

2. The Theory of Kernel Principal Components Analysis

The elementary idea of Kernel Principal Components Analysis is that map the each inputting data vector X_k, $X \in R^M$ to a high dimensional of characteristic space T through a nonlinear transform φ, and then put up feature extraction by Principal Components Analysis in the feature space T.

$$\varphi : R^M \to T, X_k \to \varphi(X_k), \quad k = 1,2,\cdots,N \tag{1}$$

In the feature space T, we can assume

$$\sum_{k=1}^{N} \varphi(X_k) = 0$$

then the covariance matrix is

$$C^{'} = \frac{1}{N}\sum_{j=1}^{N} \varphi(X_j)\varphi(X_j)^T \tag{2}$$

$$\lambda v = C^{'} v \tag{3}$$

Manuscript received September 5, 2012
Manuscript revised September 20, 2012

in formula 3 the eigenvalue λ and eigenvector $\nu \in T$, moreover ν can be linearly made by $\varphi(X_i)(i = 1,2,\ldots,N)$, at the same time, $e_i(i = 1,\cdots,N)$ is existence factor , so

$$\nu = \sum_{i=1}^{N} e_i \varphi(X_i) \tag{4}$$

at the same reason these is :

$$\lambda(\varphi(X_k)\cdot \nu) = \varphi(X_k)\cdot C^{`}\nu(k = 1,2,\ldots,N) \tag{5}$$

by above three equtions can acquire:

$$\lambda \sum_{i=1}^{N} e_i(\varphi(X_k)\cdot\varphi(X_i)) = \frac{1}{N}\sum_{i=1}^{N} e_i\left(\varphi(X_k)\cdot\sum_{j=1}^{N}\varphi(X_j)\right) * (\varphi(X_j)\cdot\varphi(X_i)) \quad (k = 1,2,\cdots,N) \tag{6}$$

Define a matrix K that is $N \times N$

$$K_{ij} = k(X_i \cdot X_j) = (\varphi(X_i)\cdot\varphi(X_j)) \tag{7}$$

Then simplified the formula can obtain:

$$N\lambda\bar{e} = K\bar{e} \tag{8}$$

the required eigenvalues and eigenvectors can gained by solving formula 8.
Here choosing order number is d can get polynomial nuclear

$$K(X_i \cdot X_j) = (X_i \cdot X_j)^d \tag{9}$$

In this paper, the eigenvalues of the matrix K are shown by $\lambda_1 \le \lambda_2 \le \ldots \le \lambda_N$, the eigenvectors that is corresponding with eigenvalues can be expressed as $\xi_1,\xi_2,\ldots,\xi_N$, and λ_p is remembered as the first nonzero value. When $\xi_1,\xi_2,\ldots,\xi_N$ are made standardization, there is a condition:

$$(\nu^k \cdot \nu^k) = 1, k = p,\ldots,N \tag{10}$$

then get the unitary condition of $\xi_1,\xi_2,\ldots,\xi_N$:

$$I = \sum_{i,j=1}^{N} e_i^k e_j^k (\varphi(X_i)\varphi(X_j)) = \sum_{i,j=1}^{N} e_i^k e_j^k K = e^k \cdot Ke^k = \lambda_k e^k e^k \tag{11}$$

The projections of the test samples in T Space vector V^k is

$$(V^k \cdot \varphi(X)) = \sum_{i=1}^{N} e_i^k (\varphi(X_i)\cdot\varphi(X)) = \sum_{i=1}^{N} e_i^k K(X_{i,}X) \tag{12}$$

From above formula, it is known that importing the kernel function is great convenient for computing the ingroup i in the high dimension space.
The KPCA algorithm is summarized as follows:

1. Select kernel function $K(x,y)$, and go along centering on the high dimensional space, then calculating the matrix $\tilde{K}$;
2. Calculate eigenvalues and eigenvectors, and then normalize them in the high dimension space;
3. Calculation the nonlinear principal components of test samples.

3. The Basic Principle of Proximal Support Vector Machine

Proximal Support Vector Machine is based on Support Vector Machine, it is simpler and faster than traditional Support Vector Machines algorithm, which is especially suitable for large amounts of data classification and operations.
On assumption that there are N training samples, such as $(x_1,y_1),(x_2,y_2),\cdots(x_N,y_N)$, among them $x_i \in R^m, y_i \in \{-1,1\}$, so the target function of Proximal Support Vector Machine can be denoted by

$$\min_{(w,r,y\in R)} \frac{c}{2}\|y\|^2 + \frac{1}{2}(w^T w + r^2) \tag{13}$$

$$\text{subject to : } L(Aw - er) + y = e \tag{14}$$

in formula 14, c is the castigation factor, y express the sample output, w figure the normal vector of the classification hyperplane, e is the units vector, g is the parameter which can ascertain the position of two dividing-line plane relating to the origin in Proximal Support Vector Machine; A is $n \times m$ dimensional training data set, each sample is corresponding to a list A_i ;

L is a sort of A_i, L also is the diagonal matrix, and its diagonal elements showed by + 1 or-1 mark a row of A, namely the sort of the each and every sample points, viz.

$$L_{ii} = \begin{cases} 1 & A_i w \geq r+1 \\ -1 & A_i w < r-1 \end{cases} \tag{15}$$

The solution of PSVM is solved by Lagrange:

$$H(w,r,y,u) = \frac{c}{2}\|y\|^2 + \frac{1}{2}\left(w^T w + r^2\right) - u^T[L(Aw-er)+y-e] \tag{16}$$

in formula 16, $u \in R^m$ express Lagrange arithmetic operators. w of formula 16 is replaced by its dual form $A^T Du$, and linear kernel matrix $K = K\left(A, A^T\right)$ substitute linear kernel matrix AA^T, then the target function of nonlinear PSVM can be got:

$$\left.\begin{aligned} &\min_{(w,r,y\in R)} \frac{c}{2}\|y\|^2 + \frac{1}{2}\left(u^T u + r^2\right) \\ &subject \quad to \quad L\left(K\left(A, A^T\right)Lu - er\right) + y = e \end{aligned}\right\} \tag{17}$$

and then the discriminant function of nonlinear PSVM can be got

$$Y = P(A) = \operatorname{sgn}\left\{K\left(A, A^T\right)K\left(A, A^T\right)\right\} \tag{18}$$

4. Design of Experiments and Results Analysis

This paper based on the normal ORL face image database which is diffusely used in the world, where are 40 people, and each one has 10 images that the pixel of every image is 100×100, so a total of 400 images. These images gathered in different time, which have the different degree of the change of expression and posture. For example, somebody is wearing glasses while another is unwearing, someone'eyes is open someone's is closed and so on. Figure 1 show a part of the sample images of the database.

Figure 1 a part image of ORL database

In this experiment, 10 images of each person are divided into two groups, five sheets are used as the training set, other five sheets are used as the testing set, so there are 200 face images for training set and testing set respectively. This experiment respectively use the traditional PCA and KPCA to distill feature of face images, and here polynomial kernel function $k(X_i \cdot X_i) = (X_i \cdot X_j)^d$ is adopted in order to simplify calculate, different d are taken for this test. Simultaneity, we define that the ratio of the recognition rate and the CPU average counting time is c, the bigger , the more superior, which show the way that can rather coach actual production.

After attaining the characteristics of face images through the traditional PCA and KPCA, we respectively utilize Support Vector Machine, the Nearrest classification and Proximal Support Vector Machine to classify face feature. The experimental results are shown in figure 1and table 1.

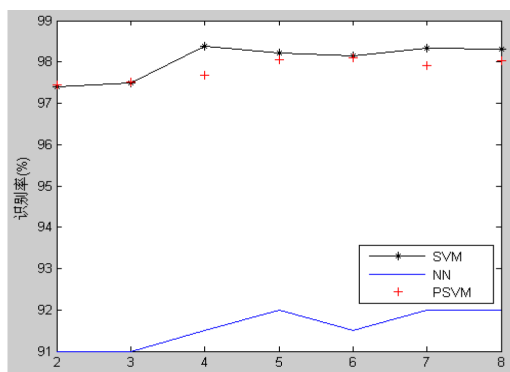

Figure 2 the recognition of KPCA in different d and different classification

In Fig.2, the different identification results are got by choosing the different kernel function parameter d to classify, that's to say optimization of the parameters can improve the recognition rate. In the same classification method, the different kernel function get the different recognition rate, such as when we adopt SVM and the order number of kernel function is four, when we use NN classification and the order number of kernel function is 5 and 7, and when we adhibit PSVM to classify and the order number of kernel function is 6, we can correspondingly get the highest recognition rate. And in the whole of the recognition rate, when we adopt SVM and the order number of kernel function is four, the recognition rate is the highest rate. But the recognition rate of PSVM is always closed to the rate of SVM.

Table 1 the compare of the capabilities

method	***recognition rate (%)***	***CPU average count time (s)***	***ratio*** c
KPCA+NN	92.39	70.4	1.31
PCA+NN	88.5	9.3	9.51
KPCA+SVM	98.38	120	0.91
PCA+SVM	92.86	90.5	1.02
KPCA+PSVM	97.56	23.6	4.13
PCA+PSVM	92.45	8.4	11.01

In Tab.1, in the same classification method, the recognition effect of KPCA is always better than the traditional PCA, because KPCA considered the nonlinear relationship among the pixels. The CPU average count time of the traditional PCA is always shorter than KPCA's, the ratio c of the traditional PCA is bigger than KPCA, so PCA is more suitable for coaching actual production than KPCA. To the same way of feature pick-up, PSVM does not only get higher recognition rate than NN, but its recognition rate is colsed to the recogniton rate of SVM; and in three classification methods, the CPU average count time of PSVM is the shortest, c is the biggest, the CPU average count time and c of SVM take second place, the CPU average count time of NN is longest, c is the smallest, which proved PSVM is the best claasification method in three classifications, it holds the high recognition rate of SVM, as well as the CPU average count time it spent is the shortest.

5.Conclusion

This paper makes full use of the high order related information among each pixel of the face image, then transforms the inseparable problems of the low dimensional space to the linear separability problems of the high dimensional space. Through the above experiment we can see, the different identification results can be got when we choosed the different types of the kernel function parameters to classify, which proved that recognition rate can be improved with the optimization selection of parameters. And the recognition rate of KPCA algorithm is higher than the traditional PCA algorithm, which proved KPCA is more effective than PCA. The recognition rate of PSVM is on the verge of SVM's, which
spend the shortest CPU average count time, so its practicality is better than SVM and NN.

References

[1] Turk M, Pentland A. Eigenfaces tor Recognition Journal of Cognitive Neuroscience, 1991, 3(I): 71-86

[2] Scholkopf B, Sinolla A, Muller K. Nonlinear Component Analysis as a Kernel Eigenvalue Problem. Neural Computation,1998,10(5):1299

[3] Yong Xu, Qiang Yang, Jingyu Yang. Based on the rapid feature extraction of kernal and recognition. J. Journal of the PLAuniversity of technology, 2005.

[4] Lili Li, Yimin Li, Ying Cai. The application of KPCA and SVM in face recognition. J. Journal of chemical industry, 2006, 5: 44-46.

[5] Hichem Sahbi.Kernel PCA for similarity invariant shape recognition. J. ScienceDirect Neurocomputing 70(2007)3034-3045.

[6] Guoqi Cui, Feng Jiao and Shiguang Shan.Face Recognition Based on Support Vector Method. R. The 5th Asian Conference on Computer Vision, Melboume, Australia, 23-25 January 2002.

[7] ZHAO W, Krishanaswamy A, Chellappa R, et al. Discriminant analysis of principal components for face recognition[C]//3rd IEEE International Conference on Automatic Face and Gesture Recognition. Washington, DC: IEEE Computer Socitey, 1998: 336-341.

[8] Roman Rosipal, Mark Girolami.Kernel PCA for Feature Extraction and Denoising in Nonlinear Regression. J. Neural Computing &Application, 2001, 10(3):231—243.

[9] Guohui He, Junying Gan. Face recognition based on Based on the Kernal Principal Component Analysis and Support Vector Machine. J. Computer engineering and design, may, 2005.

[10] Xi Zhang, Weiwu yan, Xu Zhao, Huihe Shao. The performance monitoring and fault diagnosis based on wavelet denoising nuclear principal component analysis and the neighboring support vector machine. J. Journal of Shanghai jiaotong university, 2008