Face Recognition Method Based on 2DLDA and SVM Optimated by PSO Algorithm

Dan Zou[1,a], Hong Zhang[2,b]

[1]College of Computer Science and Technology, Wuhan University of Science and Technology, China, 430065

[2]College of Computer Science and Technology, Wuhan University of Science and Technology, China, 430065

[a]zoudanwust@163.com, [b]46476522@qq.com

Keywords: Wavelet transform; Two dimensional linear discriminant analysis; Particle swarm optimization; Support vector machine; Face recognition

Abstract. Concerning the "Small Samples Size" problem in LDA algorithm and reduce the effects to the SVM face recognition rate caused by random parameters set by human. An algorithm based on combination with the PSO algorithm which was originated form artificial life and evolutionary computation to SVM's parameters election and optimization, and Wavelet Transform , two-dimensional LDA(2DLDA) was proposed. Firstly, the original images were decomposed into high-frequency and low-frequency Components by Wavelet Transform (WT). The high-frequency components were ignored, while the low-frequency components can be obtained. Then, the liner discriminant features were extracted by two-dimensional LDA (2DLDA). Finally, we use the PSO algorithm to SVM's parameters election and optimization. Experimental results based on ORL face database show the validity of the algorithm this paper proposed for face recognition and it can reach the recognition rate of 98%.

Introduction

The success of face recognition depends on the reasonable feature extraction and effective classifier [1].The traditional face feature extraction methods mainly include two categories: One is based on the geometric feature extraction method. To extract the face facial feature points (such as eye, nose, mouth, etc.) of the relative position and distance .As a result of these geometric features easy affected by illumination, expression, posture change, the stability is bad. The other one is based on statistical feature extraction method .The principal component analysis (PCA)[2] and linear discrimination analysis (LDA)[3] as two powerful tool of feature extraction and data description has been widely used in face recognition.

When the PCA and LDA are used to deal with human face image, firstly we describe face image matrix into a high dimensional vector. This will bring the computational complexity and large matrix singular problem. To solve this problem, people put forward the method directly based on 2 d vector of face image matrix without vector quantization of 2DPCA and 2DLDA method [4] .It effectively solve the problem of the computational complexity and matrix singular. The 2DLDA algorithm comes from the LDA algorithm and it can effectively solve the LDA algorithm of "small sample problem".

On the other hand, the SVM method with good generalization ability proposed by Vapid in 1995 as a new machine learning method [5] and has been the hotspot in the field of Machine Learning in the selection of classifier. The SVM used nuclear technology and the key kernel function is introduced in the Algorithm. So the kernel function plays a very important role. According to the Mercer theorem in the functional, looking for a kernel function that can transform inner product in the sample space to the inner product space in the transformation space, namely avoid nonlinear mapping for inner product [6]. The selection of kernel parameters directly affects generalization ability of kernel function. The commonly used kernel functions are: linear kernel function; Polynomial kernel

function; the radial basis kernel function and so on. SVM kernel function and parameters selection is one of the problems it is necessary to reform the development of SVM. Its essence is a optimization search process, and it directly affect the generalization ability of the model. For the same classification problem, choosing a different kernel functions or the same nuclear parameters of different kernel function has different classification results. Due to a lack of theoretical guidance, most of the traditional nuclear parameters are selected through repeated experiments by human to get a satisfactory solution.

Therefore, to solve the problem that the SVM parameter selection is easily affected by artificial settings, the particle swarm algorithm (PSO) is proposed to be applied in the SVM parameters optimization and selection. Particle swarm optimization algorithm is a kind of evolutionary computation technology based on the group intelligent and first put forward by Kennedy and Eberhart [7].The particle swarm optimization (PSO) by using particles to follow the optimal particles to search in the solution space. To find the optimal parameters by PSO, it can avoid the influence of the setting the parameters by artificial.

This paper proposed a face recognition algorithm for feature extraction using wavelet transformation and 2DLDA and using PSO to optimize the support vector machine (SVM) for classification. Not only effectively remove high frequency signal interference, but also solve the problem of matrix singularity and small sample problem, and it overcome the effects of the artificial setting parameters of the SVM, improved the recognition rate. The experimental results verify the efficient algorithm.

Wavelet transform and 2DLDA for Facial Feature Extraction

This paper uses the wavelet transform and 2DLDA algorithm for facial feature extraction. Tiny change of the expression   and   a little light changes affects only the local flow of light intensity, if we use frequency to express it can   only affects the high frequency part[8].Therefore, after removal of face image wavelet decomposition of the high frequency component, only low frequency component to express human face image will obtain good effect. L represent the face image, A represent facial features after wavelet transform. And because 2DLDA algorithm uses the image matrix to construct the matrix of discrete degree and seek for minimum discrete degree within the class, and the largest projection matrix between the data to feature extraction. Therefore it has the good class identification ability and it has been widely used in the field of face recognition. Characteristics of human face are used to extract two dimensional linear differential features further and we can get the optimal projection matrix W_{opt} and characteristic matrix. The two feature extraction methods can not only reduce the image of high frequency interference and can solve the problem of matrix singularity and small sample problem, improve the efficiency of feature extraction. The wavelet transform and 2DLDA will be introduced respectively next.

Wavelet Transform. Wavelet transform is space (time) and frequency of local transformation, and it can effectively extract information from signal [9] .

The wavelet transform is also known as "mathematical microscope" because of that it can be used for Multi-scale refinement analysis of function or signal by operation function such as expansion and translation. Through wavelet transform, the image signal is decomposed into much different spatial resolution, different frequency and the direction sub band image signal. When representing an image signal, its binary wavelet transform is equivalent to use different one-dimensional filter respectively along the x direction and y direction to get low frequency and high frequency sub band images. Wavelet transform as shown in Fig. 1:

LL	HL
LH	HH

L / H		H
L / H		
L		H

(a) one-layer Wavelet transform   (b) multi-layer Wavelet transform

Figure 1.   Schematic diagram of wavelet transform

In figure 1,LL represents low frequency information of image in the horizontal and vertical directions, HL represents low-frequency information in the vertical and high-frequency in the horizontal information of image, and LH represents low-frequency in the horizontal and high-frequency in the vertical information of image, HH represents and high frequency in the horizontal and vertical information of image. The low frequency information LL can be repeated decomposition to get triple decomposition in the horizontal and vertical direction. As shown in figure 1 (b), after the wavelet transform, if the image has obvious features in a certain frequency and direction, the output of the matching sub band image has a large energy. So the image information focuses on the few wavelet coefficients. After wavelet transform, as long as you we use suitable scale wavelet function, we can observe signals in any scale. Ignoring the part of the high frequency component can eliminate noise and achieve the goal of effective dimensionality and features extraction.

Two-dimensional Linear Discriminant Analysis. Linear discriminant analysis (LDA) is an important pattern recognition method of supervised learning. The basic idea of LDA is to seek the best projection space so that it can map the face samples to the projection and maximize the samples within the discrete degree between class and class of the ratio between discrete degrees [10]. When using LDA method for face recognition, we often transform the two-dimensional image matrix into a one dimensional vector so that it make the dimensions of the samples increased rapidly and it brings great difficulty to the LDA feature extraction. In addition, the LDA is prone to create small sample problem, and the discrete degree matrix is usually singular so it makes hard for the LDA. In order to solve the LDA defects in feature extraction. Two-dimensional linear discriminant analysis (2DLDA) method put forward By Li [11] directly used two-dimensional face image matrix for feature extraction, decreases the discrete degree matrix dimension and it greatly reduces the computation time of characteristic vector, and fundamentally solves the small sample problem.

In the representation of human face image, the LDA algorithm usually used a vector, and the 2DLDA algorithm used a matrix. Assuming that the C category is expressed as $A_1, A_2, A_3,... , A_c$, each type of sample have n samples, the total is n. A_{ij} represent m x l d matrix, and it also represent the $i(i = 1, 2, 3,..., C)$ of the class the j $(j = 1, 2, 3,..., n)$.The average of C samples total average is $\overline{A^i} = \frac{1}{n}\sum_{j=1}^{n} A_{ij}^{j}$ the total mean of C class sample is $\overline{A} = \frac{1}{N}\sum_{i=1}^{c}\sum_{j=1}^{n} A_{ij}^{i}$, therefore class C samples within the general class discrete degree matrix S_w as follows:

$$S_w = \sum_{i=1}^{c}\sum_{j=1}^{n}(A_j^i - \overline{A^i})^T(A_j^i - \overline{A^i})$$

(1)

The total C class samples between the discrete degree matrix for Sb:

$$S_b = \sum_{i=1}^{c}(\overline{A^i} - \overline{A})^T(\overline{A^i} - \overline{A})$$

(2)

2DLDA criterion function J (w) is defined as:

$$J(w) = \frac{W^T S_b W}{W^T S_w W}$$

(3)

In order to make the linear differential characteristics after projection have optimal classification ability, the optimal projection vector W_{opt} should make criterion function J(w) the maximum. We can calculate the eigenvector and eigenvalue of the generalized linear equation to get optimal identification vector W_{opt} of 2DLDA algorithm, namely:

$$S_b S_w^{-1} W_{opt} = \lambda Wopt \tag{4}$$

Among them, λ is the maximum eigenvalue of $S_b S^{-1}{}_W$, W_{opt} is the eigenvector corresponding to λ_i. It's not enough for us to usually select one discriminant vector as the projection direction.

Therefore we select the d larger eigenvalues corresponding feature vector W_1, W_2, W_3,..., W_d as projection direction to get the optimal projection matrix $W_{opt}=[w_1,w_2,w_3,…,w_d]$.We can get a m×d dimensional whole projection feature of the original image matrix if we project the training sample image on the W for feature extraction and we also project the test samples on the W. It can be effective for dimension reduction.

The SVM Parameter Selection Based on Particle Swarm Optimization Algorithm

Face recognition belongs to nonlinear problem so it increases the difficulty to identify. Support vector machine is a new machine learning algorithm and it can find global optimization solution under the condition of less training samples and nonlinear and it has a good application prospect in the field of face recognition. The performance of SVM greatly depends on the selection of parameters (mainly the penalty parameter c and kernel function parameter g) [12].

To overcome the uncertainty of kernel function parameters in support vector machine (SVM) and the optimal solution of the nuclear parameters selection problem, this paper proposed to apply particle swarm optimization algorithm originated from artificial life and evolutionary computation theory to the parameter selection in support vector machines (SVM). Using particle swarm optimization (PSO) can find the optimal parameters within the solution space and can avoid the effects of artificial parameters setting, so as to improve the recognition rate. Experimental results show that selected nuclear parameters by this method can effectively improve the classification and prediction accuracy.

The SVM Algorithm. SVM is a supervised learning algorithm based on statistical learning theory. The sample space through a nonlinear mapping is mapped to a high-dimensional or infinite dimension in the feature space. It makes the nonlinear separable problem in the original sample space into a linear separable problem in feature space.

Suppose two inseparable training sample set $\{(x_i , y_i), i=1, 2, 3, … , n\}$.Among them, $x_i \in R^m$ is the ith training sample, $y_i \in \{-1,1\}$ is the category label of the ith sample. The sample can be mapped to high-dimensional linear space H through nonlinear mapping φ： $x_i \rightarrow \varphi(x_i)$, and the training samples can be expressed as $\{(\varphi(x_i),y_i); i=1,2,3,…,n\}$.The solution of the optimal separating hyperplane can be expressed as a constraint optimization problem [13]:

$$\min \tfrac{1}{2} \| w \|^2 \tag{5}$$

The constraint conditions is $\forall i, y_i(w \circ \phi(x_i)+b)-1 \geq 0$ 。 This is a convex optimization problem. Among them, w is the normal vector of the optimal separating hyperplane, b is a constant, $\|\circ\|$ often represent the Euclidean distance. The introduction of Lagrange multilayer a=a_i(i=1,2,3, …) make the equation $a_i[y_i(w \circ \phi(x_i)+b)-1]=0$,and we get Lagrange Eq.6:

$$L(w,b,a) = \tfrac{1}{2}\| w \|^2 - \sum_{i=1}^{n} a_i y_i \left(w \circ \phi(x_i)+b\right) + \sum_{i=1}^{n} a_i \tag{6}$$

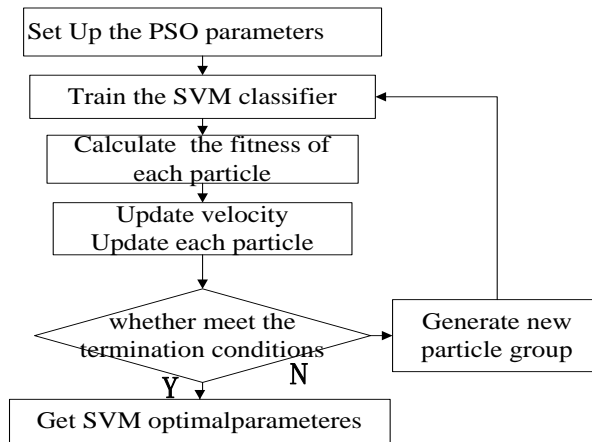We can get W_{opt} dual expression derived from before is:

Figure 2.   The PSO algorithm flow chart to optimize parameters of SVM

The SVM used the RBF as kernel function and there are only two crucial parameters that is penalty parameter c and kernel function parameter g. Particle swarm optimization algorithm introduced the two parameters. Fig. 4 shows the process using PSO to optimize the SVM parameters. The steps can be described as:

   1)   Set up the PSO parameters. Initialize the search space and search point particle's position and speed, the maximum number of iterations.

   2)   Use the SVM recognition rate to calculate the fitness of each particle.

   3)   Update the particle position and velocity according to the equation (10) and (11), and update individual extremum and groups extremum according to the new particle fitness value.

   4)   Determine whether meet the termination conditions (or minimum error requirement), so we can stop the iteration, output optimal solutions, otherwise go to step 2).

We used the particle swarm algorithm in support vector machine to find the optimum parameters in the solution space so that the result can be relieved by the influence of parameters set by artificial experience.

The Experiment Result and Analysis

The subjects of this paper are from ORL database. ORL face database is based on changes of the expressions and gestures .The database has 40 people and each person has ten face images of different attitude and expression. So the total is 400 .The images are taken from different angles and the resolution is 112 x 92, grayscale is 112. Each class has first five for the training sample and last five samples for testing. ORL face image database samples are shown as follows:
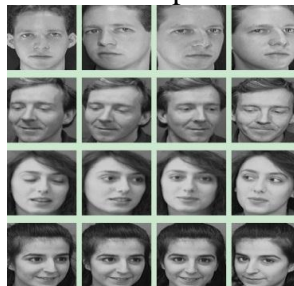

Figure 3.   ORL face image database samples

The experiment process is shown in Fig. 4. First of all, we extract features of face image database samples by using 4 layer wavelet transform to extract low frequency feature and   2DLDA without vector quantization method. It can effectively solve the computational complexity and matrix singular problem. This experiment used 2DLDA to extract identification feature. The low frequency

component between class and discrete degree matrix of class of Face image are calculated separately in the 2DLDA method. Then we calculate the optimal projection matrix W_{opt} of the face image so that the low frequency component of face image can project on W_{opt} so that we can get 2DLDA feature of the face images. And 2DLDA feature of face image is converted into one column vector, finally we used particle swarm optimization for better classification performance of support vector machine.
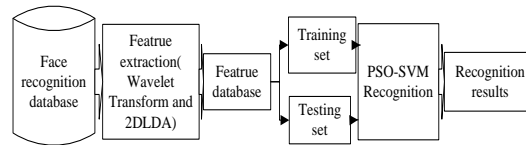

Figure 4.   The experiment process

The support vector machine adopts RBF kernel function and "one to many" SVM classification strategy. We preset c = 128, g = 0.02 to complete facial recognition. Then we set the parameters c, g according to the experience and compare it with parameters selected by PSO. The c, g parameters and recognition rate are shown in Fig. 5:
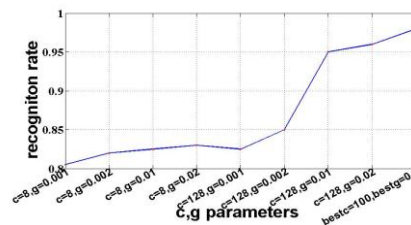

Figure 5.   c,g parameters and recognition rate

Particle swarm algorithm initialize the population size of group 20. The maximum value of parameter c is100, the minimum value of parameter c is 0.1 and the maximum value of g parameter is100, the minimum value of g parameter is 0.01. We take the recognition rate as fitness function and can get the best parameters after much iteration. The optimization process of c, g parameters in the particle swarm optimization algorithm is as shown in Fig. 6.
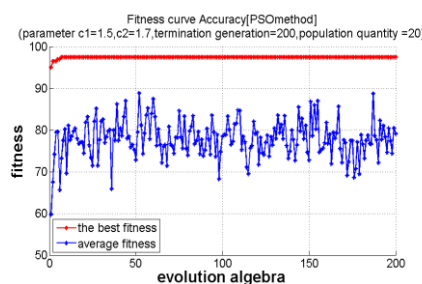

Figure 6.   The optimization process of c, g parameters

When LDA is used to deal with facial image. It brings the computational complexity and large matrix singular problem. This experiment used the LDA and 2D vector of face image matrix without 2DPCA and 2DLDA method to compare. This experiment runs 2DPCA + the nearest neighbor, 2DLDA + the nearest neighbor, 2DLDA+ SVM, LDA + SVM and 2DLDA + PSO + SVM these five kinds of algorithm. The recognition rate of the method used in this paper and other algorithm is shown in Fig. 7.
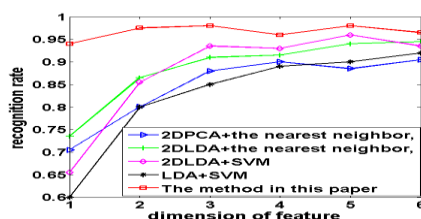
Figure 7.   The method in this paper compared with other algorithms

Experiment shows that: we can see from the figure 7 that the recognition rate of 2DLDA is better than that of 2DPCA under the condition of same number of principal components due to the effective usage of the spatial structure information of face image. And 2DLDA algorithm is superior to the one dimensional LDA algorithm. Combined with figure 5 we can see that the parameter selection has an important impact on the performance of support vector machine (SVM). The recognition rate is only 82.5% when we set $c = 8$, $g = 0.01$ according to the experience. The support vector machine (SVM) using PSO algorithm for parameters optimization has a higher recognition rate and it can reach 98%. Figure 6 shows the process of parameter optimization. We can find parameters corresponding to the best fitness through multiple iterations. It will take more time than other methods because it used the particle swarm algorithm, so you will need to weigh the consideration for the occasions of high real-time requirements.

Conclusion

In this paper, particle swarm optimization algorithm is applied to the support vector machine parameters selection process. It greatly reduce the effects of artificial   parameters setting and it combined with the wavelet transform and two-dimensional differential analysis (2DLDA) for feature extraction of facial feature so that it effectively remove high frequency signal interference and solve the small sample problem of LDA. The experimental results show that the algorithm is superior to other algorithms in the experiments and the recognition rate reached 98%.So it illustrates the effectiveness of this algorithm applied to face recognition.

Acknowledgements

References

[1] Lu Z, Zhang L. Face recognition algorithm based on discriminative dictionary learning and sparse representation [J]. Neurocomputing, 2016, 174:749–755.

[2] Luan X, Fang B, Liu L, et al. Extracting sparse error of robust PCA for face recognition in the presence of varying illumination and occlusion[J]. Pattern Recognition, 2014, 47(2): 495-508.

[3] Vinay A, Shekhar V S, Murthy K N B, et al. Face Recognition Using Gabor Wavelet Features with PCA and KPCA-A Comparative Study [J]. Procedia Computer Science, 2015, 57: 650-659.

[4] Mahanta M S, Plataniotis K N. 2DLDA as matrix-variate formulation of a separable 1DLDA[J]. Pattern Recognition Letters, 2015, 68: 169-175

[5] Hong Zhang, Xingyu Gao, Ping Wu, Xin Xu, A cross-media distance metric learning framework based on multi-view correlation mining and matching, World Wide Web,2016, Volume 19, Issue 2, pp 181-197

[6] Hou C, Nie F, Zhang C, et al. Multiple rank multi-linear SVM for matrix data classification[J]. Pattern Recognition, 2014, 47(1): 454-469.

[7] KENNEDY J, RUSSELL E. Particle swarm optimization [C]// Proceedings of IEEE international conference on neural networks. 1995, 4(2): 1942-1948.

[8] Liu D, Niu D, Wang H, et al. Short-term wind speed forecasting using wavelet transform and support vector machines optimized by genetic algorithm [J]. Renewable Energy, 2014, 62: 592-597.

[9] Baccar D, Söffker D. Wear detection by means of wavelet-based acoustic emission analysis [J]. Mechanical Systems and Signal Processing, 2015, 60: 198-207.

[10] Moulin C, Largeron C, Ducottet C, et al. Fisher linear discriminant analysis for text-image combination in multimedia information retrieval[J]. Pattern Recognition, 2014, 47(1): 260-269.

[11] Zhang hong, Wu fei, Zhang xiaolong. Multimedia data clustering based on fusion of relationship matrix [J]. Journal of computers, 2011, 34 (9): 1705-1711.

[12] Sudheer C, Maheswaran R, Panigrahi B K, et al. A hybrid SVM-PSO model for forecasting monthly streamflow [J]. Neural Computing and Applications, 2014, 24(6): 1381-1389.

[13] Chen Y, Zhou X S, Huang T S. One-class SVM for learning in image retrieval[C]//Image Processing, 2001. Proceedings. 2001 International Conference on. IEEE, 2001, 1: 34-37.

[14] Subasi A. Classification of EMG signals using PSO optimized SVM for diagnosis of neuromuscular disorders [J]. Computers in biology and medicine, 2013, 43(5): 576-586.

[15] Tang M, Chen F. Facial expression recognition and its application based on curvelet transform and PSO-SVM [J]. Optik-International Journal for Light and Electron Optics, 2013, 124(124): 5401-5406.