Face Recognition Method Based on Within-class Clustering SVM

Yan Wu, Xiao Yao and Ying Xia
Department of Computer Science and Engineering
Tongji University
Shanghai, China

Abstract **- A face recognition method based on Within-class Clustering SVM (CCSVM) is presented in this paper in order to decrease the negative effect caused by noisy training samples in the recognition process. Based on the discontinuity of finite samples distribution in the high dimension space and the existence of noisy samples, first, we re-cluster samples within the class, find out the cluster centres to form the virtual classes, and then divide virtual classes of all the classes by SVM. Experiment results show that this method follows the distribution law of points in high-dimensional space and can achieve better performance than some traditional methods.**

Keywords **- face recognition, within-class cluster, support vector machine (SVM)**

I. INTRODUCTION

Face recognition is an important research topic in Biometrics and has already become very hot in the field of pattern recognition. Many scholars at home and abroad are devoted into researches in this field. Major statistical pattern recognition methods nowadays include discriminant function method, K-nearest neighbor classification method, non-linear mapping method and feature analysis method, etc. Support Vector Machine (SVM) [1] is the statistical learning theory put forward by Vapnik based on structural risk minimization principle and it has been applied in classification and regression problems. SVM method has been paid more and more attention to because of its outstanding characteristics and promising experimental performance. However, the face samples of the same class present great randomness because of a variety of noises in reality, such as sunlight, postures, etc. Many noisy samples (dirty samples) in high-dimensional space differ greatly from other samples of the same class, so these samples can not be clustered together and show great clustering distinctiveness. Recent research results show that samples of the same kind in high-dimensional space are located in a non-linear manifold [2-3], however, the limited number of sample points determines its discontinuity, this is, there are several cluster areas within the class in space. Therefore, the key is to determine the cluster center. Up to now, many clustering methods have been proposed, such as Separation-Based K-Means [4] algorithm, Density-Based Spatial Clustering of Application with Noise (DBSCAN) [5] algorithm, Bottom-up Clustering using Representative (CURE) [6] algorithm and Gridding-and-Density-based Clustering In QUEst(CLIQUE)[7] algorithm. Each method has its own merit and also has some limitations.

In this paper a face recognition method based on within-class cluster and SVM is proposed. This method is along with the thinking of CURE algorithm and the merits of SVM algorithm. Conducting cluster analysis within each class, several cluster areas can be found as virtual classes according to its spatial distribution law and all virtual classes can be classified with SVM classifier. Experiments show that this algorithm achieves satisfactory recognition effect.

II. CLUSTER ANALYSIS BASED ON THE COVERAGE AND MERGING

Definition 1 Virtual center point: In high-dimensional space, for a point p and a threshold R, within a cluster area determined by p (center) and R (radius) the average of all the points is $x_0 = \frac{1}{card(\rho < R)} \sum_{l}^{card(\rho<R)} x_l$, where x_l is the lth sample point, ρ is the distance between a sample point and point p , $card(\rho < R)$ is the point set satisfying condition $\rho < R$.

Definition 2 Representative areas: Each virtual center point represents the hyper ball area with this point as the center and R as the radius, so this area is called representative area.

Definition 3 Adjacent center: Given radius R, if the distance of two virtual points p and q satisfies $dist(p,q) <= \theta$, then p and q are the adjacent centers, where θ is the merger threshold.

The key to cluster analysis is to find the cluster center and it can be divided to three steps: Step 1, determine coverage radius and merger threshold; Step 2, search for the candidate cluster centers and eliminate the ones that don't meet the density requirements; Step 3, merge the adjacent candidate centers and decide the final cluster center.

The threshold radius for the coverage represents the distribution law for the sample points in high-dimensional space and can be determined by the analysis for the samples of the same kind. The training sample set S with n training samples can be divided into virtual training set S_1 and virtual testing set S_2 and for each $x_i \in S_2, i = 1,2...n$, calculates

$$\rho_i = \min\left\|x_j - x_i\right\|, x_j \in S_1, j = 1,2...m \quad (1)$$

Radius R for coverage is determined as the average distance between the virtual training set and all the points from the virtual testing set, so it is shown as:

$$R = \frac{1}{card(S_2)} \sum_{i}^{card(S_2)} \rho_i \qquad (2)$$

Merger threshold θ is determined by average distance between couples of virtual center points and the formula is:

$$\theta = \{\eta \mid \max(\eta_1, \eta_2 ... \eta_t), \\ \eta_1 < t, \eta_2 < t ... \eta_t < t\} \qquad (3)$$

Where $\eta_1, \eta_2 ... \eta_t$ are the distances between virtual center points and t is the mean vector of all the current virtual center points.

After R is determined, cluster centers can be searched out by the following steps:

(1) Initialization: $n = 1$, initialize candidate center set as $candcenter = \phi$, set the number of iteration as $iterate$;

(2) Get sample points $x_i, x_i \in S$ from the sample queue of current class;

(3) Calculate the Euclidean distances ρ_j from x_i to each present virtual center points $center_j$. If $\rho_j \le R$, then x_i belongs to $center_j$'s representative area, else switch to (4).

(4) x_i becomes the new virtual center point and add x_i into $candcenter$ set.

(5) If the sample queue for the current class is not empty, then repeat (2) to (4), else switch to (6);

(6) Obtain the number num_j of the sample points in the representative area of every candidate center. If $num_j \le 1$, then this virtual center should be eliminated from $candcenter$ to be an unclassified sample point. $n = n + 1$ and if $n \le iterate$, then update virtual center vector and switch to (2), else switch to (7);

(7) Every unclassified point becomes a new virtual center. If two virtual center points are adjacent, then the corresponding representative areas should be merged, that is to say that all points in two representative areas should be merged, and then the updated virtual center points will become final cluster centers.

III. RECOGNITION METHOD CASED ON WITHIN-CLASS CLUSTERING SVM (CCSVM)

The basic principle of SVM is to map the input space to high-dimensional space through non-linear transformation and get the optimal linear classification plane according to minimal structure criterion in the new space and this non-linear transformation is completed by defining proper inner-product function.

Suppose the Linear separable sample set $(x_i, y_i), i = 1...n$ ($x \in R^n, y \in \{+1,-1\}$, and y are class labels), the general form for linear discriminant function is $g(x) = w \bullet x + b$ and classification plane equation is $w \bullet x + b = 0$. Normalize the discriminant function and for all the samples in two classes make $\|g(x)\| \ge 1$, so the nearest sample to the classification plane can $\|g(x)\| = 1$ to make the classification margin be $2/\|w\|$, therefore, the maximum margin means the minimum $\|w\|$; Requiring that all samples should be correctly classified by classifier, this is:

$$y_i[(w \bullet x) + b] - 1 \ge 0, i = 1,2...n \qquad (4)$$

So the classification plane, which meets the above mentioned conditions and make $\|w\|$ minimum, is the optimal classification plane.

The optimal classification plane problem can be considered as solving the constrained optimization problem that is restricted by (4), the minimum value for the function is:

$$\phi(w) = \frac{1}{2}\|w\| = \frac{1}{2}(w \bullet w) \qquad (5)$$

This can be resolved by Lagrange Multiplier Method.

For non-linear problem, all the points should be considered to be mapped into the high-dimensional space in order to be linear separable in this space and only conduct inner-product operation for linear discrimination in high-dimensional space. There is no need to know the non-linear transformation form, so avoiding complicated calculation in high-dimensional transformation, the problem is greatly simplified. According to Hilbert-Schmidt principle, it can be used as inner-product function only if one operation can satisfy Mercer condition.

SVM is two-class classifier and for multiple-class pattern recognition problem, a combination scheme for multiple SVM classifiers can be applied, such as 1-a-r multiple-class classification [8]. Each class in sample set can be clustered based on the coverage and mergence method proposed in section II. Corresponding to the multiple cluster centers, the samples of the same kind can be divided to several virtual classes and all of them can be classified by SVM. Because of the increase of the class numbers, after re-clustering within class, if 1-a-r method [9] proposed by Knerr is applied, then $K(K-1)/2$ classifiers should be constructed and the decision-making efficiency will decrease.

In order to solve this problem, for the training samples in K virtual classes, let the sample in the first class positive and the samples in 2nd, 3rd, … , Kth classes negative to train the first SVM. The ith SVM(SVM i) is trained by the samples which take the ith class as positive and the $i+1$ th, $i+2$ th, … K th classes as negative. Finally taking the sample of $K-1$ th class as positive and the K th class negative the $K-1$ th SVM can be trained to be SVM$(K-1)$.

Therefore, for a K-classification problem, efficiency for classification can be greatly improved because only $K-1$ two-class classifiers have to be trained.

IV. EXPERIMENTS AND ANALYSIS

ORL, Yale and YaleB face database have been applied in the experiments so as to prove the validation of the algorithm. In order to assure the same dimension for different training samples of each face pattern, face images for each person are pre-processed and the feature dimension can be reduced through Principle Component Analysis (PCA).

A. Experiment results in ORL database

ORL database includes images of 40 different persons and 10 112x92 gray images for each person. This database is mainly influenced by posture and expression. For ORL database, 40 face patterns are applied and k images for each person are randomly chosen as training samples and the rest $10-k$ for each are training samples.

The traditional K-nearest neighbor (NN) method, SVM method and CCSVM method put forward in this paper are applied in the experiments, in which the polynomial kernel $K(x,x_i)=[(x\bullet x_i)+1]^2$ is chosen for kernel function, and multiple-class classification mentioned in Section III is chosen for SVM and CCSVM methods. We conduct several experiments with different k values and calculate the mean of the testing results under the same k value. The results are shown in Fig. 1. We can see that with insufficient samples multi-cluster areas are not easy to generate in the same-class-sample because different face samples of the same kind only have tiny expression and pose changes .Therefore all the methods used in experiment differ a little from each other.

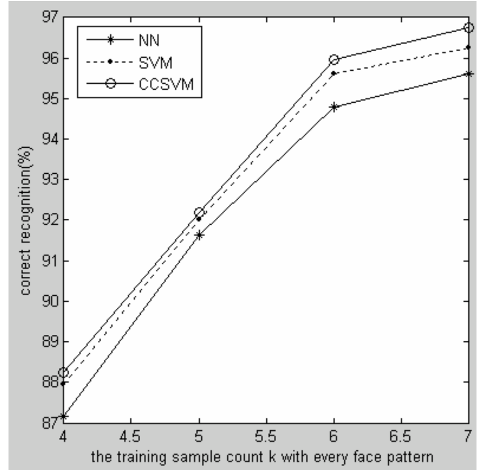


Fig. 1 Correct recognition rate in ORL database

Experiment in ORL database for SVM and CCSVM methods with different kernel functions when $k=20$ and $k=7$ are compared in Table I, in which polynomial kernel with $d=2$, radius kernel function and Sigmoid kernel function: $K(x,x_i)=\exp\left\{-\frac{|x-x_i|^2}{256\sigma^2}\right\}$ and $K(x,x_i)=\tan\left(\frac{b(x\bullet x_i)}{256}-c\right)$ are adopted in the experiment, and σ^2, b and c are set to 0.3, 2 and 1 respectively. We can see that CCSVM displays better recognition effect on the basis of these kernel functions when compared with SVM in the ORL database through relatively small abnormal samples. Time complexity in the recognition process is also compared and we find that it is not significantly increased with the increase of total class numbers and this proves the feasibility of this algorithm.

TABLE I. COMPARISON FOR RECOGNITION RATE WITH DIFFERENT KERNEL FUNCTION (ORL DATABASE)

	SVM		CCSVM	
	Correct recognition rate (%)	Running time (s)	Correct recognition rate (%)	Running time (s)
Polynomial	91.33	0.1360	92.74	0.1390
Radius basis	92.15	0.1240	92.89	0.1310
Sigmoid	91.84	0.1170	93.50	0.1200

B. Experiment results in YALE database

In order to test the effectiveness of the method, Yale database is used in the experiment. YALE face database includes images of 17 different persons and 33 images for each person. We construct the experiment with 17 face patterns in YALE database and use k images for training and the remaining 33-k images for testing. Compared with other databases, face images in YALE database are influenced much more by different illumination and expressions. Same experiment as in ORL database is conducted and the result is shown in Fig. 2.

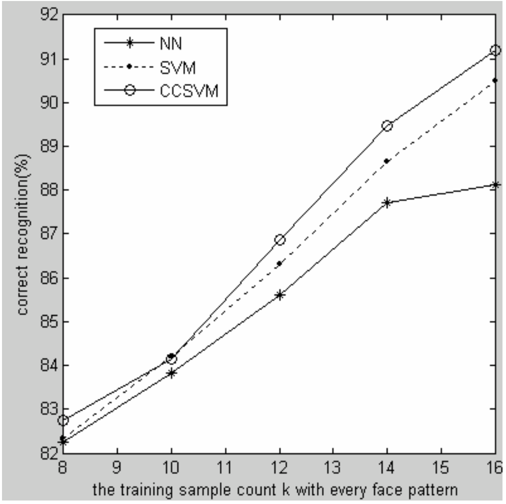
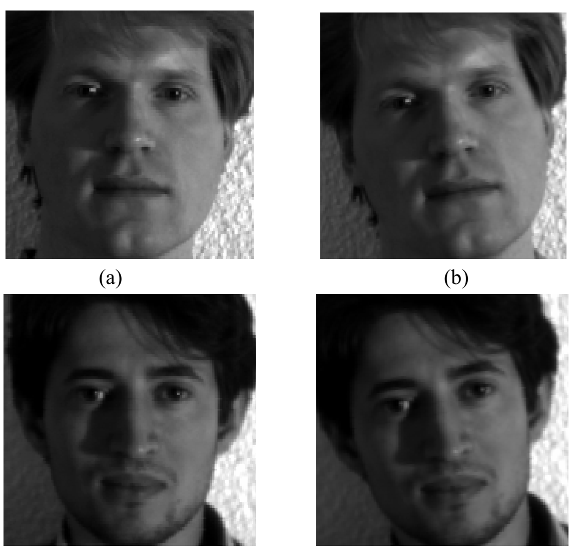


Fig. 2 Correct recognition rate in YALE database

We can see from Fig. 2 that some sample points influenced much by noise (such as the sample points with more illumination angle) will participate in the training process in YALE database with the increase of *k*. These samples may be much far way from the samples of the same kind and samples under the same illumination condition in space may be often clustered together (such as the face samples in Fig. 3) so as to generate new cluster area. If these "abnormal" samples are trained together with other "normal" samples, misclassification will be generated with samples under the same illumination condition. However, CCSVM method can separate these "abnormal" samples from samples of the same kind by clustering to generate two virtual classes and train them respectively. This method better follows the distribution law of the samples in high-dimension space and greatly improves the recognition effect.



(a) (b) (c) (d)

Fig. 3 Sample images under the same illumination condition

SVM and CCSVM with different kernel function are compared in Table II when $k = 20$ and $k = 7$ in YALE database. Parameter selections are the same as those in ORL database. We can see that CCSVM displays more advantages in recognition rate compared with SVM because of the increase of "abnormal" samples (such as the illumination samples in Fig. 3). If images (c) and (d) in Fig. 3 are testing samples, SVM will misclassify them into the class represented by sample (a) and (d), but with CCSVM method they will be correctly classified.

TABLE II. COMPARISON FOR RECOGNITION RATE WITH DIFFERENT KERNEL FUNCTION (YALE DATABASE)

	SVM		CCSVM	
	Correct recognition rate (%)	Running time (s)	Correct recognition rate (%)	Running time (s)
Polynomial	94.26	0.1480	94.71	0.1500
Radius basis	94.87	0.1350	95.28	0.1370
Sigmoid	95.58	0.1110	96.03	0.1240

C. Experiment results in YaleB database

YaleB database is composed of illumination images with illuminations from different directions, such as vertical, horizontal and mixed polarized lights. In order to compare the robustness of NN, SVM and CCSVM influenced by illumination, we use 25 face images of each person affected by illumination from all directions to be a sample set and then divide it into 4 subsets randomly. One of them is used as training sample set and the others are used as testing sample sets, and then we get the average recognition rates which are shown in Table III. We can see that the samples of the same person can not be clustered together in high-dimensional space because images in this database are greatly influenced by illumination, however, multi-cluster area are generated according to different polarized angle (for example, samples of the same kind with the polarized angle ranging from 20 to 30 degrees may be clustered together, and they keep a relatively far distance from samples with the polarized angle less than 10 degrees). CCSVM displays much more advantages compared with SVM and recognition rate have been increased significantly (for example, recognition can be increased to 7.8% when Subset 3 is used as training set). CCSVM based on virtual class in different cluster area will obtain relatively satisfactory recognition effect even without the denoising feature extraction process.

TABLE III. AVERAGE CORRECT RECOGNITION RATE WITH DIFFERENT SUBSETS AS TRAINING SAMPLE SET (YALEB DATABASE)

	Subset 1	Subset 2	Subset 3	Subset 4
NN	66.7	62.4	53.2	76.6
SVM	75.1	74.3	67.8	82.9
CCSVM	80.5	78.4	75.6	88.2

V. CONCLUSIONS

A face Recognition method based on Within-class Cluster and SVM is proposed in this paper. Based on the distribution discontinuity of the samples in space and the existence of noisy samples, this method forms many cluster areas and searches out the corresponding cluster centers by cluster analysis, dividing each class into several virtual classes, all virtue class can be classified by SVM. Many experiments show that CCSVM follows the distribution law of feature points in high-dimensional space and achieves better recognition effect.

REFERENCES

[1] [1] O. Edgar, F. Robert, G. Federico. “Training support vector machines: an application to face detection”. IEEE Conference on Computer Vision and Pattern Recognition, San Juan, Puerto Rico.1997.

[2] [2] J. B. Tenenbaum, V. DE Silva, J. C. Langford. “A global geometric framework for nonlinear dimensionality reduction”. Science,2000,290(5500):23 19-2323.

[3] [3] S. Roweis, lL. K. Saul. “An introduction to locally linear embedding”.

[4] [4] L. Kanfan, PJ. Rousseeuw. “Finding Groups in Data: an Introduction to Cluster Analysis”. New York: John Wiley& Sons. 1990.

[5] [5] M. Ester, H. Kriegel, J. Sander, X. Xu. “A density based algorithm for discovering clusters in large spatial databases with noise”. In: Simoudis E, Han JW, Fayyad UM . eds. Proceedings of the 2nd International Conference on Knowledge Discovery and Data Mining. Portland: AAAI Press, 1996, 226-231.

[6] [6] G. Sudipto, R. Rajeevand R. Kyuseok. “Cure: an efficient clustering algorithm for large databases”. Information Systems, 2001, 6(3):128-137

[7] [7] R. Agrawal, J. Gehrke, D. Gunopolos, P. Raghavan. “Automatic subspace clustering of high dimensional data for data mining application”. In : Haas LM . Tiwary A, eds. Proceedings of the ACM SIGM OD International Conference on Management of Data. Seattle: ACM Press. 1998. 94-105.

[8] [8] C. W. Hsu, C. J. Lin. “A comparison of methods for multiclass support vector machines” . IEEE Transactions on Neural Networks,2002,13(2) :415-425.

[9] [9] U. KreBel. “Pairwise classification and support vector machines”.Advances in Kernel Methods-Support Vector Learning. Cambridge, MA;MIT Press,1999.255-268.