# Face Recognition using Multi-class SVM

Ping Du, Yankun Zhang, Chongqing Liu
Inst. of Image Processing and Pattern Recognition,
Shanghai Jiao Tong University, Shanghai 200030, China
pdu901@mail1.sjtu.edu.cn

## Abstract

*In this paper, an approach to face recognition is proposed, in which SVM combined with nearest center classification (NCC) is used as the classifier. The philosophy behind this is based on that idea that their discriminative capabilities are not totally overlapped so that NCC may work on the samples that SVMs fail. Firstly, the principal component analysis is used to reduce dimension and extract feature. Then support vector machine (one-to-other scheme) combined with nearest center classifier used for classification. We conduct the experiment on the base of ORL face database with our method and three other decision rules for their comparison. The experiment result is presented and discussed, which shows the effectiveness of the strategy described.*

## 1. Introduction

Face images have received considerable attention from both the computer vision and signal processing fields. This interest is mainly motivated by the broad range of potential applications for systems able to recognize the face they contain. Examples include surveillance, personal identification, access control, mugshot recognition, and human computer interaction [1].

The recognition of faces is a well established field of research and a large number of algorithms have been proposed in the literature. Popular approaches include the ones based on Eigenfaces [2], dynamic link matching [3], active appearance models [4]. And other methods, such as neural networks [5], and HMM [6], are also described in the face recognition literature.

Support vector machine is a novel technique for pattern recognition. Because its high performance in tackling small sample size, nonlinear, high dimension and its good generalization, people attach more and more importance to it recently. The earlier studies of SVMs in face recognition have been reported in [7-9]. In this paper, a face recognition method is proposed that utilizes one-to-other SVMs combined with nearest center criterion to classify multi-class faces in the feature space.

This paper is organized as follows. Section 2 introduces the basic concept of PCA. Section 3 presents the canonical theory of SVM and our strategy of decision rule, which combines the SVM one-to-other scheme with NCC. Experiment result and discussion are then described in section 4. Finally, conclusions are given at the end of the paper.

## 2. Principal Component Analysis

PCA [2] is a standard technique for dimension reduction and feature selection. The $N \times N$ face image can be expressed as a point in the $N \times N$ dimensional space. Then the ensemble of face images can be expressed as congregation of point in image space. The purpose of PCA is to find the appropriate vectors that can describe the distribution of face images in image space and form another space with lower dimension. Let the training set of face images be $\Gamma_1, \Gamma_2, \Gamma_3, ...\Gamma_M$. The average face of the set is defined by $\Psi = \frac{1}{M}\sum_{n=1}^{M}\Gamma_n$. Each face differs from the average by the vectors $\Phi_i = \Gamma_i - \Psi$. The eigenvalues and eigenvectors are obtained from covariance matrix,

$$C = \frac{1}{M}\left[\sum_{n=1}^{M}\Phi_n\Phi_n^T\right] \tag{1}$$

and a new face image is transformed into its face components by the operation described in [2].

The weights form a vector $\Omega = [\omega_1, \omega_2, ...\omega_M]$ that describes the contribution of each eigenvector in representing the input face image, treating the eigenvector as a basis set for face images.

$$\omega_k = u_k^T(\Gamma - \Psi), k = 1,2,..., M \tag{2}$$

Where $u_k$ is eigenvector. These weights may be used in a face classification algorithm to find which of predefined face classes that describe the face.

## 3. The SVMs for face recognition

PCA for feature extraction and dimension reduction has been described in the previous section. We will use SVMs to be the classifier to classify the samples in the feature space. Our strategy of face recognition, which is based on the combination of SVMs and NCC, is introduced in this section following the canonical SVM theory.

### 3.1 Basic theory of support vector machine

The basic idea of SVM is to map the linear non-separable input vectors into some higher dimensional space such that a more suitable hyperplane can be found with minimal classification errors [9-11].

We start with training data,

$$D = \{(x_i, y_i)\}_{i=1}^{l}, \, where \; y_i \in \{-1,1\}, x_i \in R^N \quad (3)$$

then map the training data into some other inner product space $F$ via a nonlinear map,

$$\Phi : R^N \to F \quad (4)$$

The separating hyperplane in the space $F$ must satisfy the following constraints,

$$y_i(w^T z_i + b) \geq 1, \quad z_i \in F, \; i = 1,2,\cdots,l \quad (5)$$

If the optimal hyperplane $H$ is $w_0^T z + b_0 = 0$,

then the distance between the closest vector to the hyperplane $H$ is,

$$\rho(w,b) = \min_{\{x|y=1\}} \frac{z^T w}{\|w\|} - \max_{\{x|y=-1\}} \frac{z^T w}{\|w\|} \quad (6)$$

with its maximum,

$$\rho(w_0, b_0) = \frac{2}{\|w_0\|} = \frac{2}{\sqrt{w_0^T w_0}} \quad (7)$$

So the optimal separating hyperplane is determined by the vector $w$, which minimizes the functional.

$$\Phi(w) = \frac{1}{2}(w^T w) \quad (8)$$

$$subject \;\; to \;\; y_i(w^T z_i + b) \geq 1, \; i = 1,2,\cdots,l$$

It's modified for the non-separable case to,

$$\Phi(w) = \frac{1}{2} w^T w + \gamma \sum_{i=1}^{l} \xi_i \quad (9)$$

where the $\xi_i$ are measure of the misclassification error.

In terms of Lagrange multipliers, $w_0$ can be written as

$$w_0 = \sum_{i=1}^{l} \lambda_i y_i z_i$$

so the decision function,

$$f = \mathrm{sgn}[\sum_{i=1}^{l} \lambda_i y_i (z^T z_i) + b] \quad (10)$$

The theorem of functional analysis shows that a positive-semi definite symmetrical function $K(u,v)$ can solely define a Hilbert space $H_k$, $K$ is the reproducing kernel of feature space $H_k$,

$$K(u,v) = \sum_k \alpha_k \varphi_k(u) \varphi_k(v) \quad (11)$$

which represents a inner product in the feature space,

$$z_i^T z = \phi(x_i)^T \phi(x) = K(x_i, x) \quad (12)$$

The decision function can thus be written as

$$f = \mathrm{sgn}[\sum_{i=1}^{l} \lambda_i y_i K(x_i, x) + b]. \quad (13)$$

### 3.2 SVM in multi-class classification

The formulation of SVM in previous section was based on a two-class problem, hence SVM is basically a binary classifier. Several different schemes can be applied to the basic SVM algorithm to handle the *K*-class pattern classification problem.

The schemes which have been proposed in [12] for solving the multi-class problem are as listed below:

Using *k* one-to-rest classifiers is the simplest scheme, and it does give reasonable results. *K* classifiers will be constructed, one for each class. The $\mathrm{K}^{\mathrm{th}}$ classifier will be trained to classify the training data of class *k* against all other training data. The decision function for each of the classifier will be combined to give the final classification decision on the *K*-class classification problem,

$$f(x) = \arg\max_k \sum_{i=1}^{l} \lambda_i^k y_i K^k(x_i, x) + b^k \quad (14)$$

Using *k*(*k*-1)/2 pairwise classifiers with majority voting or pairwise coupling as the voting scheme.

The schemes require a binary classifier for each possible pair of classes. The decision function of the SVM classifier for $\mathrm{K}^{\mathrm{i}}$-to-$\mathrm{K}^{\mathrm{j}}$ and $\mathrm{K}^{\mathrm{j}}$-to-$\mathrm{K}^{\mathrm{i}}$ has reflectional symmetry in the zero plane. Hence only one of these pairs of classifier is needed. The total number of classifiers for a *K*-class problem will then be *K*(*K*-1)/2. The training data for each classifier is a subset of the available training data, and it will only contain the data for the two involved classes. The data will be reliable accordingly, i.e. one will be labelled as +1 while the other as -1. These classifiers will then be combined with some voting scheme to give the final classification results, such as majority voting or pairwise coupling.

### 3.3 The combination of SVMs with NCC

The construction of $k$ one-to-rest classifiers is conducted on the first stage of classification in the experiment. In the training process, the $K^{th}$ class is labeled as +1 and all the other samples are labeled as –1 to train the $K^{th}$ classifier. However, instead of (14), we use (13) as the $K^{th}$ classifier and the nearest center classifier is integrated into the make of the final classification decision.

For a sample to be tested $\alpha$, if when $f_i(\alpha) \neq +1, \forall i = 0,1,...,k$, $\alpha$ can be regarded as a new class; else if when $i = i_0$, $f_i(\alpha) = +1$, we can say that $\alpha \in K^{i_0}$; else when $i = i_0, i_1,...i_n, n \geq 1$, $f_i(\alpha) = +1$, we will use nearest center criterion to make further classification,

$$\alpha \in K^i, where\ \min_i \|x_i - \alpha\|,\ i = i_0, i_1,..i_n,\ x_i \text{ is the center of } K^i$$

## 4. Result and Discussion

This experiment has been conducted on the database of Olivetti Research Laboratory (ORL). It's made up of 400 images of 40 individuals, 10 of each person with various luminance, expression and pose. It's grayscale image and the resolution is $112 \times 92$.

In our experiment, 200 images (5 for each individual) are selected as training set, from which we calculate the eigenfaces and learn discrimination functions by SVMs , and the left 200 ones are for the test of recognition.

**Figure 1 Some of the images in the ORL database**

We use the first 30 eigenfaces to represent the feature space and Radial Basis Function (RBF) as kernel function to train the SVM classfiers. Table 1 shows the performance of different classifiers, including nearest center, BP neural network and SVM combined with nearest center.

**Table 1 Error rates with different classifiers**

| The Algorithm of Classifier | Error Rate |
|---|---|
| BP Neural Network | 10.0% |
| Nearest Center Classification | 7.5% |
| SVM | 5.5% |
| SVM combined with Nearest Center Classification | 4.0% |

It's obvious that the performance of SVM classfier is better than that of BP neural network and nearest center criterion. When SVM combined with nearest center, the error rate decreases 1.5%, which means that about 27.3% misclassification of pure SVMs can be eliminated by subsequent nearest center classification. It shows that although SVMs are superior to nearest center classification, but their discrimitive capabilities do not totally overlaped, based on which we can improve the prformance of classifiers by the combination and SVMs and nearest center criterion.

## 5. Conclusions

We have presented the face recognition experiment using SVMs followed by nearest center classification. As showed in the comparison with other techniques, it appears that the strategy can be effectively trained and tested for the face recognition. The experimental results show that the combination of SVMs with nearest center is a better discriminative algorithm than BP neural network, sole SVMs or the nearest center classification approach for face recognition.

## References


[1] Chellappa R,Wilson C L, and Sirohey S. Human and machine recognition of faces: A survey. Proceedings of the IEEE, 1995, 83(5):705～740.

[2] Turk M. Pentland A. Eigenfaces for recognition, Journal of cognitive neuroscience, 1991,3(1):71～86.

[3] M. Lades, J.C. Vorbruggen, J. Buhmann, J. Lange, C. von der Malsburg, R. P. Wurtz, and W. Konen. Distortion invariant object recognition in the dynamic link architecture. IEEE Transactions on the Computers, 42(3):300-311, Mar 1993.

[4] T.F. Cootes, G.J. Edwards, and C.J. Taylor. Active appearance models. In European Conference in Computer Vision, pages 484-498, New York, NY, 1998.

[5] G. W. Cottrell and M. Fleming, "Face recognition using unsupervised feature extraction," in Proc. Int. Neural Network Conf., vol. 1, Paris, France, July 9–13, 1990, pp. 322–325.

[6] Nefian, A.V., Hayes, M.H., III. Hidden Markov models for face recognition, Acoustics, in Proc. ICASSP 98, vol. 5, pp 2721 - 2724

[7] P. J. Phillips. Support vector machines applied to face recognition. In M. S. Kearns, S. Solla, and D. Cohen, editors, Advances in Neural Information Processing Systems 11, volume 11, pages 803--809. MIT Press, 1998.

[8] Guodong Guo; Li, S.Z.; Kapluk Chan. Face recognition by support vector machines, Proceedings of Fourth IEEE

International Conference on Automatic Face and Gesture Recognition, 2000, Page(s): 196 -201
[9] Cortes C, Vapnic V. Support Vector Networks. Machine Learning, 1995, 20 :1-25.
[10] Burges C J C. A tutorial on support vector machines for pattern recognition. Data Mining and Knowledge Discovery, 1998 2(2)
[11] M.Pontil and A.Verri. Support vector machines for 3-D object recognition. IEEE Trans. on Pattern Analysis and Machine Intelligence.1998,20:637～646.
[12] Hastie, T. and Tibshirani, R. Classification by pairwise coupling. Technical report, Stanford University and University of Toronto. 1996