SVM and Adaboost-based Classifiers with Fast PCA for Face Reocognition

Min Yao
College of Information Engineering
Shanghai Maritime University
Hai Gang Avenue 1550, Shanghai 201306, China
minyao@shmtu.edu.cn

Changming Zhu
College of Information Engineering
Shanghai Maritime University
Hai Gang Avenue 1550, Shanghai 201306, China
cmzhu@shmtu.edu.cn

***Abstract*—Face recognition has been receiving continuous academic and commercial attention for the last decades. In this paper, we construct two face recognition systems adopting SVM and Adaboost as the classifiers with fast PCA for facial feature representation. The detailed discussions about algorithm realization are given. Comparison between the two systems and analysis of them are provided through several experiments on ORL face database. We also test them on face images under various conditions.**

Keywords—face recognition; fast PCA; SVM; Adaboost

I. INTRODUCTION

Over the last decades, face recognition has developed into a major research area and a specialized application area within the larger field of pattern recognition and computer vision [1], [2]. Different from many other pattern recognition issues, face recognition system remains to be specifically defined and depicted, faced with real application conditions and corresponding considerations. For example, there are probably numerous identities (classes) to be recognized, but few samples or even only one sample is provided for each identity. The face recognition process mainly consists of two phases which are feature representation and classification. Selected features are made use of to reduce the dimension of the input face images and should be discriminant to make recognition effective. Classification verifies a new face image as one of the individual identity in the database.

Considerations about feature selection mainly focus on whether the output features are simple and representative enough for discriminate efficiently between the classes. Different research approaches have already been presented for feature reduction like independent component analysis (ICA) [7], multidimensional scaling [8], etc. One of the most common technique is Principal Component Analysis (PCA) [3] which is a statistical method used for reducing the dimension of a data set while retaining the majority of the variation present in the data set. It has been widely used in various information processing and feature extraction areas [4]-[6] and proves efficient for subsequent object classification.

In a simple yet canonical scenario, face matching may be implemented as the subspace projection followed by a classifier [9]. In this paper, we constructing two face recognition systems using SVM (Support Vector Machine) [10], [11] and Adaboost (Adaptive Boosting) [12] algorithm respectively as the classifier, and meanwhile incorporating fast PCA-based facial feature representation into each of the classifiers. We analyze the detailed usage and setting of PCA and the two classifiers, study important considerations about the face recognition systems, and further apply the proposed systems to a famous face database. The data are provided of the parameters and the performance of the two systems run on faces with pose, expression and other face variances.

II. FRAMEWORK AND FEATURE REPRESENTATION

A. Face Recognition Framework

The framework of the face recognition system proposed in this paper is shown in Fig. 1, which includes image pre-processing, face detection, face representation, face classification and the final output.

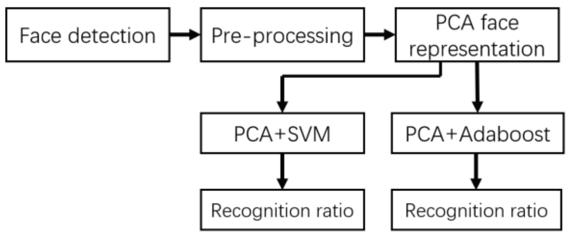

Fig.1. The framework of the face recognition system in our work. It mainly contains four steps including face detection, pre-processing (lighting normalization), face representation (PCA feature extraction) and recognition (SVM and Adaboost classification).

The first face detection step is considered finished when using the ORL face database. In this database, each face image is focused only on the individual facial area and is well-normalized in terms of image size, face location and in-plane rotation. Therefore, in the pre-processing step, we only do some changes to the face pattern not to the shape. We do lighting normalization using histogram equalization [13]. In the face representation step, PCA is adopted for its simplicity, effectiveness and efficiency. In the classification step, classifiers using SVM and Adaboost are executed for comparison study.

B. Fast PCA-based Facial Feature Representation

ORL face database contains 400 face images of 40 persons. Each person has 10 images varying in facial pose or expression. The first 5 images from every individual identity are collected, used for training out eigenfaces and the rest 5 images of each identity are used for face-matching test. We use PCA to create a representative facial feature space out of the training set. PCA takes the area containing the face images as random vectors and thus employs K-L (Karhunen-Loeve Transform) to get an orthogonal basis transformation. As a result, the bases corresponding to larger eigen values have more similar patterns with face images. Via the linear combination of these bases, PCA can describe, represent and approximate each face, which can be used for face reconstruction and identification. The identification is implemented by projecting the tested face image into the subspace expanded with the above-mentioned feature bases and then compare the location of the face image in the subspace with that of the face images from the database. The construction is realized by recovering the face image in the original face space. Each individual face image can be connected to a set of coefficients when projected to the subspace and thus these coefficients can represent the facial features.

Specifically speaking, suppose that there are n face images in the database, which are separately transformed to vector X_1, X_2,...,X_n by stretching each image into one vector. The average face image can be calculated then by

$$X_{\text{ave}} = \frac{1}{n}\sum_{i=1}^{n} X_{\text{i}} \quad , \tag{1}$$

from which we can get the difference between each face image and the average image:

$$X'_{\text{i}} = X_i - X_{ave}, i = 1,2,..., N \ . \tag{2}$$

Then the covariance matrix can be obtained as

$$C = \frac{1}{\text{n}}\sum_{i=1}^{n} X'_i \, (X'_i)^T \quad . \tag{3}$$

From the covariance matrix C, the eigen values λ_k and the corresponding eigen vectors μ_k can be further calculated.

The traditional PCA is usually used for reducing the dimensions of input information and keep the main features of the input. However, input information of high dimensions will cause serious computational cost. For example, one face image in ORL database is of size 112×92. We choose the first 5 images of each individual for training (totally 5×40=200 samples). Then the sample matrix is $Z_{n\times d}$ (n=10304, d=200) and the scatter matrix for calculating eigen vectors will be of size 10304×10304, that is $C_{n\times n}$ (n =10304), which means huge computational complexity. Review the PCA and it can be found that $C_{d\times d}$ (d=200) has the same eigen values with $C_{n\times n}$. Thus we apply the eigen vectors calculated from $C_{d\times d}$ for feature extraction, which guarantees the effectiveness and decrease the computational cost as well.

Fig. 2 shows 20 eigenfaces represented by the most principal eigen vectors, extracted from 200 face images for training. These eigenfaces are referred to as principal components. We can find facial feature-like patterns in these eigenfaces.

Fig. 3 shows one reconstructed face image of an individual from the database using various numbers of principal components. We can see that the face images can be well reconstructed when an appropriate number of principal components are used. This shows that we can use PCA to represent facial features.

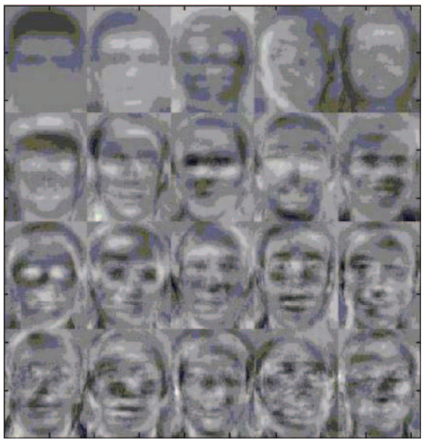

Fig.2. 20 eigenfaces calculated from 200 face images from the ORL databases, which are represented by the most principal eigen vectors of the covariance matrix.

(a) (b) (c)

Fig. 3. Face reconstruction using various numbers of principal components. The face image is of a person from the ORL face database. (a) is the original face image; (b) is the reconstructed result using 50 principal components; (c) is the reconstructed result using 100 principal components.

III. Classifier Design

In our system, we adopt the SVM and Adaboost algorithms as the core of the classifiers. We analyze the theoretical realization and the setting of the corresponding coefficients.

A. SVM Algorithm

SVM[10][11] is powerful tool for classification problems. It is widely used since it achieves an optimal linear classifier (optimal hyperplane) in feature space which is based on Structural Risk Minimization (SRM) theory. It adopts the statistical learning process to gain a linear classifier with minimum VC-dimension (machine complexity), hence the low generalization errors. Also, high dimensional feature spaces can be used through kernel functions. The learning of SVM and its recognition operations only require inner product evaluations of relative feature extraction functions or kernel functions as entitled. The evaluation itself is applicable with much fewer computational efforts than the previous direct feature function evaluation of multiple cases. There are 4 mostly used kernel functions given as follows:

978-1-5090-6193-8/16/$31.00 ©2016 IEEE

$$K(x, y) = x \cdot y, \tag{4}$$

$$K(x, y) = (x \cdot y + 1)^d, \tag{5}$$

$$K(x, y) = \exp(-\gamma \|x - y\|^2), \tag{6}$$

$$K(x, y) = \tanh(b(x \cdot y) - c), \tag{7}$$

which are named as Linear Kernel Function (LKF), Polynomial Kernel Function (PKF), Radial Basis Function (RBF) and Sigmoid Kernel Function (SKF). After choosing one of the kernel function, the parameters and the error cost coefficient C should be defined. In our system, we choose RBF due to its comparable advantages . RBF can handle non-linear separation and LFK is just a special case of RBF; RBF can achieve the same result as LFK by selecting the proper parameter (γ, C). On the other hand, SKF can also be similar to RBF under some condition while two parameters b and c remain to be defined with SKF. PKF may also cause calculation problems such as overflow when calculate the inner product of the kernel function. In our system, we achieve the best face recognition performance on the database with C=128 and γ=0.0782.

B. *Adaboost Algorithm*

AdaBoost was formulated by Freund and Schapire [12] which is considered a relatively efficient, simple, and easy learning strategy for improving the classification performance. It was first applied to face detection by Viola and Jones [14] to select Haar wavelet features and train a cascade of classifiers (two-class classification). It is a kind of adaptive boosting tree, which is able to boost weak learning algorithm into a strong one. The core idea of boosting is to lower the weight of the sample if the sample could be correctly classified and to enlarge the weight if the sample couldn't. In this way, the boosting algorithm is enabled to classify the difficult training samples in the subsequent training. This learning process is repeated, which leads to a well-performing classifier finally. Adaboost uses the boosting idea and meanwhile realizes the whole learning process adaptively, where the weak classifier output is +1 or -1. Adaboost brings each training sample a weight w_i and emphasizes the incorrectly classified sample in the weak classifier of the next iteration.

The main steps of Adaboost is as follows:

- Initialize each sample weight $w_i = \frac{1}{N}$, $i = 1, 2, ...N$
- For each sample, train weak classifiers with the weak classifier leaning algorithm $f_t(x) \in \{-1, 1\}$, and then calculate the error rate $\varepsilon_t = E_\omega[I(y \neq f_t(x))]$ and weighted coefficient $a_t = \log[(1-\varepsilon_t)/\varepsilon_t]$.
- Update sample weights $\omega_i = \omega_i \exp[a_t \cdot I(f_t(x_i) \neq y_i)]$ and execute normalization $\sum w_i = 1$.
- Output classifier $F(x) = \mathrm{sgn}[\sum^T a_t f_t(x)]$.

In order to incorporate the feature selection output into the discrete Adaboost, a weak classifier based on threshold for features of each dimension xi is used in our system during Adaboosting process which is defined as:

$$f(x_j) = a_j \cdot I(x_j \leq \theta_j) + b_j \cdot I(x_j > \theta_j). \tag{8}$$

Given the threshold θ_j, a_j and b_j can be evaluated with 1 or -1 by minimizing the error rate of weighted sample classification.

$$a_j = \mathrm{sgn}\{E_\omega[y \cdot I(x_j \leq \theta_j)]\} = \mathrm{sgn}\{\sum^N \omega_i y_i \cdot I(x_{i,j} \leq \theta_j)\} \tag{9}$$

$$b_j = \mathrm{sgn}\{E_\omega[y \cdot I(x_j > \theta_j)]\} = \mathrm{sgn}\{\sum^N \omega_i y_i \cdot I(x_{i,j} > \theta_j)\} \tag{10}$$

Now the error rate is

$$\varepsilon_t = E_\omega\{I[y \neq f(x_i)]\} = \sum^N \omega_i \cdot I[y_i \neq f(x_{i,j})]. \tag{11}$$

The optimal threshold θ_j for feature j can be got by exhaustive method. Since the number of samples is N, the time complexity for searching the optimal θ_j is O(N). The algorithm choosing the optimal weak classifier each time means choosing the optimal feature. Supposing that there are M features, it needs to select the feature with the lowest error rate, which means computational complexity is O(MN). Thus after T iterations, the computational complexity of the training process is O(TMN). Fig. 4 gives the performance of Adaboost on ORL face database. It shows that the error rate is relatively high at the beginning of the learning among both the training and testing samples. But after 20 iterations, the training error rate is lower than 1% and the testing error rate is also below 2%.

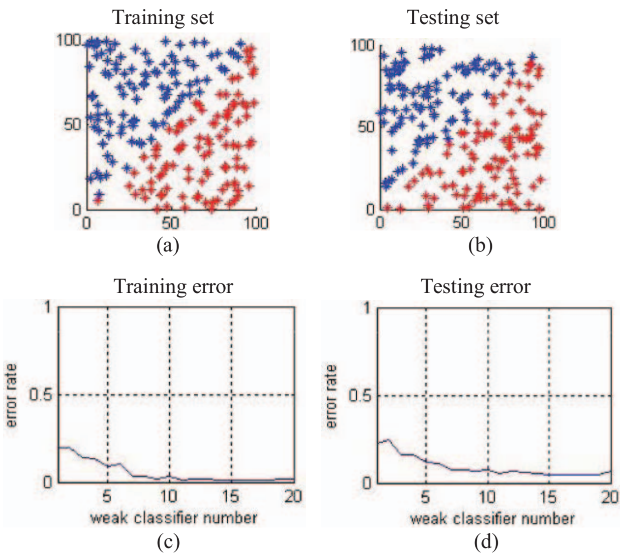

Fig. 4. Adaboost performance test on ORL face database. (a) and (b) show the training set and testing set respectively. (c) and (d) show the training error and testing error respectively.

C. *Classifier Design*

Both SVM and Adaboost algorithms serve a two-class classification. Thus we need to expand them to fit for the face recognition application. There are three common strategies: one-versus-rest strategy, one-versus-one voting strategy and one-versus-one phase-out strategy. We use one-versus-one voting strategy in our system. Assuming that there are four classes A, B, C and D, they are grouped into training set in pairs−(A, B), (A, C), (A, D), (B, C), (C, D) and this results in 6 classifiers (n(n-1)/2 classifiers if class number is n). The tested sample is then delivered through the 6 2-class classifiers and we get a set of results for voting. At first, all the voting counters are set to 0, that is vote(A)= vote(B)= vote(C)= vote(D)=0. During the voting process, if the classifier (A, B) is used and the classifier verifies the sample as class A, then vote(A)= vote(A)+1, otherwise vote(B)= vote(B)+1. The rest 5 classifiers are used in the same way to get the final voting results and the classification is determined by selecting Max{vote(A), vote(B), vote(C), vote(D)}. If more than one result has the maximum value, the class of the first max voting result is simply selected as the that of the sample.

IV. Experiments and Analysis

In order to see the performance of the two face recognition systems using SVM and Adaboost as the classifiers, experiments are conducted on the ORL face database. We separate the whole database into two sets, one of which contains the first 5 images of each individual for training and the rest for testing (i.e. 200 images for training and 200 images for testing).

The dimension of each image is reduced from the original 10304 to much smaller ones by using the fast PCA method and the faces are represented by a series of coefficients. The most discriminative facial features of the original face images are maintained mostly while the dimension largely decreases, which can be seen in Fig. 2. After dimension reduction and normalization, the output face representations are trained through SVM and Adaboost respectively. Via voting strategy, each of the test faces is identified as one of the individual in the database (i.e. face matching).

Table 1 gives the recognition rates of SVM and Adaboost classifiers with different dimension reduction extent. The number of dimension varies from 50 to 250. The corresponding recognition rates show that when the dimension of the principal components increases, the recognition rates goes higher. The general performance of SVM with PCA is relatively encouraging. The training and testing computational cost of the two systems are shown in Table 2. From the table, we can see that PCA+SVM spend more time on testing than PCA+Adaboost since SVM maintains more classifier information. On the other hand, it takes PCA+Adaboost more time to train a strong classifier due to the iteration process of feature selection.

In order to further test the robustness of the two systems, we conduct another experiment to see how they perform with different face conditions such as faces with motion blur, Gaussian blur and lighting changes. Sample images under different conditions and their corresponding matching results are shown in Fig. 5. We can see that most of the results are good except the results on motion blur images of which the features may be too obscure for the classifiers to get and it becomes hard to distinguish between face images.

V. Conclusion

In this paper, we apply SVM and Adaboost classifiers for face recognition and incorporate the fast PCA method to reduce the dimension of input face images and meanwhile maintain most of the intrinsic facial features. The whole framework and each part of it are discussed in detail. We test and compare the performance of the two systems and give the results on face images under various conditions. Generally, PCA+SVM outperforms PCA+Adaboost in our work.

TABLE I. Recognition rate (%) with different dimensions

Dimension	Method	
	PCA+SVM	***PCA+Adaboost***
50	75.0	68.0
100	83.5	73.5
150	88.0	75.0
200	90.5	78.5
250	92.5	79.0

TABLE II. Training time and testing time of the two systems

Time	Method	
	PCA+SVM	***PCA+Adaboost***
For training	30s	120s
For testing	240s	20s

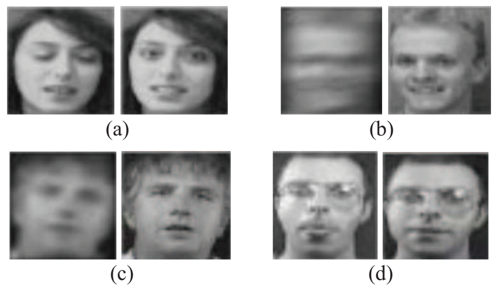

Fig. 5. Samples of the various face conditions for recognition test and the corresponding matching results. The left image of (a), (b), (c) and (d) respectively is normal image in ORL database, image with motion blur, image with Gaussian blur and image with a lighting change. The right ones are the matching results.

TABLE III. Recognition rate (%) under different face conditions

Face condition	Method	
	PCA+SVM	***PCA+Adaboost***
Normal	88.0	75.0
Motion blur	56.5	53.0
Gaussian blur	71.5	66.5
Lighting change	85.5	71.5

Acknowledgment

This work is partially sponsored by Natural Science Foundation of China (61603245 and 61602296), Natural Science Foundation of Shanghai (16ZR1414400) and Shanghai Pujiang Program (16PJ1403700).

References

[1] R. Chellappa, C.L. Wilson, S. Sirohey, "Human and Machine Recognition of Faces: A Survey", Proceedings of the IEEE, vol. 83, no.5, pp. 705-740, May.1995.

[2] Z. Zeng, M. Pantic, G. Roisman, and T. Huang, "A survey of affect recognition methods: Audio, visual and spontaneous face recognitions," IEEE Trans. Pattern Anal. Mach. Intell., vol. 31, no. 1, pp. 39–58, Jan. 2009.

978-1-5090-6193-8/16/$31.00 ©2016 IEEE

[3] S. Z. Li, X. Lu, X. Hou, X. Peng, and Q. Cheng, "Learning multiview face subspaces and facial pose estimation using independent component analysis," IEEE Trans. Image Process., vol. 14, no. 6, pp. 705–712, Jun. 2005.

[4] N. P. Chandrasiri, M. C. Park, T. Naemura, and H. Harashima, "Personal facial expression space based on multidimensional scaling for the recognition improvement," in Proc. 5th Int. Symp. Signal Process. Appl., vol. 2, pp. 943–946, 1999.

[5] M. Turk and A. Pentland, "Eigenfaces for Recognition," J. Cognitive Neurosci, vol.1, pp. 71-86, Mar.1991.

[6] Yanjun Zhang, Xuefeng Ai, "The Identification of Two-Phase Flow Regimes Using HMM Model Based on ERT System and PCA Feature Extraction", 5th Intl. Conf. Intelligent Systems Design and Engineering Applications (ISDEA), pp.1053-1055, 2014.

[7] L. Kuncheva and W. Faithfull, "PCA feature extraction for change detection in multidimensional unlabelled streaming data", 21st Intl. Conf. Pattern Recognition (ICPR), pp. 1140-1143, 2012.

[8] M. Akbar, A. Ali, A. Amira, F. Bensaali, M. Benammar, M. Hassan and A. Bermak, "An Empirical Study for PCA- and LDA- Based Feature Reduction for Gas Identification", IEEE Sensors Journal, vol. 16, no. 14, pp. 5734-5746, 2016.

[9] H. Moon and J. Phillips, "Analysis of PCA-based Face Recognition Algorithms", Empirical Evaluation Techniques in Computer Vision, K. Boyer and J. Phillips, Eds. Los Alamitos, CA: IEEE Computer Society Press, 1998.

[10] C. Cortes and V. N. Vapnik, "Support-vector networks", Machine Learning, vol. 20, no. 3, pp. 273-297, 1995.

[11] N. Cristianini and J. S. Taylor, "An Introduction to Support Vector Machines and Other Kernel-based Learning Methods", pp. 82-108, Cambridge University Press, 2004.

[12] Y. Freund and R. Schapire. A decision-theoretic generalization of on-line learning and anapplication to boosting. Computational Learning Theory: Eurocolt'95, 1:23–37, 1995.

[13] HE: http://en.wikipedia.org /wiki/Histogram_ equalization.

[14] P. Viola and M. Jones, "Rapid object detection using a boosted cascade of simple features", In Proceedings IEEE Conf. on Computer Vision and Pattern Recognition, pp. 511–518, 2001.