

Improving the Recognition of Faces using PCA and SVM Optimized by DWT

Ghali Ahmed
Computer science Department,
USTOMB, Oran, Algeria

Benyettou Mohamed
Computer science Department,
USTOMB, Oran, Algeria

ABSTRACT

Automatic recognition of people has received much attention during the recent years due to its many applications in different fields such as law enforcement, security applications or video indexing. In this paper, Wavelet transform is used for preprocessing the images in order to handle bad illumination. Principle Component Analysis (PCA) is used to play a key role in feature extractor and the SVM were used for classification. Support Vector Machines (SVMs) have been recently proposed as a new classifier for pattern recognition. We illustrate the potential of SVMs on the Cambridge ORL Face database, which consists of 400 images of 40 individuals, containing quite a high degree of variability in expression, pose, and facial details. The SVMs that have been used included the Linear (LSVM), Polynomial (PSVM), and Radial Basis Function (RBF SVM) SVMs, we obtain recognition rates as high as 97,9 in ORL face database with polynomial kernel (PSVM).

General Terms

Support Vector Machines (SVMs), Wavelet transform (WT), Principle Component Analysis (PCA).

Keywords

Wavelets transform; facerecognition; pca; svm (LSVM, PSVM, and RBF SVM).

1. INTRODUCTION

Face recognition has obtained an increasing amount of attention in pattern recognition and computer vision over the past few years. This technology can be used in wide range of applications such as identity authentication, access control, and surveillance. [1].

Various methods have been recently presented for face recognition [2], [3], [4], [5], [6] and [7].

Generally, there are two categories of methods in face recognition [8]. One approach is based on facial feature, such as eyes, nose and mouth first are located and then various feature extraction methods can be adopted to construct feature vectors of these facial features. Finally, traditional pattern recognition methods like a neural network can be used to recognize the feature vectors. The other approach takes a holistic view of the recognition problem. It extracts the statistical characterization by the statistical method directly out of the entire training sample. Holistic feature extraction of face images is adopted in this approach. The most famous technique in the second approach is PCA.

Wavelet transform provides a powerful and versatile framework for image processing. It is widely used in the fields of image denoising, compression, fusion, and so on.

The changes of expressions in the sample images of an individual result in the differences of higher frequency band of the images [9].

Support Vector Machines (SVMs) have been recently proposed by Vapnik and his co-workers [10] as a very effective method for

general purpose pattern recognition. In this research, we focus on the face recognition problem, and show that the discrimination functions learned by SVMs can give much higher recognition accuracy than the popular standard eigenface approach using MLP classifier. Principal Component Analysis (PCA) is used to extract the feature from the face image. After the features are extracted, the SVMs are learned and the disjoint test set enters the system for recognition. We perform experiment on the ORL face database of 400 images of 40 individuals containing quite a high degree of variability in expression, pose, and facial details. The SVMs that have been used included the Linear (LSVM), Polynomial (PSVM), and Radial Basis Function (RBF SVM) SVMs. We provide experimental evidence which show that Polynomial and Radial Basis Function (RBF) SVMs performs better than Linear SVM on the ORL Face Dataset.

2. WAVELET TRANSFORM

Wavelet transform technique is used for analyzing signals. The basic functions of wavelet transform are obtained from a single prototype wavelet (or mother wavelet) by translation and dilation [11]:

$$\psi_{a,b}(t) = \frac{1}{\sqrt{a}} \psi\left(\frac{t-b}{a}\right) \quad (1)$$

Where a and b are both real numbers which quantify the scaling and translation operations. Substitute a and b with 2^m and $n \times 2^m$ respectively, the basic functions become

$$\psi_{a,b}(t) = 2^{-m/2} \psi(2^{-m}t - n) \quad (2)$$

The process of one-dimensional discrete wavelet transform and reconstruction a signal $x(t)$ is defined as follows:

$$W(m,n) = \langle x(t), \Psi_{m,n}(t) \rangle \quad (3)$$

$$\chi(t) = \sum_{m,n} W(m,n) \bar{\Psi}_{m,n}(t) \quad (4)$$

The two-dimensional wavelet transform is got by applying one-dimensional wavelet transform to the rows and columns of two-dimensional data. An approximation image is derived from 1-level wavelet decomposition of an image and three detail images in horizontal, vertical and diagonal directions respectively. The approximation image is used for the next level of decomposition. Fig. 1 is the process of decomposing an image, Fig. 2 is an image from ORL Face Database, Fig. 3 is obtained after one-level wavelet transform, and Fig. 4 is obtained after four-level wavelet transform.

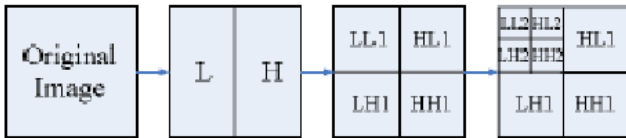


Figure 1: The process of decomposing an image



Figure 2: Original image



Figure 3: After one-level wavelet transform



Figure 4: After four-level wavelet transform

3. PRINCIPAL COMPONENT ANALYSES (PCA)

Let a face image $I(x,y)$ be a two-dimensional N by N array of intensity values or a vector of dimension N^2 . A typical image of size 256 by 256 describes a vector of dimension 65,536, or equivalently, a point in 65,536-dimensional space [12]. Consider our training set of images of 92 by 112 pixels. The main idea of principal component analysis is to find the vectors which best account for the distribution of the face images within the entire image space. Steps for Feature Extraction:

1. The first step is to obtain a set S with M face images. Each image is transformed into a vector of size N and placed into the set.

$$S = \{\Gamma_1, \Gamma_2, \Gamma_3, \Gamma_4, \dots, \Gamma_M\} \quad (1)$$

2. Second step is to obtain the mean image Ψ .

$$\Psi = \frac{1}{M} \sum_{n=1}^M \Gamma_n \quad (2)$$

3. Then find the difference ϕ between the input image and the mean image

$$\Phi_i = \Gamma_i - \Psi \quad (3)$$

4. Next seek a set of M orthonormal vectors, u_n , which best describes the distribution of the data. The k th vector, u_k , is chosen such that

$$\lambda_k = \frac{1}{M} \sum_{n=1}^M (u_k^T \Phi_n)^2 \quad (4)$$

is a maximum, subject to

$$u_l^T u_k = \delta_{lk} = \begin{cases} 1 & \text{If } l=k \\ 0 & \text{Otherwise} \end{cases} \quad (5)$$

Where u_k and λ_k are the eigenvectors and eigenvalues of the covariance matrix C

5. The covariance matrix C has been obtained in the following manner

$$C = \frac{1}{M} \sum_{n=1}^M \Phi_n \Phi_n^T \quad (6)$$

$$= A A^T$$

$$A = \{\Phi_1, \Phi_2, \Phi_3, \dots, \Phi_n\} \quad (7)$$

6. To find eigenvectors from the covariance matrix is a huge computational task. Since M is far less than N^2 by N^2 , we can construct the M by M matrix $L = A^T A$

7. Find the M eigenvector, v_1 of L .

8. These vectors (v_1) determine linear combinations of the M training set face images to form the Eigenfaces u_1

$$u_l = \sum_{k=1}^M v_{lk} \Phi_k \quad l = 1, 2, \dots, M \quad (8)$$

9. Project each of the original images into eigenspace. This gives a vector of weights representing the contribution of each Eigenfaces to the reconstruction of the given image.

$$\omega_k = u_k^T (\Gamma - \Psi) \quad \Omega^T = [\omega_1, \omega_2, \dots, \omega_M] \quad (9)$$

Where u_k is the k th eigenvector and ω_k is the k th weight in the vector

$$\Omega^T = [\omega_1, \omega_2, \omega_3, \dots, \omega_M] \quad (10)$$

4. CLASSIFICATION METHODE: SUPPORT VECTOR MACHINES (SVM)

4.1 An Overview of SVM Theory

SVM method was derived from optimal hyperplane in the linearly separable case. Consider the example of a two-class classification in Fig. 5:

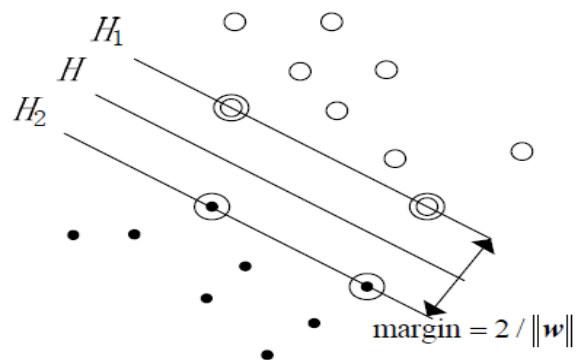


Figure 5: Optimal hyperplane

Here H is the correct classification line, H_1 and H_2 are parallel lines of H . H_1 and H_2 get through the nearest points from H . H is the only line which can both separate the data and maximize the margin (the distance between the hyperplane and the nearest data point of each class), namely optimal line. In high-dimension space, the optimal line becomes optimal hyperplane.

Assumed $(x_i, y_i), i = 1, \dots, n \in \mathbb{R}^d$ to be the sample set, $y_i \in \{+1, -1\}$ is class labels, the linear discriminate function of the d -dimension space is $g(x) = w \cdot x + b$, the corresponding hyperplane equation is:

$$(w \bullet x) + b = 0 \quad (1)$$

And then the optimal hyperplane problem turns into the following constrained quadratic programming problem:

Minimize

$$\phi(w) = \frac{1}{2} \|w\|^2 = \frac{1}{2} (w \bullet w) \quad (2)$$

Subject to

$$y_i [(w \bullet x_i) + b] - 1 \geq 0, i = 1, 2 \dots l. \quad (3)$$

The optimal classification function can be obtained by solving the above problem as follow:

$$f(x) = \text{sgn} \left(\sum_{SV} y_i a_i^0 (x_i \bullet x) - b_0 \right) \quad (4)$$

Here x_i is the support vectors (SVs), a_i^0 is the corresponding Lagrange coefficient, b_0 is the threshold, which can be obtained via (4) with any one SV.

The above conclusions can be generalized to the linear nonseparable case.

4.2 SVM Classification

SVM method is based on the following idea: The input vectors are mapped into a high-dimension space Z via certain non-linear transformation, in which the optimal hyperplane can be constructed. If the dot product in the optimal classification function is replaced with a inner product function (kernel function), the corresponding discriminate function will turn into the following:

$$f(x) = \text{sgn} \left(\sum_{SV} y_i a_i K(x_i, x) - b \right) \quad (5)$$

This is just the discriminate function in the high-dimension space, namely Support Vector Machine.

Various nonlinear discriminate SVM can be constructed adopting different kernel function $K(x_i, x)$. The following are some choices of the kernel function proposed by Vapnik:

- 1) Linear Kernel:

$$K(x, x_i) = x \cdot x_i \quad (6)$$

- 2) Polynomial function

$$K(x, x_i) = [(x \bullet x_i) + 1]^d \quad (7)$$

The corresponding SVM is a classifier of polynomial of degree d .

- 3) Radial basis function

$$K(x, x_i) = \exp \left\{ -\frac{|x - x_i|^2}{\sigma^2} \right\} \quad (8)$$

The SVM is a Gaussian RBF classifier.

- 4) Sigmoid function

$$K(x, x_i) = \tanh(v(x \bullet x_i) + c) \quad (9)$$

4.3 Multi-Class Recognition

A multi-class pattern recognition system can be obtained by combining two classes SVMs. Generally speaking, there are two methods for achieving the purpose. The first method is the one-against-all strategy to classify between each class and all the remaining; the second method is the one-against-one strategy to classify between each pair. We propose to construct a bottom-up binary tree for classification. Suppose there are eight classes in the data set, the decision tree is shown in Fig. 6, where the numbers 1–8 encode the classes. We can notice that the numbers encoding the classes are arbitrary without any means of ordering. By comparing between each pair, one class number is chosen representing with the “winner” of the current two classes. The selected classes from the lowest level of the binary tree will come to the upper level for another round of tests. At last, on the top of the tree the unique class will appear.

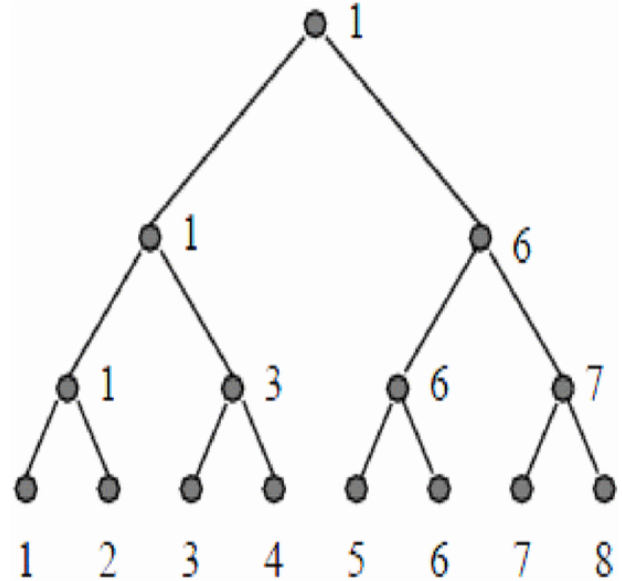


Figure 6: The binary tree structure for 8 classes face recognition.

5. EXPERIMENTS AND RESULTS

Our experiments are based on Cambridge ORL face database and Matlab11.0 is used as programming tool. This face database contains 40 individuals, and each individual has 10 images with variations in pose, illumination, facial expression and accessories. The size of each image is 92×112 pixels; with 256 grey levels per pixel. We show some images from ORL face database in Fig.7.



Figure 7: ORL face database

Our procedures are as follows:

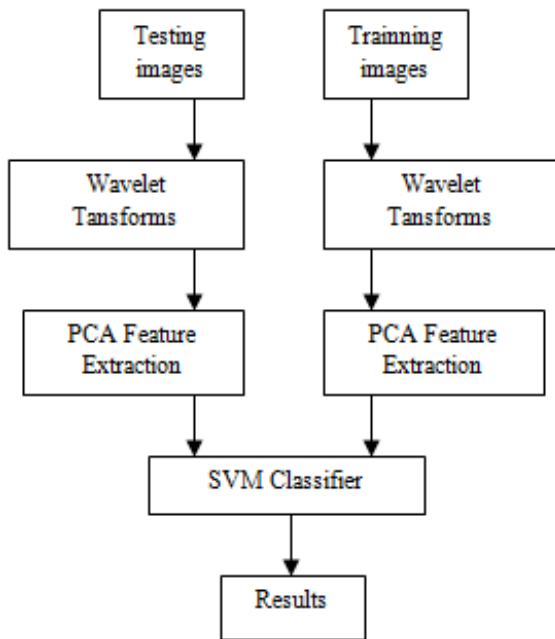


Figure 8: Process of recognition

In this study, we used PCA on wavelet subband for feature extraction. For classification we used SVM. We tested the results on ORL database. Six training sets are created by varying training images per person. For wavelet we used lower frequency subband (A4) after four level decomposition using db4 wavelet. At first step we have performed experiment using SVMs based classifier with PCA like extraction features. The SVMs that have been used included the Linear, Polynomial, and Radial Basis Function (RBF) SVMs, and in the second step we have performed experiment using PCA-SVMs that have been used included the Linear, Polynomial, and Radial Basis Function (RBF) SVMs with wavelet transform. Table. 1 and Fig.10 shows the recognition that use Linear, Polynomial and Radial Basis Function (RBF) SVMs combined with wavelet

transform and PCA obtained highest classification rate. We obtained highest classification rate of 97.9% using Wavelet-PCA and Polynomial kernel of SVM with 5 images of each person in the training set.



Figure 9: Five individuals with five images for each person

Techniques	Pose count per individual in training / number of test images					
	1/(360)%	2/(320)%	3/(280)%	4/(240)%	5/(200)%	6/(160)%
PCA-LSVM	52,1	72,6	81,2	84,2	88,6	87,5
PCA-PSVM	55,2	75,3	84,2	86,6	91,2	89,5
PCA-RBFSVM	46,3	69,6	80,7	84,5	90,6	87
DWT-PCA-LSVM	68,8	89,1	95	95,7	96,4	95,5
DWT-PCA-PSVM	66,9	86,2	94,9	95,8	97,9	97,2
DWT-PCA-RBFSVM	69,4	89,3	95,3	96,2	97,1	96,4

Table1: Recognition rates of applied techniques according to increasing pose count

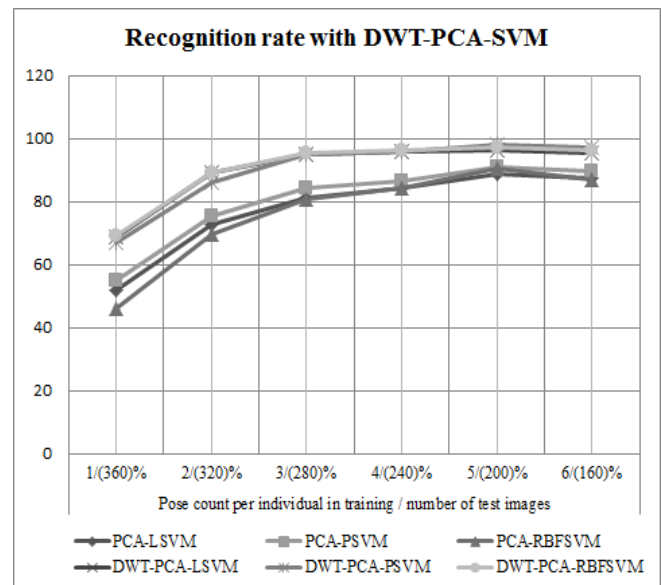


Figure 10: Recognition rates with increasing dimension of features for Linear, Polynomial and RBF functions SVM, with and without Wavelet Transform

6. CONCLUSION

In this study, we have applied two feature extraction methods (Wavelets and PCA) on the ORL face database. For classification step we used SVM (with three types of kernels, Linear, Polynomial and Radial Basis Function).

The results of face recognition can be easily influenced by some variations such as lighting, expression and pose in the face images. Wavelet transform provides a powerful method for image processing, it is universally used in the fields of image de-noising, compression, fusion, etc. Our experiment results demonstrate that this method can get better recognition rates.

As future work, we have been applying an ANN (Artificial Neural Network) classifier on extracted feature vectors. Our preliminary results also indicate that better recognition rates are obtained with Wavelet approach in contrast to PCA based Eigenfaces approach.

7. REFERENCES

- [1] R. Chellappa, C. L. Wilson, and S. Sirohey. Human and machine recognition of faces: A survey. *Proc. IEEE*, 83:705-741, May 1995.
- [2] B. Moghaddam, T. Jebara, A. Penland, "Bayesian face recognition," *Pattern Recognition*. New York, vol. 33, pp. 1171-1182, 2000.
- [3] Qingshan Liu, Rui Huang, Hanqing Lu, Songde Ma, "Face recognition using kernel based fisher discriminant analysis," *Proceedings of the 5th IEEE International Conference on Automatic Face and Gesture Recognition*, pp. 187-191, May, 2002.
- [4] O. Deniz, M. Castrillon, M. Hernandez, "Face recognition using independent component analysis and support vector machines," *Pattern Recognition Letters*, vol. 24, pp. 2153-2157, Sep. 2003.
- [5] G. D. Guo, S. Z. Li, K. Chan, "Face recognition by support vector machines," *Proceedings of the 4th IEEE International Conference on Automatic Face and Gesture Recognition*, pp. 196-201, Mar. 2000.
- [6] Jing Zhang, Xue-Dong Zhang, Seok-Wun Ha, "A novel approach using PCA and SVM for face detection," *Proceedings of 4th International Conference on Natural Computation*, vol. 3, pp. 29-33, 2008.
- [7] Lam.A, Shelton.C.R, "Face recognition and alignment using support vector machines," 2008 8th IEEE International Conference on Automatic Face & Gesture Recognition, 6 pp. Sep. 2008.
- [8] S. Ranganath and K. Arun, "Face Recognition Using Transform Features and Neural Network," *Pattern Recognition*, Vol..30, Oct.1997,pp. 1615-1622.
- [9] C. Nastar and N. Ayache, "Frequency-Based NonrigidMotion Analysis: Application to Four Dimensional Medical Images," *IEEE Trans. Pattern Analysis and Machine Intelligence*, Vol. 18, no. 11, Nov. 1996,pp. 1,067-1,079.
- [10] V. N. Vapnik. *Statistical learning theory*. John Wiley & Sons, New York, 1998.
- [11] Yugang Jiang, Ping Guo. "Face Recognition by Combining Wavelet Transform and k-Nearest Neighbor." *Journal of Communication and Computer*, Vol.2, Sep.2005, pp.50-53.
- [12] J. Ruiz-del-Solar and P. Navarrete, Eigenspacebased face recognition: a comparative study of different approaches, *IEEE Transactions on Systems, Man and Cybernetics, Part C*, Vol. 35, Issue 3, Page(s):315-325, Aug. 2005.