Face Recognition Technique Using PCA, Wavelet and SVM

Manisha Satone[1] / Gajanan Kharate[2]

[1]Sinhagad college of engineering, Sinhagad road, Pune

[2]University of Pune, Ganeshkhind, Pune

Abstract:

We present a method of face recognition in which features are extracted by applying Principal component analysis on wavelet subband. Support vector machine and nearest distance methods are used for classification. Results are tested on ORL database and obtained highest classification accuracy 97.5% for 6 images per person in training set.

Keywords: Face recognition, Wavelet, Principal component analysis, Support vector machine.

1. Introduction

Biometric-based technologies include the identification based on physiological characteristics such as face, fingerprints, hand geometry, hand veins, palm, iris, retina, ear, voice and behavioral traits such as gait, signature and keystroke dynamics [1]. These biometric technologies require some voluntary action by the user. However, face recognition can be done passively without any explicit action or participation of the user, since face images can be acquired from a distance by a camera. The face reorganization system is more appropriate for security and surveillance purposes. Facial images can be easily obtained with a couple of inexpensive fixed cameras; they cannot be modified or forged. Face images are not affected by background sound noise. Face recognition algorithms with preprocessing of the images may compensate for noise, slight variations in orientation, scale and illumination.

Face recognition by computer can be divided into two approaches [2], namely, constituent-based and face-based. In constituent-based approach, recognition is based on the relationship between human facial features such as eyes, mouth, nose, profile silhouettes and face boundary [3]. The success of this approach highly depends on the accuracy of the facial feature extraction schemes. Every human face has similar facial features; a small deviation in the extraction may introduce a large classification error. Face-based approach [4] uses the face as a whole for recognition. Many face based recognition algorithms have been developed and each has its strength. Principal Component Analysis (PCA) [5] has been proven to be an effective face-based approach. Sirovich and Kirby [6] first proposed a method using Karhunen-Loeve (KL) transform to represent human faces. In their method, faces are represented by a linear combination of weighted eigenvectors, known as eigenfaces. Turk and Pentland [7] developed a face recognition system using PCA. However, common PCA-based methods suffer from two limitations, namely, poor discriminatory power and large computational load. In view of the limitations of the existing PCA-based approach, we applied a PCA on wavelet subband of face image. Face image is decomposed into a number of subbands with different frequency components using the wavelet transform (WT). Low frequency subband at 4^{th} level is used for PCA. After each decomposition level resolution of image decreases which decreases computation and increases speed.

In this study, we used ORL database to test the performance of proposed method. Wavelet and PCA are used for feature extraction whereas support vector machine (SVM) and nearest distance classifier are used for classification of images.

This paper is organized as follows. Section 2 introduces Wavelet and PCA which were used for feature extraction. In section 3 classification method, SVM is explained. Section 4 shows the experimental results and section 5 presents conclusion.

2. Introduction of PCA and Wavelet

2.1 PCA

Principal component analysis is based on the second order statistics of the input image. It is a standard technique used in statistical pattern recognition and signal processing for data dimensionality reduction and feature extraction. As the pattern often contains redundant information, mapping it to a feature vector can get rid of this redundancy and yet preserve most of the intrinsic information content of the pattern. These extracted features have greater role in distinguishing input patterns. Every test image can be transformed to low dimensional feature vector to be projected onto the eigenface space which was obtained from the training set. This feature vector can then be compared with the set of feature vectors obtained from the training set. The face classifier can use different classification techniques to classify the images. Some important details of the PCA are highlighted as follows.

Let $X = \{X_n \quad R^d \mid n = 1,.., N\}$ be an ensemble of vectors.

In imaging applications, they are formed by row concatenation of the image data, with *d* being the product of the width and the height of an image.

Let A_x be the average vector in the ensemble.

$$A_x = \frac{1}{N}\sum_{n=1}^{N} X_n \tag{1}$$

Where N is the total number of images.

After subtracting the average from each element of X, we get a modified ensemble of vector

$$\overline{X_n} \equiv X_n - A_x \tag{2}$$

$$\overline{X} = \{\overline{X_n}, n = 1.....N\} \tag{3}$$

The auto-covariance matrix M for the ensemble X is defined by

$$M = Cov(\overline{X}) \tag{4}$$

where *M* is a *d* x *d* matrix, with elements

$$M(i,j) = \frac{1}{N}\sum_{n=1}^{N} \overline{X}_n(i)\overline{X}_n(j) \tag{5}$$

$$1 \le i, j \ge d$$

The eigenvectors of the matrix *M* form an orthonormal basis for R^d. This basis is called the K-L basis. Since the auto-covariance matrix for the K-L eigenvectors are diagonal, it follows that the coordinates of the vectors in the sample space *X* with respect to the K-L basis are un-correlated random variables. Let $\{Y_n, n=1,.., d\}$ denote the eigenvectors of M and let *K* be the d x d matrix whose columns are the vectors $Y_1,.., Y_d$. The adjoint matrix of the matrix *K*, which maps the standard coordinates into K-L coordinates, is called the K-L transform. In many applications, the eigenvectors in *K* are sorted according to the eigenvalues in a descending order. The PCA of a vector *y* related to the ensemble *X* is obtained by projecting vector *y* onto the subspaces spanned by *d'* eigenvectors corresponding to the top *d'* eigenvalues of the autocorrelation matrix *M* in descending order. Where *d'* is smaller than *d*. This projection results in a vector containing *d'* coefficients $a_1, .., a_d$'. The vector y is then represented by a linear combination of the eigenvectors with weights $a_1, .., a_d$'.

Basically, eigenface is the eigenvector obtained from PCA. In face recognition, each training image is transformed into a vector by row concatenation. The covariance matrix is constructed by a set of training images. The significant features (eigenvectors associated with large eigenvalues) are called eigenfaces. The projection operation characterizes a face image by a weighted sum of eigenfaces. Recognition is performed by comparing the weight of each eigenface between unknown and reference faces.

PCA has been widely adopted in human face recognition and face detection since 1987. However, inspite of PCA's popularity, it suffers from two major limitations: poor discriminatory power and large computational load.

2.2 Wavelet:

Wavelets are mathematical functions that cut up data into different frequency components. They have advantages over traditional Fourier methods in analyzing physical situations where the signal contains discontinuities and sharp spikes. Wavelet Transform (WT) has been a very popular tool for image analysis in the past ten years. The advantages of WT are good time and frequency localizations. In the proposed system, WT is chosen to be used in image frequency analysis and image decomposition because:

• By decomposing an image using WT, the resolutions of the subband images are reduced. In turn, the computational complexity will be reduced dramatically by working on a lower resolution image.

• Wavelet decomposition provides local information in both space domain and frequency domain[8].

Wavelet transform can be performed for every scale and translations, resulting in continuous wavelet transform (CWT), or only in multiples of scale and translation intervals, resulting in discrete wavelet transform (DWT). Since, CWT provides redundant information and requires a lot of computation; generally DWT is preferred. A two dimensional wavelet transform is derived from two one-dimensional wavelet transform by taking tensor products. The implementation of WT is carried out by applying a one-dimensional transform to the rows of the original image data and the columns of the row transformed data respectively.

In this paper, an image is decomposed into four subbands using daubechies db4 wavelet transform. The low frequency subband (*LL)* at 4^{th} level, which is a coarser approximation of the original image is selected for PCA.

3. SVM:

Support Vector Machine (SVM) method is a classification method which separates the two data sets by searching for an optimal separating hyperplane (OSH) between them. It is proposed by Vapnik [9]. If data is not linearly separable then it is transformed into new space using kernel and then finds the OSH. If data is not separable then it searches the OSH which maximizes the margin and minimizes the misclassifications. Figure 1 shows two data sets separated by hyperplane. The vector *w* is a normal vector, b is constant and x is an input data point vector.

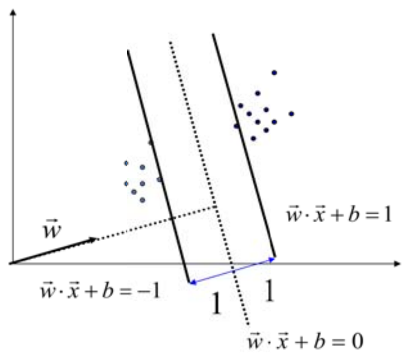

Fig.1 Data sets separated by hyperplanes

The width of margin is given by $\frac{2}{\|w\|}$

Our aim is to maximize marginal width, so

max $\frac{2}{\|w\|}$

$s.t.(w.x+b)\geq 1, \forall x$ of class 1

$(w.x+b)\leq -1, \forall x$ of class 2

Let y_i indicate classes

$$y_i = \{1, -1\} \quad (6)$$

Then $y_i = (w.x_i + b) \geq 1, \forall x_i$ (7)

So the optimization problem becomes

$$\max \frac{2}{\|w\|} \quad \text{s.t. } y_i = (w.x_i + b) \geq 1, \forall x_i \quad (8)$$

or

$$\min \frac{1}{2}\|w\|^2 \quad (9)$$

s.t. $y_i(w.x_i + b) \geq 1, \forall x_i$

This is constrained optimization problem and can be solved by the Lagrangian Multiplier method. For SVM Lagrangian multiplier equation can be written as

$$L_D = \sum_i \alpha_i - \frac{1}{2}\sum_{i,j} \alpha_i \alpha_j y_i y_j x_i x_j \tag{10}$$

α is the Lagrangin multiplier and optimization problem is maximize L_D and find values of α. There is one value of α associated with each support vector. An instant x can be classified by determining the side of the decision boundary it falls. It is by checking the sign of following equation.

$$sign(\sum_{i=1}^{l} \alpha_i y_i (x_i x) + b) \tag{11}$$

If data is not linearly separable then it is transformed into high dimensional space using kernel function. The learning task involves maximization of the following objective function

$$L_D = \sum_i \alpha_i - \frac{1}{2}\sum_{i,j} \alpha_i \alpha_j y_i y_j k(x_i, x_j) \tag{12}$$

Now classification depends on

$$sign(\sum_{i=1}^{l} \alpha_i y_i k(x_i, x) + b) \tag{13}$$

Real world data sets contain noise and cannot be separated by an optimal hyperplane. In this case slake variable ξ is used and constraint becomes

$$y_i(w.x_i + b) \geq 1 - \xi_i, \forall x_i \tag{14}$$

$$\xi_i \geq 0$$

Objective function penalizes for misclassified instances and those within the margin. The optimization problem becomes

$$\min \frac{1}{2}\|w\|^2 + C\sum_i \xi_i \tag{15}$$

C is a trade off parameter between margin width and misclassification.

For multiclass classification one-verses-all approach is used. k-1 classes are combined out of k classes into a single class and trains it against the remaining class. To train all classes, the procedure is repeated for each class in k, thus the training results in k SVM's.

Polynomial kernel function is given by

$$K(x_i, x) = (x_i.x + 1)^p$$

Where p is 1 for linear and 2 for quadratic polynomial kernel.

4. Experimental results:

Features are extracted using wavelet and PCA and for classification SVM and nearest distance classifier are used. ORL database is used to test the proposed method. For feature extraction we use MATLAB software 7.0. For SVM data mining software Weka 3.7.9 [10] is used. ORL database [11] has 400 images of 40 persons and there are 10 poses for each person. Results are tested by varying the number of training images of each person and remaining poses are chosen as the test set. Training set includes 40,80,120,160,200 and 240 images.

Images are decomposed up to four levels using db4 wavelet transform. Low frequency band (LL) at 4^{th} level is used for PCA calculation and feature extraction. The 4^{th} level subband reduces the resolution 92 x112 of original image to 13 x 12. Using PCA on a low resolution image, reduces the computations. For SVM with Weka, Polynomial kernels are used. Parameter 'C' is chosen as 1.0. We obtained highest recognition rate of 97.5 % using db4 wavelet and linear polynomial kernel. Results are shown in table 1.

Table 1.
Recognition rates according to increasing pose count for ORL database.

Technique	Number of training images per person/number of test images					
	1/360	2/320	3/280	4/240	5/200	6/160
Wavelet-PCA-SVM (Poly-Linear)	73	89.6	95	96.2	96.2	97.5
Wavelet-PCA-SVM (Poly-Quad)	70.8	86.5	94.2	95.8	96	96.8
Wavelet-PCA-ND	75.5	88.4	94.6	95.8	96	96.8

5. Conclusion:

In this study, we used PCA on wavelet subband for feature extraction. For classification we used SVM and nearest distance approaches. 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. For SVM we used linear polynomial kernel and quadratic polynomial kernel. We obtained highest classification rate of 97.5% using linear polynomial kernel with 6 images of each person in the training set.

References:

[1] Perronnin, F., Dugelay, J.-L. An introduction to biometrics and face recognition In: Proc. IMAGE'2003: Learning, Understanding, Information Retrieval, Medical, Cagliari, Italy, June 2003.
[2] R. Chellappa, C. L. Wilson and S. Sirohey. Human and machine recognition of faces: a survey. Proceedings of the IEEE, Vol. 83, No. 5,705-740, 1995.
[3] F. Goudail, E. Lange, T. Iwamoto, K. Kyuma and N. Otsu. Face recognition system using local autocorrelations and multiscale integration", IEEE Trans. PAMI, 1996,Vol. 18, No. 10, 1024-1028.
[4] M. Kirby and L. Sirovich. Application of the Karhunen-Loeve procedure for the characterization of human faces", IEEE Trans. PAMI., 1990 Vol. 12, 103-108.
[5] M. I. Shah and D. C. Sorensen. symmetry preserving singular value decomposition. SIAM J. Matrix Anal. Appl., 28(3):749–769, 2006.
[6] L.Sirovich and M. Kirby. Low-dimensional procedure for the characterization of human faces. J. Opt. Soc. Am. A, Vol. 4, No. 3, 519-524, 1987.
[7] M. Turk and A. Pentland. Eigenfaces for recognition. J. Cognitive Neuroscience, Vol. 3, 71-86,1991.
[8] I. Daubechies.Ten Lectures on Wavelets. CBMS-NSF series in Applied Mathematics, Vol.61,SIAM Press, Philadelphia, 1992.
[9] Vapnik, V. N. Statistical learning theory. New York: 1998 John Wiley & Sons.
[10] Mark Hall, Eibe Frank, Geoffrey Holmes, Bernhard Pfahringer, Peter Reutemann, Ian H. Witten 2009, The WEKA Data Mining Software: An Update; SIGKDD Explorations, Volume 11, Issue 1.
[11] ORL Database of Faces. http://www.cl.cam.ac.uk/research/dtg/attarchive/facedatabase.html.