



Evaluation of face recognition techniques using PCA, wavelets and SVM

Ergun Gumus^a, Niyazi Kilic^{b,*}, Ahmet Sertbas^a, Osman N. Ucan^b

^a Istanbul University, Engineering Faculty, Computer Eng. Dept., 34320 Avcilar, Istanbul, Turkey

^b Istanbul University, Engineering Faculty, Electrical and Electronics Eng. Dept., 34320 Avcilar, Istanbul, Turkey

ARTICLE INFO

Keywords:

Face recognition
Wavelet transform
Principal Component Analysis
Support Vector Machines

ABSTRACT

In this study, we present an evaluation of using various methods for face recognition. As feature extracting techniques we benefit from wavelet decomposition and Eigenfaces method which is based on Principal Component Analysis (PCA). After generating feature vectors, distance classifier and Support Vector Machines (SVMs) are used for classification step. We examined the classification accuracy according to increasing dimension of training set, chosen feature extractor–classifier pairs and chosen kernel function for SVM classifier. As test set we used ORL face database which is known as a standard face database for face recognition applications including 400 images of 40 people. At the end of the overall separation task, we obtained the classification accuracy 98.1% with Wavelet–SVM approach for 240 image training set.

© 2014 Elsevier Ltd. All rights reserved.

1. Introduction

As a special study of pattern recognition, face recognition has had crucial effects in daily life especially for security purposes. Face recognition task is actively being used at airports, employee entries, criminal detection systems, etc. For this task many methods like Elastic Matching (Zhang, Yan, & Lades, 1997), Neural Network based approaches (Lawrence, Giles, Tsoi, & Back, 1997), Independent Component Analysis (Yuen & Lai, 2002) and Eigenfaces (Turk & Pentland, 1991) have been proposed and tested. Most of these methods have trade off's like hardware requirements, time to update image database, time for feature extraction, response time.

Generally face recognition methods are composed of a feature extractor (like PCA, Wavelet decomposer) to reduce the size of input and a classifier like Neural Networks, Support Vector Machines, Nearest Distance Classifiers to find the features which are most likely to be looked for. In this study, we chose wavelet decomposition and Eigenfaces method which is based on Principal Component Analysis (PCA) as main techniques for data reduction and feature extraction. PCA is an efficient and long term studied method to extract feature sets by creating a feature space. PCA also has low computation time which is an important advantage. On the other hand because of being a linear feature extraction method, PCA is inefficient especially when nonlinearities are present in the underlying relationships (Kursun & Favorov, 2004).

Wavelet decomposition is a multilevel dimension reduction process that makes time–space–frequency analysis. Unlike Fourier

transform, which provides only frequency analysis of signals, wavelet transforms provide time–frequency analysis, which is particularly useful for pattern recognition (Gorgel, Sertbas, Kilic, Ucan, & Osman, 2009).

In this study, we used available 40 classes in the ORL face recognition dataset (ORL Database of Faces, 1994). Eigenfaces and Discrete Wavelet Transform are used for feature extractor. For the classification step, we consider Support Vector Machines (SVM) and nearest distance classification and all results obtained are evaluated.

The content of this work is organized as follows. Section 2 introduces feature extraction methods which are Eigenfaces and wavelet transform methods. In Section 3, classification methods which are SVM and nearest distance criterions are given. Section 4 shows the experimental results and Section 5 presents conclusions.

2. Feature extraction methods

2.1. Eigenfaces method

Eigenfaces method is a kind of Principal Component Analysis (also known as Karhunen–Loeve transform) which is specialized with face images. In PCA, every image in training set is identified with feature vectors that are provided from projection of the image to the basis in image space. Generally PCA method classifies images according to distance between feature vectors. Standard classifiers include nearest distance criterion, Euclidean distance and nearest mean classification.

Using PCA for Eigenfaces method, feature vectors identifying each image can be obtained as follows:

* Corresponding author. Tel.: +90 5443217319; fax: +90 2124737064.
E-mail address: niyazik@istanbul.edu.tr (N. Kilic).

- (i) Let's assume that we have N face images with m rows and m columns. $\Gamma_1, \Gamma_2, \dots, \Gamma_N$ represent those images in column vectors with $(m^2 \times 1)$ dimension. We need to calculate a mean face image (Ψ), from these vectors like,

$$\Psi = \frac{1}{N} \cdot \sum_{i=1}^N \Gamma_i \quad (1)$$

- (ii) After calculating mean face image, each face image's distance from mean face image should be calculated as Φ_i column vector,

$$\Phi_i = (\Gamma_i - \Psi) \quad (2)$$

- (iii) Φ_i column vectors are gathered in matrix $D = [\Phi_1, \Phi_2, \dots, \Phi_N]$ with dimension $(m^2 \times N)$ and a covariance matrix C is formed as,

$$C = D \cdot D^T \quad (3)$$

Calculating m^2 eigenvalues and m^2 eigenvectors of covariance matrix C causes a great computational complexity. To avoid this complexity as stated in Turk and Pentland (1991) we can choose covariance matrix C in $(N \times N)$ dimension like,

$$C = D^T \cdot D \quad (4)$$

- (iv) At this step we calculate N eigenvalues (λ_k) and N eigenvectors (v_k) of C to form eigenface space. $V = [v_1, v_2, \dots, v_N]$ represents a matrix including eigenvectors of C with dimension of $(N \times N)$. We can obtain eigenface space $U = [u_1, u_2, \dots, u_N]^T$ by,

$$U = V \cdot D^T \quad (5)$$

All row vectors of U are "eigenfaces" of face images in training set. Face images with higher eigenvalues have more contribution to eigenface space. For this reason, systems with low computational capability sort eigenvectors of face images according to their corresponding eigenvalues in decreasing order and choose first Z eigenvectors to form a smaller eigenface space.

- (v) Matrix $W = [w_1, w_2, \dots, w_N]$ with dimension $(N \times N)$ includes N column vectors corresponding to each face image in training set. These vectors are called "feature vectors" and they represent each image's specific characteristics. W can be obtained like,

$$W = U \cdot D \quad (6)$$

After obtaining eigenface space and feature vectors, we can compare a test image with the faces in training set by following steps:

- (a) Γ_T is a column vector which represents our test image with $(m^2 \times 1)$ dimension. At this stage distance of test image from mean face image should be calculated as Φ_T column vector,

$$\Phi_T = (\Gamma_T - \Psi) \quad (7)$$

- (b) After calculating Φ_T , we must project it on our eigenface space in order to obtain its feature vectors in format of column vector w_T with dimension $(N \times 1)$,

$$w_T = U \cdot \Phi_T \quad (8)$$

- (c) To find out which image in training set resembles our test image, we need to find similarity of w_T to each w_i in matrix W . Various classifiers can be used at this step. Techniques we used are Support Vector Machines and "Nearest Distance" criterion that are stated in Section 3.

2.2. Wavelet transform method

Wavelet decomposition has been applied in many studies for image processing successfully. In our study, we use multilevel (4-level) DWT (Discrete Wavelet Transform) type wavelet decomposition on our image set.

By application of wavelet transform on an image we obtain four sub-bands including the approximation (low frequency component which is low-low) and the details (high frequency components like low-high, high-low and high-high). Approximation (a) is a smaller scaled form of input image and details are horizontal (h), vertical (v) and diagonal (d) Gorgel et al., 2009. After 1-level DWT, an image (I) can be represented with its sub-bands like,

$$I = I_a^1 + \{I_h^1 + I_v^1 + I_d^1\} \quad (9)$$

To reduce the dimension of data we work on, we can apply DWT on I_a^1 N times to have an N -level decomposition. In our study with 4-level DWT, an image is represented like,

$$I = I_a^4 + \sum_{i=1}^4 \{I_h^i + I_v^i + I_d^i\} \quad (10)$$

After four decomposition levels, we obtain $Y/16 \times Z/16$ sized approximation from $Y \times Z$ sized input image. For our image set approximation is a feature set with 42 elements.

3. Classification methods

3.1. Support Vector Machines (SVM) method

SVM is a classification method that aims to separate two data sets with maximum distance between them. It is proposed by Vapnik (1998). This method separates two data sets by searching for an optimal separating hyperplane (OSH) between them. Bounds between data sets and OSH are called "support vectors".

Each point in total data set is referred as $x_i \in \mathbb{R}^n$, $i = 1, 2, \dots, N$ and belongs to a class $y_i \in \{-1, 1\}$. For linear classification we can identify two classes and the OSH separating them like,

$$w \cdot x_i + b \geq 1, \quad y_i = 1 \quad (11)$$

$$w \cdot x_i + b \leq -1, \quad y_i = -1 \quad (12)$$

We can generalize (11) and (12) with the form,

$$y_i \cdot [(w \cdot x_i) + b] \geq 1, \quad i = 1, \dots, l \quad (13)$$

The distance between support vectors are pre-defined as:

$$d = \frac{2}{\|w\|} \quad (14)$$

The bigger d is, a better separation between two classes can be achieved. For this reason to maximize d we need to minimize norm of w . This problem can be solved using Lagrange function,

$$L(w, b, \alpha) = \frac{\|w\|^2}{2} - \sum_{i=1}^l \alpha_i \cdot \{y_i \cdot [(w \cdot x_i) + b] - 1\} \quad (15)$$

Here α_i represents Lagrange multipliers. Solving (15) by minimizing according to w and b , maximizing according to $\alpha_i \geq 0$ values, most suitable OSH parameter w can be obtained in (16) according to condition $\sum_{i=1}^l \alpha_i \cdot y_i = 0$, $\alpha_i \geq 0$, $i = 1, \dots, l$,

$$w = \sum_{i=1}^l \alpha_i \cdot y_i \cdot x_i, \quad \alpha_i \geq 0, \quad i = 1, \dots, l \quad (16)$$

Distance of any data point x to OSH is defined as:

$$d(w, b; x) = \frac{|w \cdot x + b|}{\|w\|} \quad (17)$$

We can get a more generalized form of (17) by replacing w with its value shown in (16),

$$d(x) = \frac{\left(\sum_{i=1}^l \alpha_i \cdot y_i \cdot x_i\right) \cdot x + b}{\left\|\sum_{i=1}^l \alpha_i \cdot y_i \cdot x_i\right\|} \quad (18)$$

Sign of distance calculated in (18) shows us to which class point x belongs and $|d|$ shows distance of x to OSH. As $|d|$ increases a better classification result can be obtained.

Linear separation of data sets can not be achieved successfully all the time. In such cases a simple conversion of feature space is done. Point x in first data space is expanded to a feature space with higher dimension and linear separation is retried. This expansion process is realized with operator $\Phi(\cdot)$ OSH function turns into the form:

$$f(x) = w \cdot \phi(x) + b \quad (19)$$

By replacing w with its value in (16) we can get a more generalized form as:

$$f(x) = \sum_{i=1}^l \alpha_i \cdot y_i \cdot (\phi(x_i) \cdot \phi(x)) + b \quad (20)$$

In a high dimensional space realization of $(\phi(x_i) \cdot \phi(x))$ multiplication is intractable. For this reason “Kernel Functions” in $K(x_i, x) = (\phi(x_i) \cdot \phi(x))$ form are used. In such processes there are two widely used kernel functions:

(i) Polynomial Kernel Function:

$$K(x_i, x) = (x_i \cdot x + 1)^p \quad (21)$$

(ii) RBF Kernel Function:

$$K(x, x_i) = \exp \left[-\gamma \|x - x_i\|^2 \right] \quad (22)$$

We examined classification performances of both kernels in this study.

3.2. Nearest Distance criterion

According to the Nearest Distance criterion each distance $\text{Dist}(i)$ between a test sample x and a training sample x_i can be calculated as:

$$\text{Dist}(i) = \sum_{j=1}^d |x(j) - x_i(j)|, \quad i = 1, 2, \dots, N \quad (23)$$

The training image with the minimum $\text{Dist}(i)$ value is the most similar one to the test image where d is the dimension of the extracted feature vector using PCA or Wavelet method.

4. Experimental results

We applied each feature extraction method with both SVM and Nearest Distance Classifiers on the ORL face database. We extracted PCA feature vectors with an application program coded in Visual Basic 6.0 (also nearest distance classification is done by this program) and Wavelet vectors using Matlab 7.0. SVM classification is achieved by data mining software called Weka 3.5 (Weka, 2007). Tests were done on a PC with Intel Pentium D 2.8-GHZ CPU and 1024-MB RAM.

In this study, standard ORL images (10 poses for each of 40 people) were converted into JPEG image format without changing their size. For both feature extraction methods a total of six training sets were composed that include varying pose counts (from 1 to 6) for each person and remaining poses are chosen as the test set. Our training sets include 40, 80, 120, 160, 200 and 240 images according to chosen pose count. For each person, poses with the same indices are chosen for the corresponding set.

For PCA based Eigenfaces approach, size of each feature vector is determined by the size of eigenface space. As the training set grows, this size reaches up to 240 (six pose for each person). For a simpler and more feasible classification process we utilize only first 40 elements of each feature vector. (In creation of eigenface space, eigenvectors were re-arranged by sorting their corresponding eigenvalues.) Thus we use these 40 features for SVM and nearest distance classification.

For Wavelet approach, we use a HAAR type level-4 wavelet decomposition to obtain feature vectors with the size of 1×42 . Thus we can make a fair comparison between classification results of PCA and Wavelet approaches.

During SVM classification with Weka, we chose parameter “c” as 1.0. While using RBF kernel we chose gamma as 0.1. For polynomial kernel exponent was 1 for linear and 2 for quadratic forms.

Table 1 illustrates the average recognition rates of classification methods using PCA and Wavelet feature extractors according to increasing pose count. Here RBF, POLY-LINEAR and POLY-QUAD stand for RBF kernel, Linear polynomial kernel and Quadratic polynomial kernel for SVM, respectively. ND stands for “Nearest Distance” Classifier.

As one can notice from Table 1, as the pose count increases accuracy of recognition improves. In case of six poses for each individual (totally 240 training images) a peak level (98.1%) in recognition performance is obtained with Wavelet–SVM (Quadratic polynomial kernel) method.

For all kinds of classification methods HAAR type level-4 wavelet decomposition based approach has shown better recognition results than PCA based approach. Considering the weighted mean

Table 1
Recognition rates of applied techniques according to increasing pose count.

Technique	Pose count per individual in training/(number of test images)						Weighted mean
	1/(360) (%)	2/(320) (%)	3/(280) (%)	4/(240) (%)	5/(200) (%)	6/(160) (%)	
Wavelet–SVM (RBF)	69.4	89	95.3	96.2	96.5	97.5	88.5
Wavelet–SVM (POLY-LINEAR)	68.8	89.3	95	95.8	96.5	97.5	88.3
Wavelet–SVM (POLY-QUAD)	66.9	86.2	93.9	95.4	97.5	98.1	87.2
Wavelet–ND	71.1	88.1	94.6	95.4	97	97.5	88.5
PCA–SVM (RBF)	55.2	75.3	84.2	86.6	89.5	91.2	77.5
PCA–SVM (POLY-LINEAR)	54.1	74.6	83.2	86.2	89.5	90.6	76.7
PCA–SVM (POLY-QUAD)	46.3	69.6	80.7	84.5	87	90.6	72.9
PCA–ND	54.7	71.2	82.1	86.6	87	90.6	75.7

Table 2

Recognition rates for ORL Face database.

Eigenface (Lawrence et al., 1996)	SOM + CN (Lawrence et al., 1996)	WT + SVM (Luo et al., 2005)	This study
81.8%	88.2%	94.8%	95.3%

of recognition rates for six training sets, Wavelet-SVM (Quadratic polynomial kernel) has the lowest mean recognition rate (87.2%) and PCA-SVM (RBF kernel) has the highest mean recognition rate (77.5%) for PCA based approaches.

For SVM classification, according to weighted means (percentage of correctly classified test images) we obtained highest mean recognition rates with RBF kernel and lowest mean recognition rates with Quadratic polynomial kernel.

We did not observe much gap between the mean recognition rates of nearest distance classifier and SVM classifiers. For Wavelet based approach, SVM (RBF kernel) and Nearest Distance Classifiers have highest means of 88.5%.

Considering previous studies (Lawrence, Giles, Tsoi, & Back, 1996; Luo, Zhang, & Pan, 2005) on ORL database we obtained higher accuracy rates with wavelet transform based recognition for small training sets (3 samples for each individual in training set, 7 in test set). This is shown in Table 2.

We have also evaluated recognition rates according to gender. ORL database have 4 female (40 poses) and 36 male (360 poses) individuals. Female individuals with the ID's 8, 10, 32 and 35 can be seen in Fig. 1 from top to bottom.

We chose Wavelet-SVM with RBF kernel approach (due to highest mean recognition rate) in order to measure recognition rates for each gender. False drops for female individuals can be seen in Table 3. X^Y denotes that individual X was chosen as perfect match for Y times. Females are confused with males (27 times) more than other females (only 2 times for individual #8).

In total false drop count is calculated as 29 out of 156 test images which leads us to a recognition rate of 81.41% for female individuals. Considering male individuals, false drop count is 149 out of 1404 test images. Thus we get a recognition rate of 89.38% for male individuals.

5. Conclusion

In this study, we have applied two feature extraction methods (PCA and Wavelets) on the ORL face database. For classification step using the extracted features, we used both SVM (with three types of kernels) and nearest distance classification approaches. We have created six training sets to compare classification accuracies of these approaches with various pose counts per individual. For PCA based Eigenfaces method we used all eigenfaces to create feature vectors. For wavelet approach, we used level-4 Haar type wavelet decomposition. In SVM classification step, we used both RBF and Polynomial Kernels with optimum parameters for high classification rates. We obtained highest recognition rate as 98.1% with Wavelet-SVM (Quadratic polynomial kernel) method. Considering weighted means of recognition rates, Wavelet based recognition gave better results than PCA based approach.

We also compared recognition rates according to gender and noticed that male individuals have higher recognition rate (89.38%) than females (81.41%).

**Fig. 1.** Female individuals of ORL database.**Table 3**

False drops for female individuals according to increasing pose count.

ID	Pose count per Individual in training/(number of test images)					
	1/(36)	2/(32)	3/(28)	4/(24)	5/(20)	6/(16)
8	–	–	–	–	–	–
10	7 ¹ , 9 ¹ , 23 ³ , 38 ¹	–	–	–	8 ¹	8 ¹
32	15 ²	2 ¹ , 14 ¹	2 ¹	–	–	–
35	18 ² , 25 ⁵	3 ¹ , 25 ² , 40 ¹	3 ¹ , 25 ² , 40 ¹	15 ¹	–	–

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.

Acknowledgement

This work is supported by Istanbul University Scientific Research Project Fund. No: UDP-6204.

References

- Gorgel, P., Sertbas, A., Kilic, N., Ucan, O. N., & Osman, O. (2009). Mammographic mass classification using wavelet based support vector machine. *Istanbul University-Journal of Electrical and Electronics Engineering*, 9(1), 867–875.
- Kursun, O., & Favorov, O. (2004). SINBAD automation of scientific discovery: From factor analysis to theory synthesis. *Natural Computing*, 3(2), 207–233.
- Lawrence, S., Giles, C., Tsoi, A., & Back, A. (1996). Face recognition: A hybrid neural network approach. Technical report UMIACS-TR-96-16. University of Maryland.
- Lawrence, S., Giles, C., Tsoi, A., & Back, A. (1997). Face recognition: A convolutional neural network approach. *IEEE Transactions on Neural Networks*, 8, 98–113.
- Luo, B., Zhang, Y., Pan, Y.H., (2005). Face recognition based on wavelet transform and SVM. In *Proceedings of the 2005 IEEE international conference on information acquisition, June 27–July 3, 2005, Hong Kong and Macau, China*.
- ORL Database of Faces, (1994). <http://www.cl.cam.ac.uk/research/dtg/attarchive/facedatabase.html>.
- Turk, M. A., & Pentland, A. P. (1991). Eigenfaces for recognition. *Journal of Cognitive Neuroscience*, 3(1), 71–86.
- Vapnik, V. N. (1998). *Statistical learning theory*. New York: John Wiley & Sons.
- Weka 3, (2007). Data Mining with Open Source Machine Learning Software in Java. The University of Waikato, <http://www.cs.waikato.ac.nz/ml/weka/>.
- Yuen, P. C., & Lai, J. H. (2002). Face representation using independent component analysis. *Pattern Recognition*, 35(6), 1247–1257.
- Zhang, J., Yan, Y., & Lades, M. (1997). Face recognition: Eigenface, elastic matching and neural nets. *Proceedings of the IEEE*, 85(9).