



Feature Selection Using Genetic Algorithm for Face Recognition Based on PCA, Wavelet and SVM

Manisha Satone¹ and Gajanan Kharate²

¹Dept of Electronics and Telecommunication Engineering, Sinhagad College of Engineering, Pune.

²Dean , Pune University, Pune.

Abstract: Many events, such as terrorist attacks, exposed serious weaknesses in most sophisticated security systems. So it is necessary to improve security data systems based on the body or behavioral characteristics, called biometrics. With the growing interest in the development of human and computer interface and biometric identification, human face recognition has become an active research area. Face recognition offers several advantages over other biometric methods. Nowadays Principal Component Analysis (PCA) has been widely adopted for the face recognition algorithm. Yet still, PCA has limitations such as poor discriminatory power and large computational load. This paper proposed a novel algorithm for face recognition in which a low frequency component of the wavelet is used for PCA representation. Best features of PCA are selected using the genetic algorithm (GA). Support vector machine (SVM) and nearest neighbor classifier (ND) are used for classification. Classification accuracy is examined by changing number of training images, number of features and kernel function. Results are evaluated on ORL, FERET, Yale and YaleB databases. Experiments showed that proposed method gives a better recognition rate than other popular methods.

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

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 recognition 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.

Many researchers have proposed and developed the machine based face recognition algorithms. They discussed some of the possible problems for the machine based face recognition systems, which are facial expression change, illumination change, aging, rotation, size of the image, occlusions, etc. Modification of the partial information because of obstruction can dramatically affect face recognition performances, for example the face of the person with and without spectacles. Due to these factors large variations of facial images of the person are obtained and it is the challenge for the researcher to propose the accurate algorithm for face recognition.

Research has been conducted vigorously in this area for the past four decades, and huge progress has been made. Still there is scope for improvement. Encouraging results have been obtained and current face recognition systems have reached a certain degree of maturity when operating under constrained conditions. However, they are far from achieving the ideal of being able to perform adequately in all the various situations. The ultimate goal of researchers in this area is to develop the sophisticated face recognition system which minimizes the false recognition under all above discussed constraints and to emulate the human vision system.

Received: August 30th, 2013. Accepted: January 29th, 2014

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.

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. The face image is decomposed into a number of subbands with different frequency components using the wavelet transform (WT). Low frequency subband at 4th level is used for PCA. All feature vectors of PCA are not important for face recognition and their selection should depend on the application and database used. In this study, we used Genetic algorithm to select a best subset of eigenfeatures in order to improve face recognition performance. After each decomposition level, using wavelet transform, resolution of image decreases which decreases computation and increases speed. For classification of face images we have used SVM and the results are compared with a nearest neighbor classifier.

This paper is organized as follows. Section 2 and 4 introduce PCA and wavelet transform which are used for feature extraction. Section 3 explains Genetic algorithm which is used for selection of features. In section 5 and 6 explanation of SVM and nearest neighbor classifier are given. Explanation of block diagram is given in Section 7. Section 8 is for databases used. Section 9 shows the experimental results and section 10 presents conclusions.

2. 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 spaces 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. Some important details of the PCA are highlighted as follows [8].

$$\text{Let } X = \{X_n \in R^d \mid n = 1, \dots, N\} \quad (1)$$

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.

$$A_x = \frac{1}{N} \sum_{n=1}^N X_n \quad (2)$$

Where N is the total number of images.

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

$$\bar{X}_n \equiv X_n - A_x \quad (3)$$

$$\bar{X} = \{\bar{X}_n, n = 1, \dots, N\} \quad (4)$$

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

$$M = Cov(\bar{X}) \quad (5)$$

Where M is a $d \times d$ matrix, with the elements

$$M(i, j) = \frac{1}{N} \sum_{n=1}^N \bar{X}_n(i) \bar{X}_n(j) \quad (6)$$

$$i \geq 1, j \leq d$$

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 of 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 uncorrelated random variables.

Let $\{Y_n, n=1, \dots, d\}$ denote the eigenvectors of M and let K be the $d \times d$ matrix whose columns are the vectors Y_1, \dots, 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 $a_1, \dots, a_{d'}$. The vector y is then represented by a linear combination of the eigenvectors with weights $a_1, \dots, a_{d'}$.

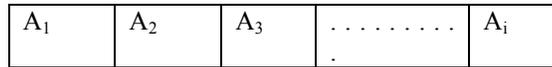
Basically, eigenface is the eigenvector obtained from PCA. In face recognition, each training image is transformed into a vector by row concatenation. The covariance matrix is constructed from a set of training images. The significant features are called eigenfaces. The projection operation characterizes 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, in spite of PCA's popularity, it suffers from two major limitations: poor discriminatory power and large computational load.

3. Feature selection using a genetic algorithm

Feature selection is the first and one of the most important steps in pattern recognition. In the eigenspace, all the eigenvectors are not equally informative. Generally, in most of the techniques the order of the eigenvalues determines the importance of eigenvectors, but this order is not always suitable to describe the data. For example, the first few eigenvectors seem to encode lighting while other eigenvectors seem to encode features such as glasses or mustaches. Although many of the eigen-features are very important for face recognition, they might actually confuse the classifier in other applications. In this study, we used Genetic

algorithm to select a best subset of eigen-features in order to improve face recognition performance. It reduces computation and increases recognition rate. Moreover selection of features depends upon database used.

Genetic Algorithms are search algorithms based on natural selection and natural genetics [9,10]. A binary encoding scheme is used to represent the presence or absence of a particular eigenvector from the linear expansion of the training images. Chromosome representation is as shown in figure 1.



1 bit for each feature, $A_i = 0$: the feature is not selected and $A_i=1$: the feature is selected

Figure 1. Chromosome representation

Each chromosome is checked to see how good it is at solving the problem at hand and assign a fitness score accordingly. The fitness score is a measure of how good that chromosome is at solving the problem. Here we used recognition rate as a fitness score of chromosomes. To find the recognition rate for fitness score, we used separate training and testing sets. Training set consists of one frontal image of each of 20 persons and another 20 images of same 20 persons are used for testing. Two parent chromosomes are selected from the current population. The chance of being selected is proportional to the chromosome's fitness. The roulette wheel method is used to select parent chromosomes. The population's total fitness score is represented by a pie chart, or roulette wheel. To each member of the population a slice of the wheel is assigned. The size of the slice is proportional to the individual fitness score. The member is selected by spinning the ball.

Depending on the crossover rate, the bits from each chosen chromosome are crossover at a randomly chosen point. Single-point crossover, two-point crossover and uniform crossover can be used. In this study we used single point crossover.

Suppose

Parent 1: 1 1 0 0 1 0 1 0 1 1

Parent 2: 1 0 0 1 0 1 1 1 0 1

And crossover point is selected after 4th bit then child chromosomes will be

Child 1: 1 1 0 0 0 1 1 1 0 1

Child 2: 1 0 0 1 1 0 1 0 1 1

Step through the chosen chromosomes bits and flip depending on the mutation rate. This is the chance that a bit within a chromosome will be flipped. This is usually a very low value for binary encoded genes, say 0.001. Find the two worst individuals and replace them with the offsprings.

This process of:

- (i). Selecting two parents
- (ii). Allowing them to create two offspring
- (iii). Immediately replacing the two worst individuals in the population with the offspring is repeated until a stopping criterion is reached. We used count equal to 15 to repeat above process. Lastly features selected in the chromosome having highest fitness score are used for further processing of PCA and then for classification.

4. Wavelet Transform

A face recognition system, generally, consists of two major stages, a feature extraction and a classification stage. The accuracy of the system depends strongly on features extracted to represent the face images and classification methods used to discriminate among the faces. The purpose of feature extraction is to provide useful information which represents the face images without redundancy. Moreover it can greatly reduce the dimensionality of the original image representation.

The wavelets are a set of functions that result from a shift and dilation of the original waveform. All wavelet transforms are forms of time-frequency joint representation with coefficients in a linear combination of the wavelet functions [11,10]. A family of wavelets can be obtained by scaling wavelet function ψ by s and translating it by u .

$$\psi_{u,s}(t) = s^{-1/2} \psi\left(\frac{t-u}{s}\right) \tag{7}$$

The DWT of a 1-D signal $f[n]$ with period N is computed as

$$Df[n, a^j] = \sum_{m=0}^{N-1} f[m] a^{-j/2} \psi\left(\frac{m-n}{a^j}\right) \tag{8}$$

Where m and n are integers.

The value of a is equal to 2 for a dyadic transform. The information corresponding to the scales larger than a^j is also required, which is computed by a scaling filter and is given by

$$SFf[n, a^j] = \sum_{m=0}^{N-1} f[m] a^{-j/2} \phi\left(\frac{m-n}{a^j}\right) \tag{9}$$

Where $\phi(n)$ is the discrete scaling (low-pass) filter.

For the 2-D data, the implementation is carried out by applying a 1-D transform to all the rows of the input image data, and then repeating on all of the columns. The 2-D transform uses a family of wavelet functions and its associated scaling function to decompose the original image into different subbands, namely the low-low (A), low-high (H), high-low (V) and high-high (D) sub-bands. The band LL is a coarser approximation of the original image. The bands LH and HL record the changes of the image along horizontal and vertical directions. While the HH band shows the high frequency component of the image. Each of the channels of the wavelet decomposition can be further decomposed using the discrete wavelet transform, thus a multi-level representation of the face is obtained. A 4 level decomposition of LL subband is shown in figure 2.

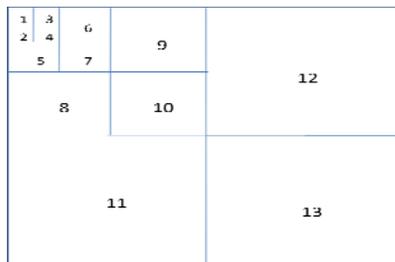


Figure 2. A 4- level decomposition of LL subband

In this paper ORL, FERET, Yale, Yale B databases are used to perform the evaluation of the proposed method. We used Daubechies 4 (db4) wavelet for our experiment because of its symmetry, compact support and the use of overlapping windows to reflect all changes between pixel intensities. Named after Ingrid Daubechies, the Daubechies wavelets are a family of orthogonal wavelets defining a discrete wavelet transform and characterized by a maximal number of vanishing moments. With each wavelet type, there is a scaling function which generates an orthogonal multiresolution analysis. The names of the Daubechies family wavelets are written dbN, where N is the order of the wavelet.

For choosing the WT subband, we use two criteria. The first is to reduce the computational complexity. As discussed above the computational complexity of PCA-based method is in the cubic order of image resolution, d , or number of training images, N , depending on which value is smaller. To minimize the computational complexity, we choose to work on those subbands with lower resolution. Accordingly, subbands 1 to subband 4 are chosen. Nastar et al. [13] have investigated the relationship between variations in facial appearance and their deformation spectrum. They found that facial expressions and small occlusion affect the intensity manifold locally. Under the frequency-based representation, only the high frequency spectrum is affected. Whereas changes in illumination affect the intensity manifold globally, in which only the low frequency spectrum is affected. When there is a change in human face, all frequency components will be affected. Since the LL subband image is a coarser approximation of the original image, we have chosen an LL subband image after 4 level decomposition and we call it as A4 subband image. Since the resolution of A4 subband is less compared to level1, level2 and level3, it will reduce the computation.

5. Support Vector Machine

Support Vector Machines (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 [14]. 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 [15]. Figure 3 shows two data sets separated by hyperplane. Vector w is a normal vector, b is constant and x is an input data point vector.

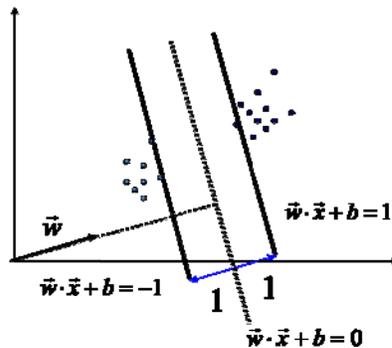


Figure 3. Data sets separated by hyperplanes

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

The aim is to maximize marginal width, so

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

$$\begin{aligned} s.t. (w \cdot x + b) &\geq 1, \forall x \quad \text{Of class 1} \\ (w \cdot x + b) &\leq -1, \forall x \quad \text{Of class 2} \end{aligned}$$

Let y_i indicate classes

$$y_i = \{1, -1\} \quad (10)$$

$$\text{Then } y_i = (w \cdot x_i + b) \geq 1, \forall x_i \quad (11)$$

So the optimization problem becomes

$$\max \frac{2}{\|w\|} \quad (12)$$

$$\text{S.t. } y_i = (w \cdot x_i + b) \geq 1, \forall x_i$$

$$\text{or } \min \frac{1}{2} \|w\|^2 \quad (13)$$

$$\text{S.t. } y_i (w \cdot x_i + b) \geq 1, \forall x_i$$

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

$$L_D = \sum_i \alpha_i - \frac{1}{2} \sum_{i,j} \alpha_i \alpha_j y_i y_j x_i x_j \quad (14)$$

α is the Lagrangian multiplier.

The optimization problem is maximizing 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.

$$\text{sign}\left(\sum_{i=1}^l \alpha_i y_i (x_i \cdot x) + b\right) \quad (15)$$

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

$$L_D = \sum_i \alpha_i - \frac{1}{2} \sum_{i,j} \alpha_i \alpha_j y_i y_j k(x_i, x_j) \quad (16)$$

Now classification depends on

$$\text{sign}\left(\sum_{i=1}^l \alpha_i y_i k(x_i, x) + b\right) \quad (17)$$

Where $k(x_i, x)$ is the kernel function.

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

$$\begin{aligned} y_i(w \cdot x_i + b) &\geq 1 - \xi_i, \forall x_i \\ \xi_i &\geq 0 \end{aligned} \quad (18)$$

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

$$\min \frac{1}{2} \|w\|^2 + C \sum_i \xi_i \quad (19)$$

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.

The polynomial kernel function is given by

$$K(x_i, x) = (x_i \cdot x + 1)^p \quad (20)$$

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

6. Nearest neighbor classifier

Let X, Y is the feature vectors of length n. Then we can calculate the distances between these feature vectors using Manhattan distance (L1 metrics) as in equation 21.

$$d(X, Y) = L_{p=1}(X, Y) = \sum_{i=1}^n |x_i - y_i| \quad (21)$$

In this paper results of SVM classifier are compared with a nearest neighbor classifier.

7. Proposed system block diagram

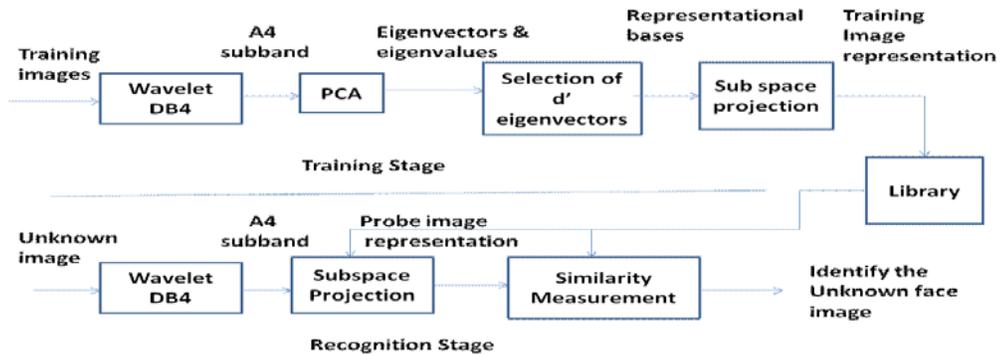


Figure 4. Block diagram of proposed system

A face recognition system is developed to demonstrate the application of the proposed method. The system follows the face-based approach and it consists of two stages, namely, training and recognition stages. Training stage computes the representational bases for images in the domain of interest (that is reference images) and converts them into training image representations. The training image representations of each image are stored in the library. Using the representational bases recognition stage translates the testing image into probe image representation. Testing image is then matched with reference images which are stored in the library to identify the face image. Figure 4 shows the proposed system block diagram.

A. Training Stage

There are five major steps in the training stage. First, a 4-level Wavelet Transform (WT), using the Daubechies wavelet db4 is applied to decompose the reference images. In the second step, PCA is applied on the A4 subband images. The collection of the A4 subband images will form a new ensemble. An auto-correlation matrix will be constructed for the ensemble as discussed in section 2. The output of this step will be a set of eigenvectors and eigenvalues. By arranging the eigenvalues in a descending order, the third step is to select the d' eigenvectors using genetic algorithm which are used as the representational bases. These d' eigenvectors span a subspace $E = \text{span}\{e_1, \dots, e_{d'}\}$. After that, the subband A4 images, x_4 , of all reference images are represented by a linear combination of M representational bases by projecting them into the M eigen subspace. The resultant of the projection of x_4 into E is called Training Image Representation. Lastly, the training image representations and the representational bases are stored in the library.

B. Recognition Stage

Recognition stage is divided into 3 steps. Firstly, an unknown image y is presented to the recognition stage, it is subtracted by the mean value of the reference images and then, a 4-level WT is applied. In the second step, subband A4 subimage is represented as a testing image. Then subband A4 image is projected into the subspace $E = \text{span}\{e_1, \dots, e_M\}$, which is obtained in the training stage and we get the Probe Image Representation. In the last step, classification between the test image and the reference images in the library is performed to determine whether the input test image matched any of the images in the library using SVM and compared with a nearest neighbor classifier.

8. Databases used

We used four databases to evaluate the performance of the proposed system. The FERET database, which has become the standard for evaluating the face recognition technologies, consists of more than 13,000 facial images of more than 1500 subjects. The diversity of the FERET database is related to gender, ethnicity and age. Images are acquired during different photo sessions hence the illumination conditions, facial expressions and the size of the face have been varied. The dataset used in our experiments consists of 2000 FERET face images corresponding to 250 subjects. The images are of size 256 x 384 with 8-bit resolution. Three sets consisting of galleries and probes were used for evaluating the performances of the algorithms. The training set is the set of known facial images, which are used for training. The images in the probe set are the unknown facial images to be recognized. The galleries corresponding to Set1, Set 2 and Set3 consist of one (fa), two (fa+rb), and three (fa+rb+ql) images per subject of 250 subjects. The fa indicates a regular frontal image, rb indicates images corresponding to the phase angle of $+10^0$, and ql indicate images corresponding to pose angle of -22.5^0 . The probe sets corresponding to Set1, Set2 and Set3 were formed from the combination of two (fa+rb), three (fb+rc+qr) and four (fb+rc+qr+ql) images per subject, respectively. The fb indicates an alternative frontal image, taken seconds after the corresponding fa. The qr represents the pose angled image to $+22.5^0$.

The ORL face database [16] consists of 400 images with 10 different images for each of the 40 distinct subjects. The variations of the images are across pose, size, time and facial

expression. All the images were taken against a dark homogeneous background with the subjects in an upright, frontal position, with tolerance for some side movements. The spatial resolution is of 92 x 112, while gray level resolution is 8-bit.

In Yale data base there are 165 images, which are of 15 different subjects with 11 images per subject. The images vary in facial expressions and lighting conditions. YaleB database consists of 10 subjects with each subject having 65 images under different lighting conditions, which are generally used to investigate the effect of illumination. The spatial resolution is of 640 x 480. The images are normalized to extract the facial regions that contain only face, so that the performances of the algorithms are not affected by the factors not related to the face. In normalization, the center of the eyes is detected and images are cropped to the size of 128 x 128 to extract the facial region.

9. Experiments and Results

We used wavelet and PCA for feature extraction. Images are decomposed up to four levels using DB4 wavelet transform. Low frequency band (LL) at 4th level is used for PCA calculation. A genetic algorithm is used for feature selection. SVM and nearest distance classifier are used for classification. Results are tested on ORL, FERET Yale and YaleB databases. For feature extraction we use MATLAB software 12 and for SVM data mining software Weka 3.7.9 [17] is used.

Experiment 1: Recognition rates by varying % of feature vectors selected using ORL database

In the first experiment we used all 400 images of ORL database. Images are decomposed using db4 wavelet and A4 subband images are used for PCA. By varying the % of eigenvectors selected images are classified using SVM. In the first case 4 images of each person are used for training and remaining 6 images of each person are used for testing. In the second case 6 images of each person are used for training and remaining 4 images of each person are used for testing. Results are compared using SVM and nearest distance classifier. For SVM results are evaluated using polynomial linear kernel, radial basis kernel and polynomial quadrature kernel. The highest recognition rate is 99.3 % using the proposed method for linear polynomial kernel and with 25 % of eigenvectors selected using GA. For more than 25 % eigenvector selection, no much improvement in recognition rate is observed. Results are shown in table 1. Compared to the results from paper [18] recognition rate is improved using the proposed method.

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

Technique	Number of training images per person/number of test images							
	4/240				6/160			
	% eigenvectors selected for total eigenvectors using a genetic algorithm							
	10%	15%	20%	25%	10%	15%	20%	25%
Wavelet-PCA-SVM (Poly-Linear)	90	95.8	96.6	97.5	91.8	95	98.7	99.3
Wavelet-PCA-SVM (Poly-Quad)	89.1	94.1	95	96.2	91.2	94.3	96.8	97.5
Wavelet-PCA-SVM (Radial Basis)	89.1	95	95	95.8	91.2	94.6	96.8	97.5
Wavelet-PCA-ND	89.1	95	95	95.8	91.2	93.1	96.8	97.5

Experiment 2: Recognition rates for normal faces

The second experiment compares the performance of the proposed approach with some of the existing approaches for normal faces. The normal faces are the front faces with average illumination and a neutral facial expression. To test the performance FERET, ORL, Yale and YaleB databases are used. For each database 3 images of each person are used for training and 4 images of each person are used for testing. 25 % of total eigenfeatures are selected using genetic algorithm and for the SVM linear polynomial kernel is used. For the methods without GA, 70 % of top eigenvectors having highest eigenvalues are used. Results are shown in table 2.

Table 2. Recognition rate for normal faces

Algorithm	Recognition rate (%)			
	ORL	FERET	Yale	YaleB
PCA-ND	88	86	88.3	95
Wavelet-ND	82	83.6	90	92.5
Wavelet-PCA-ND	90	92	91.6	96.3
Wavelet-PCA-SVM	93.7	93	93.3	96.3
Wavelet-PCA-GA-ND	94.3	94.1	93.3	97.5
Wavelet-PCA-GA-SVM	97.3	98	100	100

For normal faces, most of the algorithms can achieve a high recognition rate. All the algorithms considered for Yale and YaleB databases achieved a high recognition rate because only 15 and 10 different subjects are used respectively.

Experiment 3. Recognition rates for different sets of galleries and probes using FERET databases

In this experiment, methods used in experiment 2 are used for different sets of galleries and probes using FERET databases. Using genetic algorithm 25 % of total eigenvectors are selected. For SVM linear polynomial kernel is used. For the methods without GA, 70 % of top eigenvectors having highest eigenvalues are used. Results are shown in table 3.

Table 3. Recognition rates for different sets of galleries and probes using FERET databases

Algorithm	Recognition rate (%)		
	Set 1	Set 2	Set 3
PCA-ND	84.2	84.3	86
Wavelet-ND	81.8	83.1	83.6
Wavelet-PCA-ND	88.8	90	92
Wavelet-PCA-SVM	89.2	90.9	93.3
Wavelet-PCA-GA-ND	90.8	91.86	94.2
Wavelet-PCA-GA-SVM	96.9	97.3	97.3

Proposed method, Wavelet-PCA-GA-SVM performed well because of the combination of multiresolution features with PCA, selection of best eigenfeatures using GA and power of classification of SVM technique.

Experiment 4: Recognition rates under different facial expressions

An experiment was conducted based on expression variations of the face images with different facial expressions to assess the robustness of the proposed approach to variations in

facial expression. In this experiment, two frontal normal images per subject from FERET, ORL and Yale databases were used for training. For testing 350 images from FERET, 63 images from ORL and 75 images from Yale databases, having facial expression variations were used. Using genetic algorithm 25 % of total eigenvectors are selected. For SVM linear polynomial kernel is used. For the methods without GA, 70 % of top eigenvectors having highest eigenvalues are used. The recognition rates for different databases are given in table 4.

Table 4. Recognition rates under different facial expressions

Algorithm	Recognition rate (%)		
	ORL	FERET	Yale
PCA-ND	87.3	85.1	88.3
Wavelet-ND	80.9	83.4	88
Wavelet-PCA-ND	88.8	90	90.6
Wavelet-PCA-SVM	88.8	91.4	90.6
Wavelet-PCA-GA-ND	90.4	93.1	92
Wavelet-PCA-GA-SVM	98.4	97.7	97.7

Wavelet decomposition of LL subimage and best feature selection using GA make the proposed approach invariant to facial expression variations.

Experiment 5: Recognition rates under different illumination conditions.

We evaluated the effect of illumination variation on face recognition, using FERET and YaleB databases. In FERET database, two normal images per subject were used for training and two images per subject, which have variations in illumination, were used for testing. In Yale B database, two normal images per subject were used for training and five images per subject with illumination variations were used for testing. To remove the effect of illumination variations, first three eigenvectors (after arranging eigenvectors in descending order according to their eigen values) are not considered for GA algorithm since they contain information about illumination variation. GA is used for remaining eigenfeatures and 25 % of eigenfeatures are used for further calculations. For the methods without GA, 70 % of top eigenvectors having highest eigenvalues are used. For SVM linear polynomial kernel is used. The results of these experiments are given in table 5.

Table 5. Recognition rates under different illumination conditions.

Algorithm	Recognition rate (%)	
	FERET	YaleB
PCA-ND	54	62.2
Wavelet-ND	80.5	89
Wavelet-PCA-ND	88.2	92
Wavelet-PCA-SVM	89	92
Wavelet-PCA-GA-ND	90.2	92
Wavelet-PCA-GA-SVM	95.8	92.5

Since first three eigenfeatures have not been used for GA, illumination variation information is removed and proposed method gives best recognition rate.

10. Conclusions

In this paper, we have proposed a technique for face recognition, which uses the combination of wavelet and PCA for feature extraction. The DWT derives the multiresolution features and PCA transform these features into low dimensional space. Only the best eigenfeatures of PCA are selected using genetic algorithm. Since the eigenfeature selection depends upon the database and for classification SVM is used, there is improvement in recognition rate using the proposed method. The paper has also compared the performances of the different face recognition algorithms in terms of changes in facial expression and illumination. Experiments were conducted on different databases. The results show that proposed approach has higher classification accuracy. It is also invariant to illumination and facial expression variations. With the proposed method, the recognition rates for normal images based on FERET, ORL, Yale, and YaleB databases are 97.3%, 98%, 100% and 100% respectively.

References

- [1] Perronnin, F., Dugelay, J.-L. (2003), "An introduction to biometrics and face recognition". In: *Proc. IMAGE'2003: Learning, Understanding, Information Retrieval, Medical, Cagliari, Italy, June*.
- [2] R. Chellappa, C. L. Wilson and S. Sirohey, (1995), "Human and machine recognition of faces: a survey", *Proceedings of the IEEE*, Vol. 83, No. 5, 705-740.
- [3] F. Goudail, E. Lange, T. Iwamoto, K. Kyuma and N. Otsu, (1996). "Face recognition system using local autocorrelations and multiscale integration", *IEEE Trans. PAMI*, Vol. 18, No. 10, 1024-1028.
- [4] M. Kirby and L. Sirovich (1990), "Application of the Karhunen-Loeve procedure for the characterization of human faces", *IEEE Trans. PAMI*, 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] Xiang Xu, Wanquan Liu, Svetha Venkatesh, "An innovative face image enhancement based on principle component analysis", *International Journal of Machine Learning and Cybernetics*, December 2012, Volume 3, Issue 4, pp 259-267.
- [9] Melanie Mitchell, "An introduction to Genetic Algorithms"
- [10] Randy L. Haupt and Sue Ellen Haupt, "Practical Genetic Algorithms" second edition, John Wiley and Sons, Inc.
- [11] I. Daubechies, "Ten Lectures on Wavelets", *CBMS-NSF series in Applied Mathematics*, Vol. 61, SIAM Press, Philadelphia, 1992.
- [12] T. Kathirvalavakumar, E. Ponmalar, "Self organizing map and wavelet based image compression" *International Journal of Machine Learning and Cybernetics August 2013*, Volume 4, Issue 4, pp 319-326.
- [13] C Nastar, "The image shape spectrum for image retrieval", Technical report, No. 3206, INRIA, June 1997.
- [14] Vapnik, V. N. (1998). *Statistical learning theory*. New York: John Wiley & Sons.
- [15] Abdallah Bashir Musa, "Comparative study on classification performance between support vector machine and logistic regression" *International Journal of Machine Learning and Cybernetics*, December 2012, Volume 3, Issue 4, pp 259-267.
- [16] ORL Database of Faces, (1994). <http://www.cl.cam.ac.uk/research/dtg/attarchive/facedatabase.html>.
- [17] 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.

- [18] Ergun Gumus , Niyazi Kilic ,Ahmet Sertbas , Osman N. Ucan,'Evaluation of face recognition techniques using PCA, wavelets and SVM' Expert Systems with Applications 37 (2010) 6404–640.



Manisha Satone received her ME in Electronics from RTMU, Nanded in 1998. She is working as an Associate Professor at KKWIEER, Nashik, and is now undertaking a doctorate course. Her research interests are in image and Video Processing, Face Recognition, Microprocessor and Computer Network.



Gajanan Kharate received his Ph.D. in Electronics and Telecommunication Engineering from Pune University in 2007. He is a Principal at Matoshri College of Engineering and Research, Nashik and also serves as a Dean of Pune University. His research interests are in the areas of Digital Signal Processing, Communication Engineering, Image Processing and VLSI Design.