Face Recognition Using Particle Swarm Optimization-Based Selected Features

Rabab M. Ramadan, Rehab F. Abdel - Kader*
Electrical Engineering Department, Faculty of Engineering - Port-Said, Suez Canal University, Port Fouad 42523, Port-Said, Egypt
E-mail: r.abdelkader@scuegypt.edu.eg, rabab.ramadan58@gmail.com

Abstract

Feature selection (FS) is a global optimization problem in machine learning, which reduces the number of features, removes irrelevant, noisy and redundant data, and results in acceptable recognition accuracy. It is the most important step that affects the performance of a pattern recognition system. This paper presents a novel feature selection algorithm based on particle swarm optimization (PSO). PSO is a computational paradigm based on the idea of collaborative behavior inspired by the social behavior of bird flocking or fish schooling. The algorithm is applied to coefficients extracted by two feature extraction techniques: the discrete cosine transform (DCT) and the discrete wavelet transform (DWT). The proposed PSO-based feature selection algorithm is utilized to search the feature space for the optimal feature subset where features are carefully selected according to a well defined discrimination criterion. Evolution is driven by a fitness function defined in terms of maximizing the class separation (scatter index). The classifier performance and the length of selected feature vector are considered for performance evaluation using the ORL face database. Experimental results show that the PSO-based feature selection algorithm was found to generate excellent recognition results with the minimal set of selected features.

Keywords: *Discrete Cosine Transform, Discrete Wavelet Transform, Face Recognition, Feature Selection, Genetic Algorithm, Particle Swarm Optimization.*

1. Introduction

Face recognition (FR) has emerged as one of the most extensively studied research topics that spans multiple disciplines such as pattern recognition, signal processing and computer vision. This is due to its numerous important applications in identity authentication, security access control, intelligent human-computer interaction, and automatic indexing of image and video databases. Many approaches to face recognitions have been developed; an excellent survey paper on the different face recognition techniques can be found in [1].

The success of any FR methodology depends heavily on the particular choice of the features used by the pattern classifier. It is known that a good feature extractor for a face

* Corresponding Author

recognition system is claimed to select as more as possible the best discriminant features which are not sensitive to arbitrary environmental variations such as variations in pose, scale, illumination, and facial expressions. Feature extraction algorithms mainly fall into two categories: geometrical features extraction and, statistical (algebraic) features extraction [2], [3], [4], [5], [6], [7], [8]. The geometrical approach , represent the face in terms of structural measurements and distinctive facial features that include distances and angles between the most characteristic face components such as eyes, nose, mouth or facial templates such as nose length and width, mouth position, and chin type. These features are used to recognize an unknown face by matching it to the nearest neighbor in the stored database. Statistical features extraction is usually driven by algebraic methods such as principal component analysis (PCA), and independent component analysis (ICA) [6]. These methods find a mapping between the original feature spaces to a lower dimensional feature space. The shortage of PCA is that it treats inner-class and out-class equally [3], [4], [5] and therefore it is sensitive to light and changes of expressions. LDA has higher performance than PCA in face recognition but the traditional LDA can not provide reliable and robust solution since their separable criterion is not relevant to classification precision. Alternative algebraic methods are based on transforms such as downsampling, Fourier transform (FT), discrete cosine transform (DCT), and the discrete wavelet transform (DWT). Transformation based feature extraction methods such as the DCT and DWT were found to generate good FR accuracies with very low computational cost [8]. DCT is one of the approaches used in image compressing which is also used to extract features [9], [10]. Wavelet analysis has both a good qualities in time domain and frequency domain which is an ideal tool in analyzing unsteady signals. The DCT and the DWT feature extraction methods are explained in detail in Section 2.

Feature extraction methods commonly represent the face images with a large set of features in which features do not contribute equally to the face recognition task. Feature selection (FS) in pattern recognition involves the derivation of the feature subset from the raw input data to reduce the amount of data used for classification and simultaneously provide enhanced discriminatory power. The selection of an appropriate set of features often exploits the design criteria such as redundancy minimization and decorrelation, and minimization of the reconstruction error. For many pattern classification problems, a higher number of features used do not necessarily translate into higher recognition rate [11]. In some cases the performance of algorithms devoted to speed and predictive accuracy of the data characterization can even decrease. Therefore, feature selection can serve as a pre-processing tool of great importance before solving the classification problems. The purpose of the feature selection is to reduce the maximum number of irrelevant features while maintaining acceptable classification accuracy. Feature selection is of considerable importance in pattern classification, data analysis, multimedia information retrieval, biometrics, remote sensing, computer vision, medical data processing, machine learning, and data mining applications.

The feature selection seeks for the optimal set of d features out of m [11], [12], [13]. One possible approach would be to do an exhaustive search among all possible feature subsets $\binom{m}{d}$ and choose the best one according to the optimization criterion at hand. However, such an approach is computationally very expensive. Several methods have been previously used to perform feature selection on training and testing data, branch and bound algorithms [14], sequential search algorithms [15], mutual information [16], tabu search [17] and greedy algorithms [12]. In an attempt to avoid the prohibitive complexity FS algorithms usually involve heuristic or random search strategies. Among the various methods proposed for FS, population-based optimization algorithms such as Genetic Algorithm (GA)-based method [7], [18], [19] and Ant Colony Optimization (ACO)-based method have attracted a lot of attention [20]. These methods attempt to achieve better solutions by using knowledge from previous iterations with no prior knowledge of features.

In this paper, a face recognition algorithm using a PSO-based feature selection approach is presented. The algorithm utilizes a novel approach that employs the binary PSO algorithm to effectively explore the solution space for the optimal feature subset. The selection algorithm is applied to feature vectors extracted using the DCT and the DWT. The search heuristics in PSO is iteratively adjusted guided by a fitness function defined in terms of maximizing class separation. The proposed algorithm was found to generate excellent recognition results with less selected features.

The main contribution of this work is:

- Formulation of a new feature selection algorithm for face recognition based on the binary PSO algorithm. The algorithm is applied to DCT and DWT feature vectors and is used to search for the optimal feature subset to increase recognition rate and class separation.
- Evaluation of the proposed algorithm using the ORL face database and comparing its performance with a GA- based feature selection algorithm and various FR algorithms found in the literature.

The rest of this paper is organized as follows. The DCT and the DWT feature extraction techniques are described in Section 2. An overview of Particle Swarm Optimization (PSO) is presented in Section 3. In Section 4 we explain the proposed PSO- based feature selection algorithm. Finally, Sections 5 and 6 attain the experimental results and conclusion**.**

2. Feature Extraction

The first step in any face recognition system is the extraction of the feature matrix. A typical feature extraction algorithm tends to build a computational model through some linear or nonlinear transform of the data so that the extracted feature is as representative as possible. In this paper DCT and DWT were used for feature extraction as explained in the following Sections.

2.1. Discrete Cosine Transform (DCT)

DCT has emerged as a popular transformation technique widely used in signal and image processing. This is due to its strong “energy compaction” property: most of the signal information tends to be concentrated in a few low-frequency components of the DCT. The use of DCT for feature extraction in FR has been described by several research groups [8], [9], [10] [21], [22], [23] and [24]. DCT was found to be an effective method that yields high recognition rates with low computational complexity. DCT exploits inter-pixel redundancies to render excellent decorrelation for most natural images. After decorrelation each transform coefficient can be encoded independently without losing compression efficiency. The DCT helps separate the image into parts (or spectral sub-bands) of differing importance (with respect to the image's visual quality). DCT transforms the input into a linear combination of weighted basis functions. These basis functions are the frequency components of the input data. DCT is similar to the discrete Fourier transform (DFT) in the sense that they transform a signal or image from the spatial domain to the frequency domain, use sinusoidal base functions, and exhibit good decorrelation and energy compaction characteristics. The major difference is that the DCT transform uses simple cosine-based basis functions whereas the DFT is a complex transform and therefore stipulates that both image magnitude and phase information be encoded. In addition, studies have shown that DCT provides better energy compaction than DFT for most natural images.

The general equation for the DCT of an NxM image $f(x, y)$ is defined by the following equation:

$$F(u,v) = \alpha(u)\alpha(v)\sum_{x=0}^{N-1}\sum_{y=0}^{M-1} .\cos[\frac{\pi.u}{2.N}(2x+1)]\cos[\frac{\pi.u}{2.M}(2y+1)]f(x,y) \quad (1)$$

Where $f(x, y)$ is the intensity of the pixel in row *x* and column *y*; *u = 0,1, …., N-1* and *v=0,1, …., M-1* and the functions $\alpha(u)$, $\alpha(v)$ are defined as

$$\alpha(u), \alpha(v) = \begin{cases} \sqrt{\frac{1}{N}} & for \quad u,v = 0 \\ \sqrt{\frac{2}{N}} & for \quad u,v \neq 0 \end{cases} \quad (2)$$

For most images, much of the signal energy lies at low frequencies (corresponding to large DCT coefficient magnitudes); these are relocated to the upper-left corner of the DCT array. Conversely, the lower-right values of the DCT array represent higher frequencies, and turn out to be small enough to be truncated or removed with little visible distortion, especially as *u* and *v* approach the sub-image width and height, respectively. This means that the DCT is an effective tool that can pack the most effective features of the input image into the fewest coefficients.

The original face image can be roughly reconstructed only by few DCT coefficients. This makes the choice of the number of DCT coefficient initially used in the face recognition system very critical. The effect of the number of DCT coefficients used as features for face recognition is examined in Section *5*. This includes the effect of the number of coefficients on the quality of the reconstructed image and the recognition rate. The study is extended by examining the performance of the dynamically generated feature subset generated by the PSO feature selection algorithm.

2.2. Discrete Wavelet Transform (DWT)

Wavelets have many advantages over other mathematical transforms such as the DFT or DCT. Functions with discontinuities and functions with sharp spikes usually take substantially fewer wavelet basis functions than sine-cosine functions to achieve a comparable approximation. Wavelets have been successfully used in image processing since 1985 [8], [22], [25], and [26]. Its ability to provide spatial and frequency representations of the image simultaneously motivates its use for feature extraction. The decomposition of the input data into several layers of division in space and frequency and allows us to isolate the frequency components introduced by intrinsic deformations due to expression or extrinsic factors (like illumination) into certain sub-bands. Wavelet-based methods prune away these variable sub-bands, and focus on the space/frequency sub-bands that contain the most relevant information to better represent the data and aid in the classification between different images. There exists a large selection of wavelet families depending on the choice of the mother wavelet. In this paper FR using the DWT is based on the facial features extracted from a Haar Wavelet

Transform [8, 22]. The Haar wavelet transform is a widely used technique that has an established name as a simple and powerful technique for the multi-resolution decomposition of time series. Earlier studies concluded that information in low spatial frequency bands play a dominant role in face recognition. In 1986, Sergent [26] shows that the low frequency band and high frequency band play different roles. The low frequency components contribute to the global description, while the high frequency components contribute to the finer details required in the identification task. Sergent has also demonstrated that as human face is a non-rigid object, it has abundant facial expressions; and expressions influence local spatial components of face.

The Haar wavelet transform has been proven effective for image analysis and feature extraction. It represents a signal by localizing it in both time and frequency domains .Wavelets can be used to improve the image registration accuracy by considering both spatial and spectral information and by providing multi-resolution representation to avoid loosing any global or local information. Additional advantages of using the wavelet-decomposed images include bringing data with different spatial resolution to a common resolution using the low frequency sub-bands while providing access to edge features using the high frequency sub-bands.

As shown in Figure 1 at each level of the wavelet decomposition, four new images are created from the original *N x N*-pixel image. The size of these new images is reduced to ¼ of the original size, i.e., the new size is *N/2 x N/2.* The new images are named according to the filter (low-pass or high-pass), which is applied to the original image in horizontal and vertical directions. For example, the LH image is a result of applying the low-pass filter in horizontal direction and high-pass filter in vertical direction. Thus, the four images produced from each decomposition level are LL, LH, HL, and HH. The LL image is considered a reduced version of the original as it retains most details. The LH image contains horizontal edge features, while the HL contains vertical edge features. The HH contains high frequency information only and is typically noisy and is, therefore, not useful for the registration. In wavelet decomposition, only the LL image is used to produce the next level of decomposition.

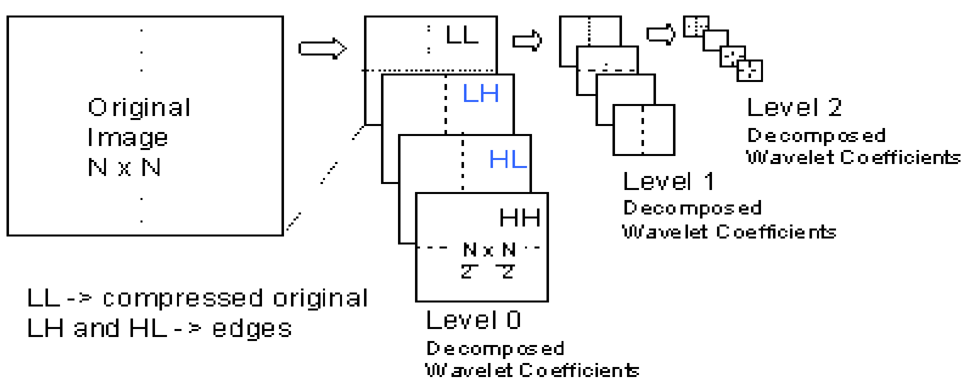
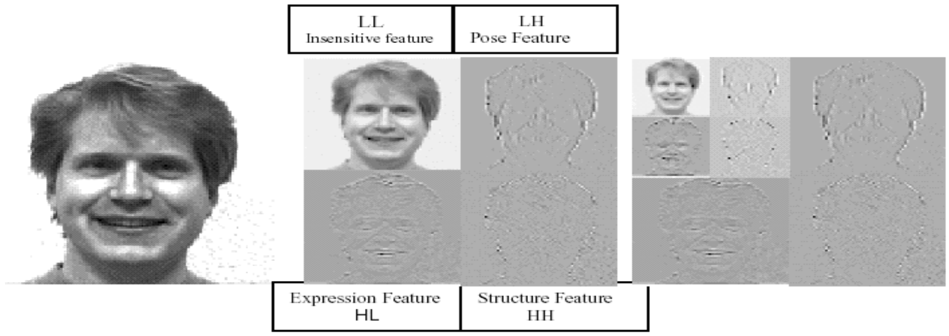


Figure 1. A 3-level wavelet decomposition of an N x N-pixel image

Figure 2 shows the decomposition process by applying the 2D Wavelet Transform on a face image. The original image (shown in Figure (2a)) is decomposed into four sub band images (shown in Figure (2b)) similarly, 2 levels of the Wavelet decomposition as shown Figure (2c) can be obtained by applying the wavelet transform on the low frequency band sequentially.



(a) (b) (c)

Figure 2. 2D Wavelet decomposition of a face image. (a)The Original Image (b) 1-Level Wavelet Decomposition (c) 2-Level Wavelet decomposition

In Figure (2b), the sub band LL corresponds to the low frequency components in both vertical and horizontal directions of the original image. Therefore, it is the low frequency sub band of the original image. The sub band LH corresponds to the low frequency component in the horizontal direction and high frequency components in vertical direction. Therefore it holds the vertical edge details. Similar interpretation is made on the sub bands HL and HH.

As the change of facial expressions mainly varies in eyes, mouth and other face muscles, from the technical point of view, it involves mainly changes of edges. Let's take Figure (2b) as an example, the horizontal features of eyes and mouth are clearer than its vertical features, the sub band HL can therefore depict major facial expression features. The sub band LH, the vertical features of outline and nose are clearer than its horizontal features, depicts face pose features. The sub band HH is therefore the most important for rigid object recognition because it depicts the structure feature of the object. But human faces indeed are non-rigid objects, the sub band HH is the unstable band in all sub bands because it is easily disturbed by noises, expressions and poses. Therefore, if wavelet transform is applied to decompose face images, the sub band LL will be the most stable sub band.

3. Particle Swarm Optimization (PSO)

PSO proposed by Dr. Eberhart and Dr. Kennedy in 1995 is a computational paradigm based on the idea of collaborative behavior and swarming in biological populations inspired by the social behavior of bird flocking or fish schooling [27], [28], [29], and [30]. Recently PSO has been applied as an effective optimizer in many domains such as training artificial neural networks, linear constrained function optimization, wireless network optimization, data clustering, and many other areas where GA can be applied [29].

Computation in PSO is based on a population (swarm) of processing elements called particles in which each particle represent a candidate solution. PSO shares many similarities with evolutionary computation techniques such as GA's. The system is initialized with a population of random solutions and searches for optima by updating generations. The search process utilizes a combination of deterministic and probabilistic rules that depend on information sharing among their population members to enhance their search processes. However, unlike GA's, PSO has no evolution operators such as crossover and mutation. Each particle in the search space evolves its candidate solution over time, making use of its individual memory and knowledge gained by the swarm as a whole. Compared with GAs, the information sharing mechanism in PSO is considerably different. In GAs, chromosomes share information with each other, so the whole population moves like one group towards an optimal area. In PSO, the global best particle found among the swarm is the only information shared among particles. It is a one-way information sharing mechanism. Computation time in PSO is significantly less than in GAs because all the particles in PSO tend to converge to the best solution quickly [29].

3.1. PSO Algorithm

When PSO is used to solve an optimization problem, a swarm of computational elements, called particles, is used to explore the solution space for an optimum solution. Each particle represents a candidate solution and is identified with specific coordinates in the D-dimensional search space. The position of the i -th particle is represented as $X_i = (x_{i1}, x_{i2}, \ldots\ldots, x_{iD})$. The velocity of a particle (rate of the position change between the current position and the next) is denoted as $V_i = (v_{i1}, v_{i2}, \ldots\ldots\ldots, v_{iD})$. The fitness function is evaluated for each particle in the swarm and is compared to the fitness of the best previous result for that particle and to the fitness of the best particle among all particles in the swarm. After finding the two best values, the particles evolve by updating their velocities and positions according to the following equations:

$$V_i^{t+1} = \omega * V_i^t + c_1 * rand_1 * (p_{i\_best} - X_i^t) + c_2 * rand_2 * (g_{best} - X_i^t) \tag{3}$$

$$X_i^{t+1} = X_i^t + V_i^{t+1} \tag{4}$$

Where $i = (1, 2, \ldots\ldots N)$ and N is the size of the swarm; p_{i_best} is the particle best reached solution and g_{best} is the global best solution in the swarm. c_1 and c_2 are cognitive and social parameters that are bounded between 0 and 2. $rand_1$ and $rand_2$ are two random numbers, with uniform distribution $U(0,1)$. $-V_{max} \leq V_i^{t+1} \leq V_{max}$ (V_{max} is the maximum velocity). In equation (3) the first component represents the inertia of pervious velocity. The inertia weight ω, is a factor used to control the balance of the search algorithm between exploration and exploitation; the second component is the "cognitive" component representing the private experience of the particle itself; the third component is the "social" component, representing the cooperation among the particles. The recursive steps will go on until we reach the termination condition (maximum number of iterations K). The pseudo code of the PSO algorithm is shown in Figure 3.

```
Initialize parameters
Initialize population
while (number of generations, or the stopping criterion is not met) {
   for (i = 1 to number of particles N)  {
       if the fitness of X_i^t is greater than the fitness of p_i_best
            then update p_i_best = X_i^t
       if the fitness of X_i^t is greater than that of g_best then
            then update g_best = X_i^t
     Update velocity vector
     Update particle position
     Next particle
    }
  Next generation
}
```

Figure 3. Pseudo code of the PSO algorithm

3.2. Binary PSO and Feature Selection

A binary PSO algorithm has been developed in [30]. In the binary version, the particle position is coded as a binary string that imitates the chromosome in a genetic algorithm. The particle velocity function is used as the probability distribution for the position equation. That is, the particle position in a dimension is randomly generated using that distribution.
The equation that updates the particle position becomes the following:

If $rand_3 < \frac{1}{1+e^{-v_i^{t+1}}}$ then $X_i^{t+1} = 1$; else $X_i^{t+1} = 0$ (5)

A bit value of {1} in any dimension in the position vector indicates that this feature is selected as a required feature for the next generation, whereas a bit value of {0} indicates that this feature is not selected as a required feature for the next generation.

4. PSO-Based Feature Selection

The task for the binary PSO algorithm is to search for the most representative feature subset through the extracted DCT or DWT feature space. Each particle in the algorithm represents a possible candidate solution (feature subset). Evolution is driven by a fitness function defined in terms of class separation (scatter index) which gives an indication of the expected fitness on future trials.

4.1 Chromosome Representation

The initial coding for each particle is randomly produced where each particle is coded to imitate a chromosome in a genetic algorithm; each particle was coded to a binary alphabetic string $P = F_1F_2 \ldots F_n$, $n = 1, 2, \ldots, m$; where *m* is the length of the feature vector extracted by the DCT or the DWT. Each gene in the *m*-length chromosome represents the feature selection, "*1*" denotes that the corresponding feature is selected, otherwise denotes rejection. The binary PSO algorithm is used to search the 2^m genospace for the optimal feature subset where optimality is defined with respect to class separation. For example, when a 10-dimensional data set ($n=10$) $P = F_1\, F_2\, F_3\, F_4\, F_5\, F_6\, F_7\, F_8\, F_9\, F_{10}$ is analyzed using binary PSO to select features, we can select any subset of features smaller than n. i.e. PSO can chose a random 6 features, $F_1\ F_2\ F_4\ F_6\ F_8\ F_9$ by setting bits *1, 2, 4, 6, 8,* and *9* in the particle chromosome. For each particle, the effectiveness of the selected feature subset in retaining the maximum accuracy in representing the original feature set is evaluated based on its fitness value.

4.2 Fitness Function

The m-genes in the particle represent the parameters to be iteratively evolved by PSO. In each generation, each particle (or individual) is evaluated, and a value of *goodness* or *fitness* is returned by a fitness function. This evolution is driven by the fitness function *F* that evaluates the quality of evolved particles in terms of their ability to maximize the class separation term indicated by the scatter index among the different classes [3]. Let $w_1, w_2, \ldots, w_L$ and $N_1, N_2, \ldots, N_L$ denote the classes and number of images within each class, respectively. Let $M_1, M_2, ..., M_L$ *and* M_o be the means of corresponding classes and the grand mean in the feature space, M_i can be calculated as:

$$M_i = \frac{1}{N_i}\sum_{j=1}^{N_i} W_j^{(i)}, \quad i = 1,2,..., L. \tag{6}$$

Where $W_j^{(i)}$, $j = 1,2,..., N_i$, represents the sample images from class w_i and the grand mean M_o is:

$$M_0 = \frac{1}{N}\sum_{i=1}^{L} N_i M_i \tag{7}$$

Where n is the total number of images for all the classes. Thus, the between class scatter fitness function F is computed as follows:

$$F = \sqrt{\sum_{i=1}^{L} (M_i - M_o)^t (M_i - M_o)} \tag{8}$$

4.3 PSO-Based Feature Selection Algorithm

An overview of the proposed PSO-based feature selection algorithm is shown in Figure 4.

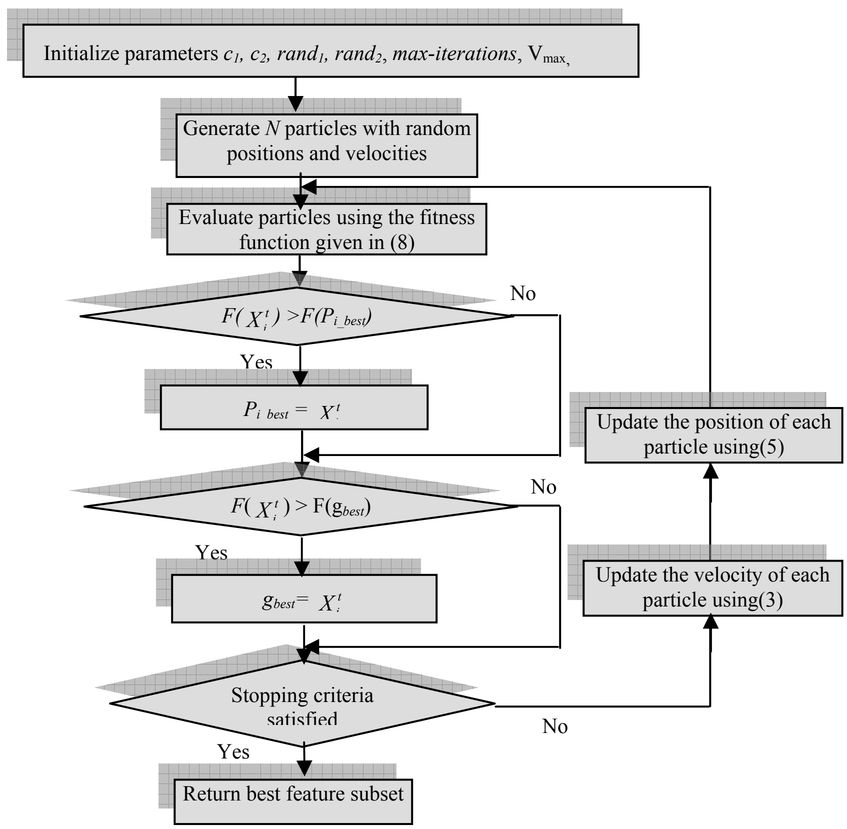


Figure 4. The PSO-based feature selection algorithm

4.4. Classifier

A typical and popular Euclidean distance is employed to measure the similarity between the test vector and the reference vectors in the gallery. Euclidean distance is defined as the straight-line distance between two points. For N-dimensional space, the Euclidean distance between two any points' p_i and q_i is given by:

$$D = \sqrt{\sum_{i=1}^{N} (p_i - q_i)^2} \tag{9}$$

Where p_i (or q_i) is the coordinate of p (or q) in dimension i.

In the application of this approach for face recognition, distances in the feature space from a query image to every image in the database are calculated. The index of the image which has the smallest distance with the image under test is considered to be the required index.

5. Experimental Results

The block diagram of the proposed FR system is shown in Figure 5. The block diagram shows the various steps of processing an input image in the training and recognition stages.

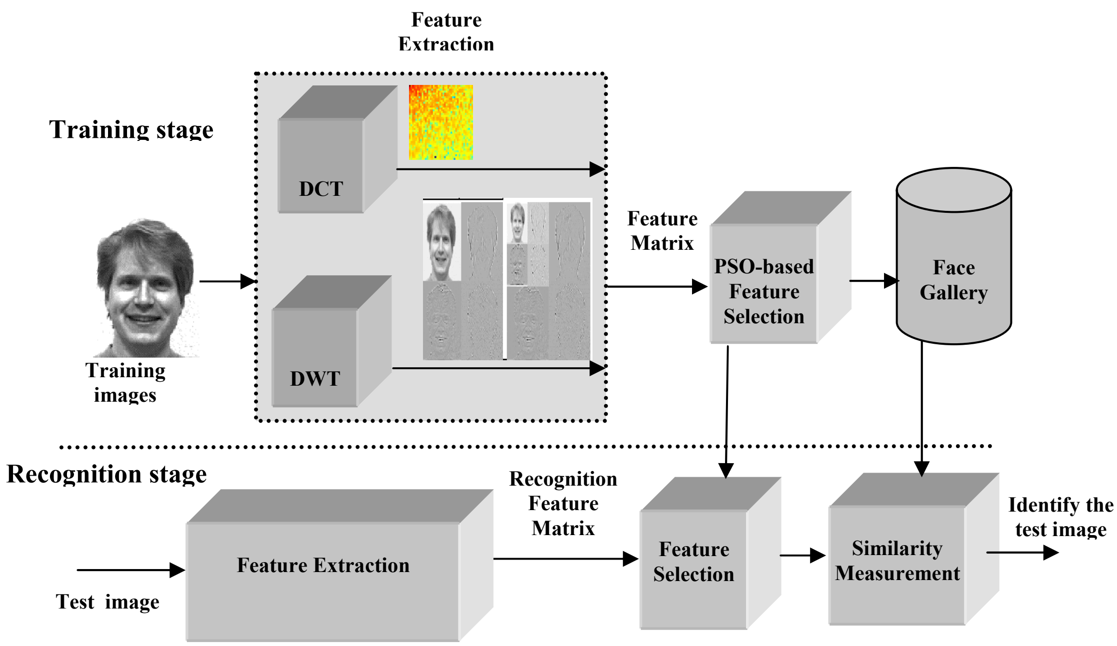


Figure 5. Block Diagram of the proposed Face recognition system.

The performance of the proposed feature selection algorithm is evaluated using the standard Cambridge ORL gray-scale face database. The ORL database of faces contains a set of face images taken between April 1992 and April 1994 at the AT&T Laboratories (by the Oliver Research Laboratory in Cambridge, UK) [20] and [21]. The database is composed of 400 images corresponding to 40 distinct persons. The original size of each image is 92x112 pixels, with 256 grey levels per pixel. Each subject has 10 different images taken in various sessions varying the lighting, facial expressions (open/ closed eyes, smiling/ not smiling) and facial details (glasses/ no glasses). All the images were taken against a dark homogeneous background with the subjects in an upright, frontal position (with tolerance for some side movement). Four images per person were used in the training set and the remaining six images were used for testing. Figure 6 shows sample images from the ORL face database.

Figure 6. Sample images from the ORL faced database.

In the experiments carried out in this Section we compare the performance of the proposed PSO-based feature selection algorithm with the performance of a GA-based feature selection algorithm. The parameters used for the binary PSO and the GA algorithms are given in Table 1 and Table 2 respectively.

TABLE 1. PSO PARAMETER SETTING

Swarm size N	30
Cognitive parameters c_1	2
Social parameter c_2	2
Inertia weight ω	0.6
Number of Iterations	100

TABLE 2. GA PARAMETER SETTING

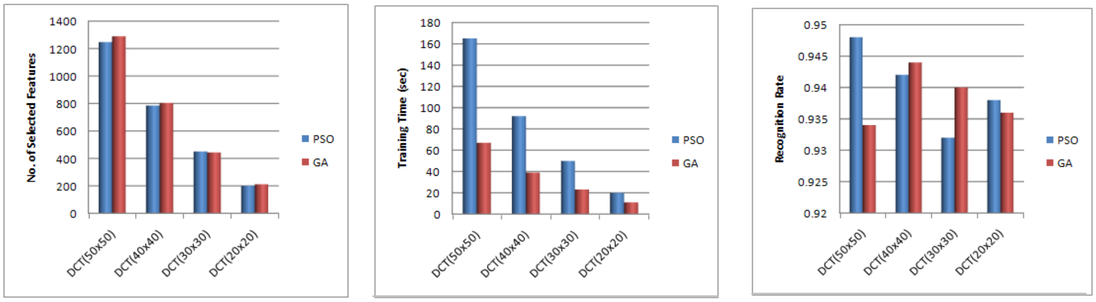
The population size	30
Crossover Probability (P_C)	0.5
Mutation Probability (P_M)	1
Number of Iterations	100

For each problem instance, 5 replications are conducted. The average recognition rate is measured together with the CPU training time and the average number of selected features for each experiment.

5.1. Experiment 1

In this experiment we test the PSO-based feature selection algorithm with feature vectors based on various sizes of DCT coefficients. The 2-dimentional DCT is applied to the input image and only a subset of the DCT coefficients corresponding to the upper left corner of the DCT array is retained. Subset sizes of 50x50, 40x40, 30x30 and 20x20 of the original 92x112 DCT array are used in this experiment as input to the subsequent feature selection phase. Figure 7 shows the No. of selected features, training time, and recognition rates for different feature vector dimensions using the PSO and GA based feature selection algorithms.

The best average recognition rate of 94.8% is achieved using the DCT (50x50) feature vector and the PSO-based feature selection algorithm. In this instance, the selection algorithm reduces the size of the original feature vector by nearly 50%. In general the PSO and GA selection algorithms have comparable performance in terms of recognition rates but in all test cases the number of selected features is smaller when using the PSO selection algorithm. On the other hand, in terms of computational time, GA-based selection algorithm takes less training time than the PSO-based selection algorithm in all tested instances. Which indicates that PSO is computationally expensive than GA but the effectiveness' of PSO in finding the optimal feature subset compared to GA compensates its computational inefficiency.



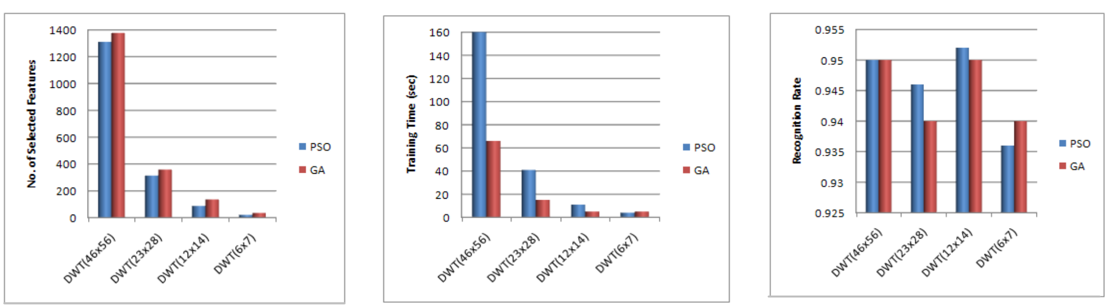
(a) (b) (c)

Figure 7. Recognition results for different DCT-feature based vectors and Feature selection Algorithms. (a) No. of Selected Features (b) Training Time (c) Recognition Rate

5.2. Experiment 2

In this experiment the DWT coefficient features have been extracted from each face image. The 2-dimentional Haar wavelet transform is applied to the input image reducing its size to 1/4 of its original size. 4-level wavelet decomposition is performed and the approximation of the input image at each decomposition level is used as a feature vector. The dimensions of the feature vectors are 46x56, 23x28, 12x14 and 6x8 corresponding to level-0, level-1, level-2 and level-3 wavelet decompositions respectively.

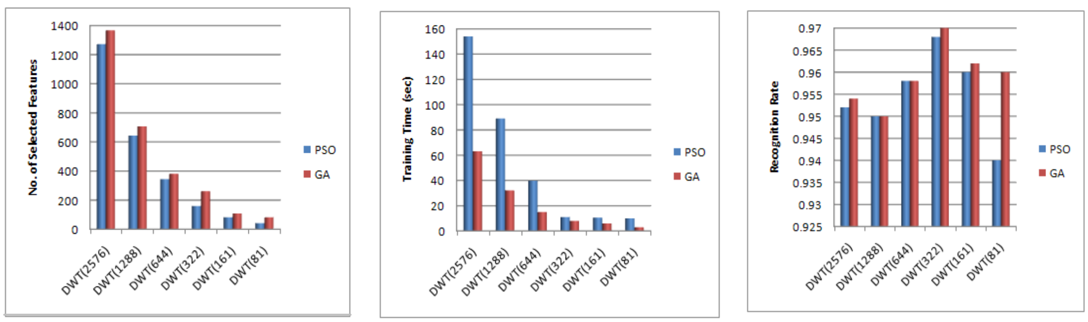
Figure 8 shows the No. of selected features, training time, and recognition rates for different feature vector dimensions using the PSO and GA based feature selection algorithms. The best average recognition rate of 95.2% is achieved using the DWT (12x14) feature vector and the PSO-based feature selection algorithm using only 88 selected features and with approximately 35% less selected features than GA. The PSO and GA selection algorithms have comparable performance in most tested instances but with less selected features using PSO.



(a) (b) (c)

Figure 8. Recognition results for different DWT-feature based vectors and Feature selection Algorithms. (a) No. of Selected Features (b) Training Time (c) Recognition Rate

Another DWT-based feature extraction approach was implemented and tested. The 2-dimentional input image is converted to a 1-dimentional array using raster scan. This is achieved by processing the image row by row concatenating the consecutive rows into an array of size 10304. Raster scanning preserves horizontal pixel correlations well but vertical correlations are lost [31]. The DWT is then applied to the 1-dimentinal image array and used as feature vector for the feature selection phase. 6-level wavelet decomposition is performed and the approximation of the input image at each decomposition level is used as a feature vector. The dimensions of the feature vectors are 2576, 1288, 644, 322, 161, and 81 corresponding to level-1, level-2, level-3, level-4, level-5 and level-6 wavelet decompositions respectively. Figure 9 shows the No. of selected features, training time, and recognition rates for different feature vector dimensions using the PSO and GA based feature selection algorithms.



(a) (b) (c)

Figure 9. Recognition results for different DWT-feature based vectors applied to 1-Dimentional Image Array and Feature Selection Algorithms. (a) No. of Selected Features (b) Training Time (c) Recognition Rate

The best average recognition rate of 97% is achieved using the DWT(322) feature vector and the GA-based feature selection algorithm using 262 selected features. This is compared an average recognition rate of 96.8% with 159 PSO-based selected features for the same test instance. Experimental results indicate that the recognition accuracy based on features extracted using the DWT applied to the 1-dimentional raster scan of the input image outperforms that of the DCT and the DWT of the 2-dimantional input image.

In Table 3, the performance of the proposed algorithm in terms of its recognition rate is compared to various FR algorithms found in the literature using the ORL database [32]. Table 3 indicates the superiority of the proposed algorithm utilizing the DWT feature extraction and PSO feature selection. As far as feature selection is concerned the algorithm selects the optimal number of elements in the feature vector which has a great influence on the training and recognition times of the algorithm.

6. Conclusion

In this paper, a novel PSO-based feature selection algorithm for FR is proposed. The algorithm is applied to feature vectors extracted by two feature extraction techniques: the DCT and the DWT. The algorithm is utilized to search the feature space for the optimal feature subset. Evolution is driven by a fitness function defined in terms of class separation.

The classifier performance and the length of selected feature vector were considered for performance evaluation using the ORL face database. Experimental results show the superiority of the PSO-based feature selection algorithm in generating excellent recognition accuracy with the minimal set of selected features. The performance of the proposed algorithm is compared to the performance of a GA-based feature selection algorithm and was found to yield comparable recognition results with less number of selected features.

Table 3. Comparison of recognition rates for various FR algorithms

METHOD	*RECOGNITION RATE*	*TEST CONDITIONS*
Hybrid NN: SOM+a convolution NN	96.2%	DB contained 400 images of 40 individuals. The classification time less than .5 second for recognizing one facial image, but training time is 4 hour
Hidden Markov model (HMMs)	87%	
SVM with a binary tree	91.21% for SVM and 84.86% for Nearest Center Classification (NCC)	They compare the SVMs with standard eigenfaces approach using the NCC
Eigenface	90%	
2D-HMM	95%	An average processing time of .22 second for face pattern with 40 classes.
DCT+PSO FS	*94.7*	Four images per person were used in the training set and the remaining six images were used for testing The training time is less than 3 minutes for all experiments. The average recognition time for recognizing an input image is *0.05 sec*
DWT +PSO FS	*96.8*	

7. References

[1] W. Zhao, R. Chellappa, P. J. Phillips, and A. Rosenfeld, "Face Recognition: A Literature Survey," *ACM Computing Surveys,* vol. 35, no. 4, pp. 399-458, 2003.

[2]R. Brunelli and T. Poggio, "Face Recognition: Features versus Templates," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 15, no. 10, pp. 1042-1052, 1993.

[3] C. Liu and H. Wechsler, "Evolutionary Pursuit and Its Application to Face Recognition," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 22, no. 6, pp. 570-582, 2000.

[4] M. A. Turk and A. P. Pentland, "Face Recognition using Eigenfaces," *Proc. of IEEE Conference on Computer Vision and Pattern Recognition*, pp. 586-591, June 1991.

[5] L. Du, Z. Jia, and L. Xue, "Human Face Recognition Based on Principal Component Analysis and Particle Swarm Optimization -BP Neural Network," *Proc 3rd Conference. on Natural Computation (ICNC 2007),*vol. 3, pp. 287-291, August 2007.

[6] X. Yi-qiong, L. Bi-cheng and W. Bo, "Face Recognition by Fast Independent Component Analysis and Genetic Algorithm," *Proc. of the 4th International Conference on Computer and Information Technology (CIT'04)*, pp. 194-198, Sept. 2004.

[7] X. Fan and B. Verma, "Face recognition: a new feature selection and classification technique," *Proc. 7th Asia-Pacific Conference on Complex Systems*, December 2004.

[8] A. S. Samra, S. E. Gad Allah, R. M. Ibrahim, "Face Recognition Using Wavelet Transform, Fast Fourier Transform and Discrete Cosine Transform," *Proc.* 46th *IEEE International Midwest Symp. Circuits and Systems (MWSCAS'03*), vol. 1, pp. 272- 275, 2003.

[9] C. Podilchuk and X. Zhang, "Face Recognition Using DCT-Based Feature Vectors," *Proc. IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP'96),* vol. 4, pp. 2144-2147, May 1996.
[10] Z. Yankun and L. Chongqing, "Efficient Face Recognition Method based on DCT and LDA," *Journal of Systems Engineering and Electronics*, vol. 15, no. 2, pp. 211-216, 2004.
[11] C.-J. Tu, L.-Y. Chuang, *J.-Y. Chang,* and C.-H. Yang, "Feature Selection using PSO-SVM," *International Journal of Computer Science* (*IAENG)*, vol. 33, no. 1, IJCS_33_1_18.
[12] E. Kokiopoulou and P. Frossard, "Classification-Specific Feature Sampling for Face Recognition," *Proc. IEEE 8th Workshop on Multimedia Signal Processing*, pp. 20-23, 2006
[13] A. Y. Yang, J. Wright,Y. Ma, and S. S. Sastry, " Feature Selection in Face Recognition: A Sparse Representation Perspective," submitted for publication, 2007.
[14] P. M. Narendra and K. Fukunage, "A Branch and Bound Algorithm for Feature Subset Selection," *IEEE Trans. Computers*, vol. 6, no. 9, pp.917-922, Sept. 1977.
[15] P. Pudil, J. Novovicova, and J. Kittler, "Floating Search Methods in Feature Selection," *Pattern Recognition Letters*, vol. 15, pp. 1119-1125, 1994.
[16] B. Roberto, "Using Mutual Information for Selecting Features in Supervised Neural Net Learning," *IEEE Trans. Neural Networks*, vol. 5, no. 4, pp. 537-550, 1994.
[17] H. Zhang and G. Sun, "Feature Selection Using Tabu Search Method," *Pattern Recognition Letters*, vol. 35, pp. 701-711, 2002.
[18] D.-S. Kim, I.-J. Jeon, S.-Y. Lee, P.-K. Rhee, and D.-J. Chung, "Embedded Face Recognition based on Fast Genetic Algorithm for Intelligent Digital Photography," *IEEE Trans. Consumer Electronics*, vol. 52, no. 3, pp. 726-734, August 2006.
[19] M. L. Raymer, W. F. Punch, E. D. Goodman, L.A. Kuhn, and A. K Jain, "Dimensionality Reduction Using Genetic Algorithms," *IEEE Trans. Evolutionary Computation*, vol. 4, no. 2, pp. 164-171, July 2000.
[20] H. R. Kanan, K. Faez, and M. Hosseinzadeh, "Face Recognition System Using Ant Colony Optimization-Based Selected Features," *Proc. IEEE Symp. Computational Intelligence in Security and Defense Applications (CISDA 2007),* pp 57-62, April 2007.
[21] F. M. Matos, L. V. Batista, and J. Poel, "Face Recognition Using DCT Coefficients Selection," *Proc. of the 2008 ACM Symposium on Applied Computing, (SAC'08)*,pp. 1753-1757, March 2008.
[22] M. Yu, G. Yan, and Q.-W. Zhu, "New Face recognition Method Based on DWT/DCT Combined Feature Selection," *Proc. 5th International Conference on Machine Learning and Cybernetics*, pp. 3233-3236, August 2006.
[23] Z. Pan and H. Bolouri, "High Speed Face Recognition Based on Discrete Cosine Transform and Neural Networks," Technical Report, Science and Technology Research Center (STRC), University of Hertfordshire.
[24] Z. M. Hafed and M. D. Levine, "Face Recognition Using Discrete Cosine Transform, " *International Journal of Computer Vision,* vol. 43, no. 3, pp. 167-188. 2001.
[25] D.-Q. Dai and H. Yan, "Wavelets and Face Recognition," in Face Recognition, K. Delac and M. Grgic, Eds. I-Tech, Vienna, Austria, 2007, pp.558.
[26] J. Sergent, "Microgenesis of face perception,". In: H.D. Ellis, M.A. Jeeves, F. Newcombe and A. Young, Editors, *Aspects of Face Processing*, Nijhoff, Dordrecht (1986).
[27] J. Kennedy and R. Eberhart, "Particle swarm optimization," *Proc. IEEE International Conference on Neural Networks*, pp. 1942-1948, 1995.
[28] R. C. Eberhart and J. Kennedy, "A New Optimizer Using Particles Swarm Theory," *Proc. Roc. 6th International Symp. Micro Machine and Human Science*, pp. 39-43, Oct. 1995.
[29] R. C. Eberhart and Y. Shi, "Comparison between Genetic Algorithms and Particle Swarm Optimization, " *Proc. 7th international Conference on Evolutionary Programming*, pp. 611-616, 1998.
[30] J. Kennedy and R. C. Eberhart, "A Discrete Binary Version of the Particle Swarm Algorithm", *Proc. IEEE International Conference on Systems, Man, and Cybernetics*, vol. 5, pp. 4104-4108, Oct. 1997.
[31] J. Modayilt, H. Chengt, and X. Lii, "Experiments in Simple One-Dimensional Lossy Image Compression Schemes," *Proc. 1997 International Conference on Multimedia Computing and Systems (ICMCS '97),* pp. 614-615, 1997.
[32] A. S. Tolba, A.H. El-Baz, and A.A. El-Harby "Face Recognition: A Literature Review," *International Journal of Signal Processing, vol.* 2, no. 2, pp. 88-103. 2006