

Multimodal Biometric Recognition Using Particle Swarm Optimization-Based Selected Features

Sara Motamed*

Artificial Intelligence, Ph.D. Student, Computer and Science, Islamic Azad University, Fuman Branch
Samotamed@yahoo.com

Ali Broumandnia

Artificial Intelligence, Assistant Professor, Computer Department, Islamic Azad University, South Tehran Branch
Broumandnia@azad.ac.ir

Azamossadat Nourbakhsh

Artificial Intelligence, Ph.D. Student, Computer Department, Islamic Azad University, Lahijan Branch
Nourbakhsh@liau.ac.ir

Received: 20/Dec/2012

Accepted: 12/Jun/2013

Abstract

Feature selection is one of the best optimization problems in human recognition, which reduces the number of features, removes noise and redundant data in images, and results in high rate of recognition. This step affects on the performance of a human recognition system. This paper presents a multimodal biometric verification system based on two features of palm and ear which has emerged as one of the most extensively studied research topics that spans multiple disciplines such as pattern recognition, signal processing and computer vision. Also, we present 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. In this method, we used from two Feature selection techniques: the Discrete Cosine Transforms (DCT) and the Discrete Wavelet Transform (DWT). The identification process can be divided into the following phases: capturing the image; pre-processing; extracting and normalizing the palm and ear images; feature extraction; matching and fusion; and finally, a decision based on PSO and GA classifiers. The system was tested on a database of 60 people (240 palm and 180 ear images). 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: Biometric, Genetic Algorithm (GA), Particle Swarm Optimization (PSO), Discrete Cosine Transform (DCT), Discrete Wavelet Transform (DWT).

1. Introduction

It is known that a good feature extractor for a human recognition system is claimed to select as much as possible the best discriminate of features which are not sensitive to arbitrary environmental variations such as variations in pose, scale and illumination. Feature selection algorithms mainly fall into two categories: geometrical features extraction and statistical (algebraic) features extraction [2-8]. Single modal biometric system depends on only one biometric feature of a person. Single modal biometric systems are less accurate and not universally accepted [1]. They are more susceptible to the factors that generate false results like environmental noise, change of biometric features with time and condition, illness or accidents and spoofing [21]. Multi-modal biometric systems [23] are expected to be more reliable due to the presence of multiple pieces of evidence. These systems are also able to meet the

stringent performance requirements imposed by various applications [24]. Multimodal systems address the problem of non-universality: It is possible for a subset of users which do not possess particular biometrics. In such instances, it is useful to acquire multiple biometric traits to verify identity. Multimodal systems also provide anti-spoofing measures by it making difficult for an intruder spoofing multiple biometric traits simultaneously. By asking the user to present a random subset of biometric traits, the system ensures that a *_live_* user is indeed present at the point of acquisition. However, an integration scheme is required to fuse the information presented by the individual modalities. Moreover, multimodal biometric system takes more than one single feature into account [31]. This helps in identifying and verifying the person with more accuracy even if one of the features gives less matching score [32,34,35]. Our multimodal biometric identification system is based on

* Corresponding Author

features extracted from palm and ear images by alternative algebraic methods, which are based on transforms called 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 rate of accuracies with very low computational cost [8]. DCT is one of 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 unsteady signals analyzing. The DCT and the DWT Feature extraction methods are explained in detail in Section 2.

Feature extraction in pattern recognition involves the derivation of feature subset from the raw input data to reduce the amount of data used for classification, and simultaneously provide enhanced discriminatory power. The extraction of an appropriate set of features often exploits the design criteria such as redundancy minimization, and minimizing the reconstruction error. For many pattern classification problems, usage of a higher number of features does not necessarily translate into higher recognition rate [11]. In some cases the performance of algorithms is devoted to speed and predictive accuracy of the data characterization can even decrease. Therefore, feature extraction can serve as a pre-processing tool of great importance before solving the classification problems. The purpose of feature extraction is reducing the maximum number of irrelevant features while maintaining acceptable classification accuracy. Feature extraction is considerably important in pattern classification, data analysis, multimedia information retrieval, biometrics, remote sensing, computer vision, medical data processing, machine learning, and data mining applications. Feature extraction seeks for the optimal set of d features out of m [11-13] one possible approach is an exhaustive search among all possible feature subsets $\binom{m}{d}$ and choosing 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 extraction on training and testing data, branch and bound algorithms [14], sequential search algorithms [15], mutual information [16], tabu search [17] and greedy algorithms [12].

To avoid the prohibitive complexity feature selection algorithms, we usually involve heuristic or random search strategies. Among the various methods proposed for feature extraction, population-based optimization algorithms such as Genetic Algorithm (GA)-based method [7],

[18], [19] and Ant Colony Optimization (ACO)-based method have been 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, palm and ear recognition algorithms 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 and guided by a fitness function definition in terms of maximizing class separation. The proposed algorithm was found to generate excellent recognition results with less selected features. Our paper is divided into 6 sections that are introduced with following sequences: To use feature selection algorithms by palm and ear recognition based on the binary PSO algorithm in Section 1. The DCT and the DWT Feature selection 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

In this section, two methods of feature extraction for building ear and palm features vector are introduced. DCT and DWT were used for feature extraction as explained in the following Sections.

2.1 Discrete Cosine Transform (DCT)

DCT, as a popular transformation technique, has been 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. DCT is 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 separating 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 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 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 [33]. The general equation for the DCT of an $N \times M$ image $f(x,y)$ is defined by the following equation:

$$f(U, V) = \alpha(U) \alpha(V) \sum_{X=0}^{N-1} \sum_{Y=0}^{M-1} \cos\left[\frac{\pi \cdot U}{2 \cdot N} (2X+1)\right] \cos\left[\frac{\pi \cdot V}{2 \cdot M} (2Y+1)\right] f(X, Y) \quad (1)$$

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

$$\alpha(U) \cdot \alpha(V) = \begin{cases} \sqrt{\frac{1}{N}} & \text{For } u, v=0 \\ \sqrt{\frac{2}{N}} & \text{For } u, v \neq 0 \end{cases} \quad (2)$$

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 [33].

The palm and ear images can be roughly reconstructed only by few DCT coefficients. This increases selecting DCT coefficient initially used in the palm and ear recognition system very critical. The effect of the number of DCT coefficients used as features for palm and ear recognition is examined in Section 5. This part includes the effect of the number of coefficients on the reconstructed image's quality 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 for providing spatial and frequency representations of image simultaneously, motivate its use for feature extraction. The decomposition of input data into several layers of division in space and frequency 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.

It represents a signal by localizing it in both time and frequency domains. Wavelets can be used to improve image registration accuracy by considering both spatial and spectral information and by providing multi-resolution representation to avoid losing any global or local information. Additional advantages of using wavelet-decomposed images include bringing data with different spatial resolution to a common resolution using 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 $N \times N$ -pixel image. The size of these new images is reduced to $\frac{1}{4}$ of the original size, i.e., the new size is $N/2 \times N/2$. 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 only contains high frequency

information, is typically noisy, and therefore, is not useful for the registration. In wavelet

decomposition, only the LL image is used to produce the next level of decomposition [33].

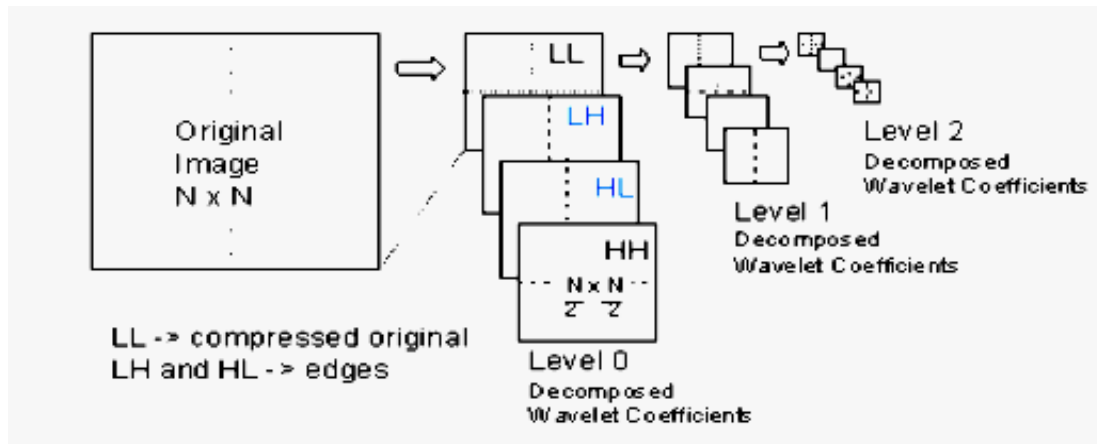


Figure 1. A 3-level wavelet decomposition of an $N \times N$ -pixel image

3. Particle Swarm Optimization (PSO)

PSO, which 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 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 GA's, 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 GA's, 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 dimensional search space. The position of the i -th particle is represented as $X_i = (x_{i1}, x_{i2}, \dots, x_{iD})$. The velocity of a particle (rate of the position change between the current position and the next) is denoted as $V_i = (v_{i1}, v_{i2}, \dots, v_{iD})$. 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:

$$V_i^{t+1} = w * v_i^t + c_1 * rand_1 * (p_{i-best} - x_i^t) + c_2 * rand_2 * (g_{best} - x_i^t) \quad (3)$$

$$x_i^{t+1} = x_i^t + v_i^{t+1} \quad (4)$$

Where $i = (1, 2, \dots, N)$ and N is the size of the swarm; pi_best is the particle best reached solution and $gbest$ is the global best solution in the swarm. $c1$ and $c2$ are cognitive and social parameters that are bounded between 0 and 2. $rand1$ and $rand2$ are two random numbers, with uniform distribution $U(0,1)$ [33]. 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 "cognitive" component representing the private experience of the particle itself; the third component is "social" component, representing the cooperation among the particles. The recursive steps will go on until

$$\text{IFrand}_3 < \frac{1}{1+e^{-vi+1}} \text{ then } x_i^{t+1} = 1 : \text{else } x_i^{t+1} = 0 \quad (5)$$

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 [33].

4.1 Chromosome Representation

The initial coding for each particle is randomly produced where each particle is coded for imitating a chromosome in a genetic algorithm; each particle was coded to a binary alphabetic string $P = F1F2...Fn$, $n = 1, 2, \dots, m$; 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 $2m$ gene space for the optimal feature subset where optimality is defined with respect to class separation. For example, when a 10- dimensional data set ($n=10$) $P = F1 F2 F3 F4 F5 F6 F7 F8 F9 F10$ is analyzed using binary PSO to select features, we can select any subset of features smaller than n , i.e. PSO can choose a random 6 features, $F1 F2 F4 F6 F8 F9$ 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 [33].

reaching to the termination condition (maximum number of iterations K).

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 a 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: [33]

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]. We have two classes and number of images within each class and we find the means of corresponding classes and the grand mean in the feature space, M_i can be calculated as:

$$M_i = \frac{1}{N_i} \sum_{j=1}^{N_i} W_j^{(i)} \quad (6)$$

where $W_j(j)$, $j=1,2,\dots,N_i$, represents the sample images from class w_i and the grand mean M_0 is:

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

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

$$F = \sqrt{\sum_{i=1}^L (M_i - M_0)^t (M_i - M_0)} \quad (8)$$

In the next step of our algorithm, we use the Euclidean distance by means of measuring the similarity between the test vector and the train vectors in each class. Equation of Euclidean distance is defined by (9):

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

Where pi (or qi) is the coordinate of p (or q) in dimension i . 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. Results and Discussion

The block diagram of the proposed is shown in Figure 2. The block diagram shows various processing steps of an input image in the training and recognition stages.

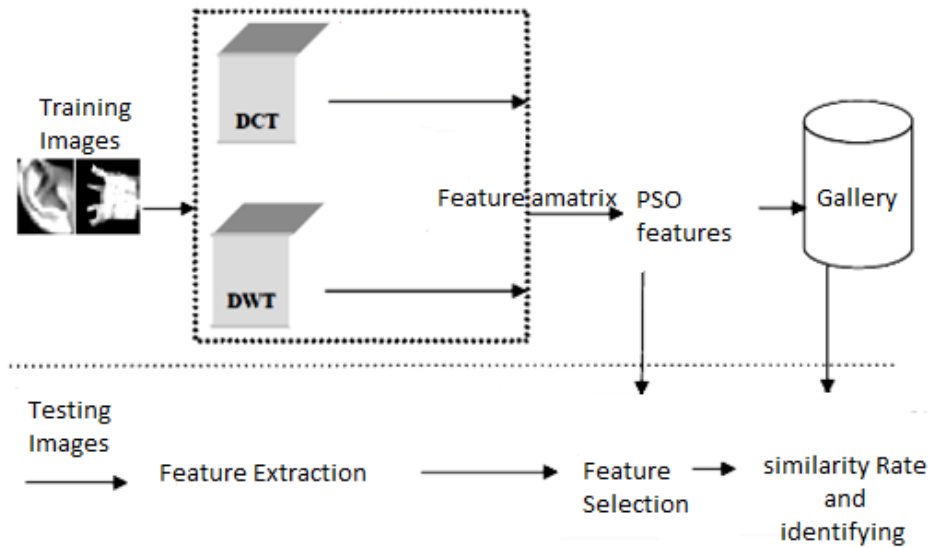


Figure 2. Block diagram of the proposed multimodal recognition system

We have constructed our database as follows: The palm and ear data consisting of 60 users [36,37]. Each user has been asked to provide three ear images and four palm impressions (of the same palm). In the preprocessing step the palm images are cropped to a size of 384×284 pixels and also ear images are cropped to a size of 150×150 pixels. The normalization step includes geometric normalization, masking and photometric normalization. In this phase, all images are scaled in a standard 80×80 size. In the next step, we remove unessential palm and ear areas with masking. Also, we rotate some

palm and ear images in each class and run our model. Different levels of masking are experimented for finding the best one to get as good performance as possible for the algorithm. Finally, the images are normalized for illumination. Then, these images are given for feature selection level. In the last section, we compare the performance of the proposed PSO-based features selection algorithm with the performance of a GA-based features selection algorithm. The parameters used for the binary PSO and the GA algorithms are given in Table 1.

Swarm size N	60
Cognitive parameters c1	2
Social parameter c2	2
Inertia weight ^w	0.6
Number of iterations	100

Table 1. (a) PSO parameter setting

The population	60
Crossover probability (pc)	0.8
Mutation probability (pm)	0.5
Number of iterations	100

(b) GA parameter setting

5.1 Experiment 1

In this test, our algorithm based on PSO had feature vectors with different subset sizes of DCT coefficients. Subset sizes 40×40 , 30×30 , 20×20 and 10×10 of the original 80×80 DCT array are used in this experiment as input to the

subsequent feature extraction phase. Table 2 showed the best average recognition rate of 96.5% which achieved by using the DCT (40×40) feature vector and the PSO-based feature selection algorithm. In general, 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 by using the PSO selection algorithm. We have found that PSO-based selection algorithm takes more time than GA-based selection

DCT	PSO	GA
40 x 40	100	70
30 x 30	80	50
20 x 20	50	30
10 x 10	20	10

Table 2. (a) Training Time (sec)

algorithm but the rate of recognition by using PSO is higher in comparison with GA. Moreover, we can claim that our method is rotate invariant.

DCT	PSO	GA
40 x 40	96.5	92.1
30 x 30	96.3	92.0
20 x 20	96.1	91.7
10 x 10	95.8	91.3

(b) Recognition Rate

In continue, DWT coefficient features have been extracted from each palm and ear image.

DWT	PSO	GA
40x40	80	60
20x20	70	60
10x10	40	30
5x5	10	10

Table 3. (a) Training Time (sec)

Table 3 shows the best average recognition rate of 97.3% by using PSO algorithm.

DWT	PSO	GA
40x40	97.3	95.2
20x20	97.1	94.6
10x10	96.8	93.3
5x5	96.3	93.1

(b) Recognition Rate

In table 4, performance of the proposed algorithm in terms of its recognition rate is compared to various feature recognition algorithms found in the literature using the POLYU and USTB databases [34,35].

Table 4 indicates the superiority of the proposed algorithm utilizing the DWT feature extraction and PSO feature selection. As far as feature selection is concerned with the algorithm, it selects the optimal number of elements in the feature vector which has a great influence on the training and recognition times of the algorithm.

Method	Recognition rate	Test condition
DCT+PSO feature selection	96.55%	Four images (two ear and two palm) per person were used in the training set and remaining images were used for testing. The average recognition time for recognizing an input image is 1.05 sec.
DWT+PSO feature selection	97.3%	
Eigen ear	80%	
Eigen palm	90%	
Eigen palm+Eigen ear	92%	

Table 4. Comparison of recognition for various feature recognition algorithm

6. Conclusion

In this paper, we used a famous algorithm called PSO in multimodal recognition systems based on the POLYU and USTB databases for palm and ear images have been used. In the step of feature extraction by using the DCT and the DWT, two feature vectors were selected. By using these techniques subset feature space was built. After feature extraction level, PSO and GA feature selection methods were used for selecting the best features and those features were the entrance of classification level. Experimental results showed PSO-based feature selection algorithm in generating excellent recognition rather than GA-based feature selection algorithm.

References

- [1] Jain, A. and Ross, A. "Information Fusion in Biometrics." In Proceedings AVBPA, Halmstad, Sweden, June 2001, pp.354-359.
- [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, 2006A.
- [13] 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] S. Cruz-Llanas, J. Fierrez-Aguilar, J. Ortega Garcia and J. Gonzalez-Rodriguez, "A Comparative Evaluation of Global Representation-Based Schemes for Face Verification", *Proc. Intl. Conf. on Image Processing, ICIAP 2003*.
- [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] Hong, L., Jain, A.k, "Integrating faces and fingerprints for personal identification," *IEEE Trans. PAMI* 20 (12), 1295-1307. 1998.
- [24] Zuev, Y., Ivanon, S., 1996. "The voting as a way to increase the decision reliability. In: Foundations of Information/Decision Fusion with Applications to Engineering Problems," Washington, DC, USA. pp.206-210.
- [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. 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] Jain, A.K., Hong, L., Pankanti, S, "Can multibiometrics improve performance?" In Proc. AutoID_99, Summit, NJ, USA. pp.59-64. 1999.
- [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.
- [32] L. Xu, A. Kryzak and C.Y. Suen, "Methods of Combining Multiple Classifiers and Their Application to Handwriting Recognition," *IEEE Trans. on Systems, Man and Cybernetics*, Vol.22, No.3, pp.418-435, May-June 1992.
- [33] R. M. Ramadan, R.Raheb, and F.Abdel-Kader "Face Recognition Using Particle Swarm Optimization-Based Selected features," *International Journal of Signal Processing*, Vol.2, No.2, June 2009.
- [34] Jain, A. K. and Ross, A., "Multibiometric Systems", In *Communications of the ACM*, Special Issue on Multimodal Interfaces, Vol.47, No.1, pp.34-40, January 2004.
- [35] Kittler, J., Hatef, M., Duin, R. P. and Matas, J.G. "On Combining Classifiers", *IEEE Transactions on PAMI* 20 (3) (1998) 226-239.
- [36] <http://www.ustb.edu.cn/resb>
- [37] http://www.comp.polyu.edu.hk/_biometrics

