REAL-TIME FACE RECOGNITION WITH ARTIFICIAL NEURAL NETWORK TRAINED BY PARTICLE SWARM OPTIMIZATION

Musa Peker[1], Huseyin Guruler[2]*

[1]Department of Information Technologies, Samandira Vocational and Technical High School, Istanbul, Turkey
[2] Department of Information Systems Engineering, Faculty of Technology, Mugla Sitki Kocaman University, Mugla, Turkey
*Corresponding Author
pekermusa@gmail.com, hguruler@mu.edu.tr

ABSTRACT

Face recognition is one of the widely used biometric method. Verification and recognition of individuals is possible via the features obtained from desired face image and compared with the facial image by various methods. Automatic face recognition which is a fundamental research area in the scope of pattern recognition, is applied in many civil, military and commercial areas for the purpose of authentication and identification. In this study a real-time face recognition system was developed. It is aimed that identification of individuals who entering any field observed with a camera. After detecting the important facial points, they are presented as input data to feed-forward neural network. Particle swarm optimization was used as learning algorithm in the network. As a result, a novel real-time face detection method, which provide high accuracy has been developed.

Keywords: Real time face detection, pattern recognition, neural network, particle swarm optimization.

INTRODUCTION

Face detection and face recognition on images is one of the important subjects in computer vision applications. Previously saved images of each person are used in face recognition of persons. Faces to be recognized can be obtained from the images obtained in controlled conditions (images on passport, credit cards, ID, etc.), may also be obtained from real time video recordings (Peker & Zengin, 2010). Face recognition problem contains finding the faces in images, determining of limits, finding of attributes and classification of faces using attributes (Peker & Zengin, 2011).

In recent years, the issue of face recognition systems has been investigated from many aspects. Variety of techniques for different aspects and details of the subject and efficient algorithms and methods have been proposed (Turk & Pentland, 1991; Swets & Weng, 1996; Chen et al., 2000, Yu & Yang, 2001).

Face recognition is a difficult process. Although numerous applications related to facial recognition has been developed, studies are ongoing. The reason for this, factors such as 3D exposure differences, different facial expressions, lighting differences, makeup, hair style, background differences and noise make face recognition quite difficult.

The purpose of this study is to accurately perform face detection and face recognition process as real time. A new approach is presented in this context. After face recognition, important attributes of face on images are determined automatically by Gabor wavelet transform. These attributes are presented as an input to the neural network trained with particle swarm algorithm. As a result of the operation of neural network, the owner of the face is determined. The most important feature of the study is the ability to find and recognize multiple faces simultaneously. The results were obtained are promising.

METHODS

a. Skin color identification algorithm

In this study, skin color based two algorithms were used in order to ideally detect human face region. These algorithms are RGB and $YCbCr$ code techniques. Because red and green color tones are more in skin color according to RGB code technique, the following equations are used.

$$R = R/(R+G+B);\ G = G/(R+G+B) \tag{1}$$

With using RGB code technique, different skin colors were examined and the most appropriate value range was attempted to be determined in order algorithm to detect skin color region ideally. Accordingly, ($R > 0.40$) and ($0.25 < D < 0.33$) ranges were determined as the optimum value ranges to detect skin color regions.
According to $YCbCr$ code technique, in $YCbCr$ color space Y represents the brightness information, Cb and Cr represent color information. Thus, the brightness information is easily obtained. RGB color space can be converted into $YCbCr$ color space by equation (2).

$$\begin{aligned} Y &= 0{,}299R + 0{,}587G + 0{,}114B \\ Cb &= -0{,}169R - 0{,}332G + 0{,}500B \\ Cr &= 0{,}500R - 0{,}419G - 0{,}081B \end{aligned} \tag{2}$$

Cb and Cr values are used in skin color region finding process. By calculating the maximum and minimum values of Cb and Cr components, the pixels between these values are marked as skin color. Minimum and maximum values of Cb and Cr components are calculated as shown in equation (3) (Kurt et al., 2007).

$$\begin{aligned} Cbmin &= Ort\_Cb - Std\_Cb * f\ ;\ Cbmax = Ort\_Cb + Std\_Cb * f \\ Crmin &= Ort\_Cr - Std\_Cr * f\ ;\ Crmax = Ort\_Cr + Std\_Cr * f \end{aligned} \tag{3}$$

Minimum and maximum values of Cb and Cr components are determined according to the average value and standard deviation of these components (Kurt et al., 2007).

b. Gabor Wavelet Transform (GWT)

Wavelet transform with Gabor main function is expressed as Gabor Wavelet Transform. Gabor wavelets show great similarities with the human visual system according to the frequency and orientation characteristics. These wavelets are used in computer vision applications, face recognition, fingerprint recognition and classification algorithms (Acar & Özerdem, 2012). Gabor wavelets constitute an excellent filter for both spatial localization and orientation. A complex Gabor wavelet (filter) is defined as multiplication of a complex sinusoid with Gaussian kernel. A two-dimensional Gabor wavelet transform is expressed by convolution of the image of I(x,y) (Acar & Özerdem, 2012; Buciu & Gacsadi, 2009):

$$J(x,y) = \iint I(x',y')g(x-x',y-y')\,dx'dy' \tag{4}$$

$g(x,y)$ function represents Gabor filter:

$$g(x,y;\lambda,\varphi,\omega,\sigma,\gamma) = exp\left(-\frac{x_1^2+\gamma^2 y_1^2}{2\sigma^2}\right) exp\left(i\left(2\pi\frac{x_1}{\lambda}+\omega\right)\right) \tag{5}$$

$$x_1 = x\cos\varphi + y\sin\varphi \tag{6}$$

$$y_1 = -x\sin\varphi + y\cos\varphi \tag{7}$$

The above-described parameters λ and φ, represent the wavelength factor of the cosine (scale) and the direction of Gabor function (angular orientation), respectively. ω indicates offset value of phase, and γ indicates spatial visual angle.

Parameters calculated from GWT (Acar & Özerdem, 2012):
Assume that the matrix obtained from each GWT matrix regarding gray tone images is an MxN-dimensional $I(i,j)$ matrix. Accordingly:

Mean : $$o_{ij} = \frac{1}{MN}\sum_i^N\sum_j^M I_{ij} \tag{8}$$

Standard deviation : $$s_{ij} = \frac{1}{MN}\sum_i^N\sum_j^M (I_{ij}-m_{ij})^2 I_{ij} \tag{9}$$

Entropy : $$e_{ij} = -\sum_{i,j} I(i,j).\log(I_{ij}) \tag{10}$$

c. *Artificial Neural Network*

Artificial neural networks (ANN) are mathematical systems consisting of many processing units weighted and connected to each other (Sen & Peker, 2013). This processing unit receives signals from other neurons; combines them, transforms, and reveals a numerical result. In general, the processing units roughly correspond to the actual neurons and interconnected in a network; this structure constitutes neural networks. In this study, feed forward neural network of the neural network models was used. There are basically three different layers in feed forward neural networks. These layers, respectively; the input layer that holds data going into artificial neural network, hidden layer or layers on which processes are done and train itself according to desired result, and finally output layer which shows output values.

d. *Particle Swarm Optimization*

In particle swarm optimization, each solution is called as particle in the search space. All the particles have relevancy value evaluated by the relevancy function to be optimized and particle velocity information directing their movements. Particles follow the existing optimum particles in the problem space (Bakbak & Peker, 2013).

PSO is initialized with random particle swarm and the optimum value is searched with update. In each iteration, each particle is updated according to the best two values. One of them is the best relevancy value found by the particle so far called pbest. This value is kept in memory for later use. Second best value is the best relevancy value found by any particle in swarm so far called gbest. It is the best global value in the swarm (Yalcin et al., 2013).

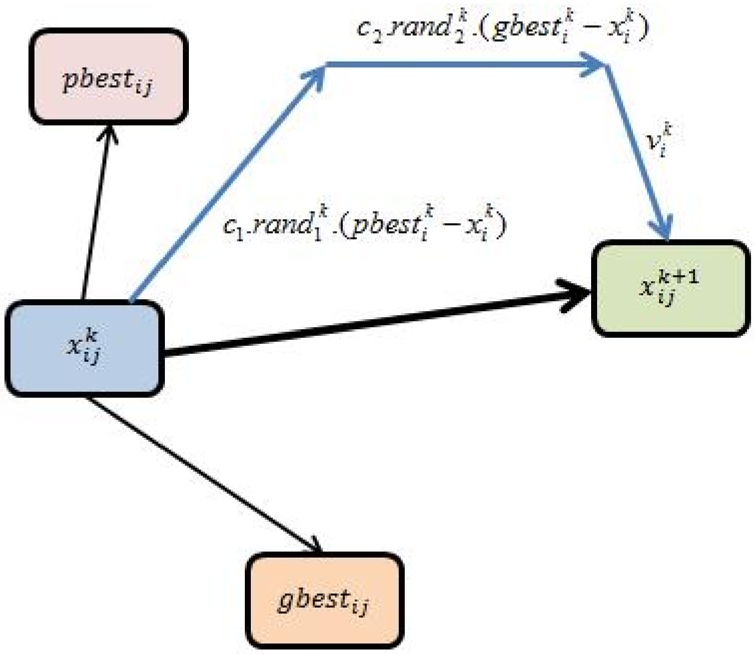


Figure 1. The velocity and position updating of a particle at kth generation (Yalcin et al., 2013).

Swarm matrix with D swarn dimension and n particle size is described as follows.

$$x = \begin{bmatrix} x_{11} & x_{12} & \dots & x_{1D} \\ x_{21} & x_{22} & \dots & x_{2D} \\ x_{31} & x_{32} & \dots & x_{3D} \\ \dots & \dots & \dots & \dots \\ x_{n1} & x_{n2} & \dots & x_{nD} \end{bmatrix}_{nxD} \quad (11)$$

According to the swarm matrix ith particle is described with

$$x_i = [x_{i1}, x_{i2}, x_{i3},,,, x_{iD}] \quad (12)$$

and the $pbest$, best relevancy value found by the particle so far, is

$$pbest_i = [p_{i1}, p_{i2}, p_{i3},,,, p_{iD}] \quad (13)$$

global good within the population

$$gbest = [p_1, p_2, p_3,,,, p_D] \quad (14)$$

ith is described as a velocity vector indicating the amount of change in each position of the particle.

$$v_i = [v_{i1}, v_{i2}, v_{i3},,,, v_{iD}] \quad (15)$$

Particle's velocity and position is updated according to the following equations, respectively.

$$\begin{gathered} v_i^{k+1} = v_i^k + c_1.rand_1^k.(pbest_i^k - x_i^k) + c_2.rand_2^k.(gbest_i^k - x_i^k) \\ x_i^{k+1} = x_i^k + v_i^{k+1} \end{gathered} \quad (16)$$

Where k is the number of iterations and i is the number of particles. If the particle swarm matrix consists of n rows, it means that i. line is being mentioned. $c1$ and $c2$ values which are the learning factors, pull the particle to $pbest$ and $gbest$ values. $c1$ and $c2$ usually selected as equal and in [0,4] range. $c1$ allows particle to move according to the particle's own experience, $c2$ allows particle to move according to the experience of other particles in the swarm.

APPLICATION AND EVALUATION

e. Face Detection

The biggest problem encountered in face detection is the existence of the areas with a color close to skin color outside the human face area (Ikizler & Duygulu, 2005). System detects these areas as a part of the face. After converted to gray level, image is converted to black-and-white picture in order to provide a fast and accurate work on color image. Equation (17) was used in the conversion of the image to gray level.

$$I = (R + G + B)/3 \quad (17)$$

Thresholding method was used to convert the image to black and white picture. According to this method; pixels of skin area determined by skin color algorithms are transformed into white color, while other areas are transformed to black. Thus, the image is transformed into binary level. In binary system, 0 refers to black while 1 refers to the white color. Binary format of the image is seen on Figure 2.a. After the image is converted to binary format, the image is filtered and ensured that image is not affected by unnecessary noises. (Figure 2). As a filter, median filter, which aims to soften the image, was used as a 3x3 matrix. Median filter is a nonlinear filter that protects the edges and eliminates random noises (Umbaugh, 1998).

As shown in Figure 2.b, although picture was filtered, noise cleaning cannot be fully ensured. Therefore, the image was passed through a scanning filter which has 3x3 matrix. This filter determines the area with maximum skin color by scanning the display screen.

Figure 2. (a) Picture with noise (b) Noise-free picture

As stated earlier, in this study RGB and $YCbCr$ code techniques, which are skin color based algorithms, were used. In this study, results of algorithms using these code techniques were compared and it was found that the algorithm in which $YCbCr$ code technique is used provided more successful outcome. The reason for this, the algorithm in which $YCbCr$ code technique is used, is less affected from the factors such as ambient brightness, dust, etc. In Figure 3, the difference between these two algorithms is seen more clearly.

Figure 3. (a) detection of the areas with skin color using RGB code technique (b) detection of the areas with skin color using $YCbCr$ code technique.

f. Feature Extraction

Gabor based feature vector related to an image was obtained with statistical values of each of the 8 wavelet matrix (2 scales and 4 orientations) related to an image. These values are combined in a vector and feature vector is obtained. From each wavelet transform matrix, standard deviation, mean and entropy values were calculated respectively. As a result, statistical values of 8 wavelet transform matrix related to each image were calculated and feature vector with 8x3 = 24 data length in total was obtained by adding these values consecutively. The resulting attributes were applied to the input of the classifier.

g. *Face Recognition with Neural Networks Trained by PSO*

For the realization of learning in artificial neural networks, weight values between the layers must be appropriately updated. In this study, unlike classical training algorithms, the PSO, a powerful optimization algorithm was preferred. In figure 4, a flow chart in which testing and training of ANN with PSO took place is presented. In the learning phase, primarily, the weights that hold the numerical value of connections between layers, take random values. These weight values represent particle values for PSO. Number of connections between the layers denotes the size of particles (Yalcin et al., 2013).

Network is established according to each particle and training examples are respectively sent to the network. After the example is presented to the network, the difference (error) between the actual value should be and the value obtained as output is calculated. After all the samples submitted to the network, total error (MSE) is calculated and the obtained value is regarded as the particle's relevancy value. In the first step, this relevancy value is assigned as pbest value of the particle; the best relevancy value among the particles is assigned as the gbest value (Yalcin et al., 2013). If relevancy value (error) is not in an acceptable level, particles are updated with pbest and gbest values. Network is re-established according to the new particle values, examples are given to the network again and the relevancy value calculation is performed. These processes continue until the best relevancy value obtained so far (gbest) reaches to the desired value or the maximum iteration (Yalcin et al., 2013).

If the error in an acceptable level, the testing process begins. This time, network is established according to the gbest particle values. Test samples are sent respectively to the input layer of the network and the resulting values is given as output of the example. If any threshold is not applied to the output of the network, last obtained gbest value gives the classification performance of the network (Yalcin et al., 2013; Yalcin, 2012).

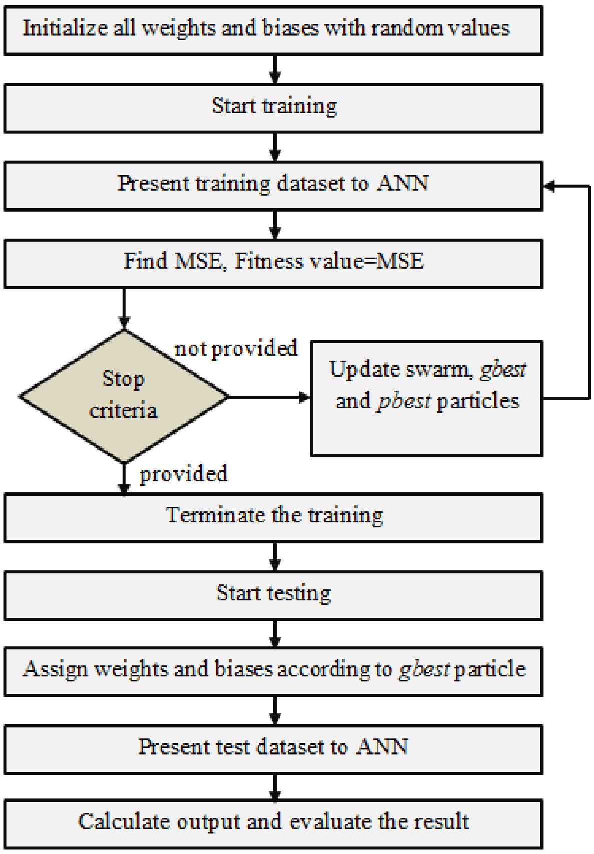


Figure 4. Flowchart for training and testing of PSONN

24 attribute values obtained in feature extraction step are presented as an input to the neural network structure. With PSO algorithm, network parameters are updated and training of network is carried out. Example of test result is presented in Figure 5. When the figure is considered, it is seen that the amount of error is decreased depending on the increase in the value of iteration.

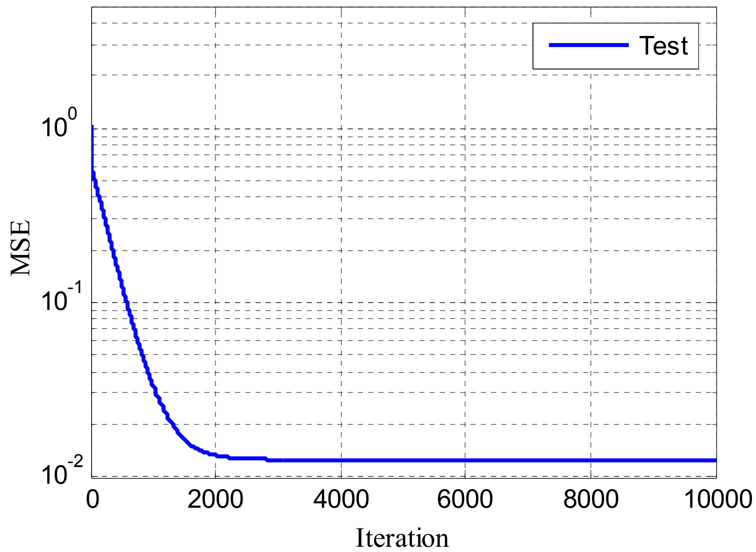


Figure 5. Iteration-error graph

The face detection performance of the developed software was investigated under various conditions. Primarily, the success of the application was evaluated as real-time. There is no change in the success of face recognition in case of an increase in the number of faces in the image. System is able to find more than one face within the field of view of the camera with success. Considering the results of the analyses presented in Table 1, the success of the face recognition process is increase when the image taken frontally and in adequate light conditions. This is observed that performance is decreased when viewing angle move away from the front. Face recognition performance of the system decreases in low light conditions. In order to solve this problem, using of quality and multiple cameras and providing of sufficient light conditions are proposed.

Table 1. Performance analysis of face detection under different conditions.

Light Condition	Angle	Distance	Success Level
Sufficient	From the front line	1,5 m	High
Sufficient	From the front line	2 m	High
Sufficient	30 degrees	1,5 m	High
Insufficient	From the front line	1,5 m	Medium
Insufficient	30 degrees	1,5 m	Medium
Insufficient	90 degrees	1,5 m	Medium

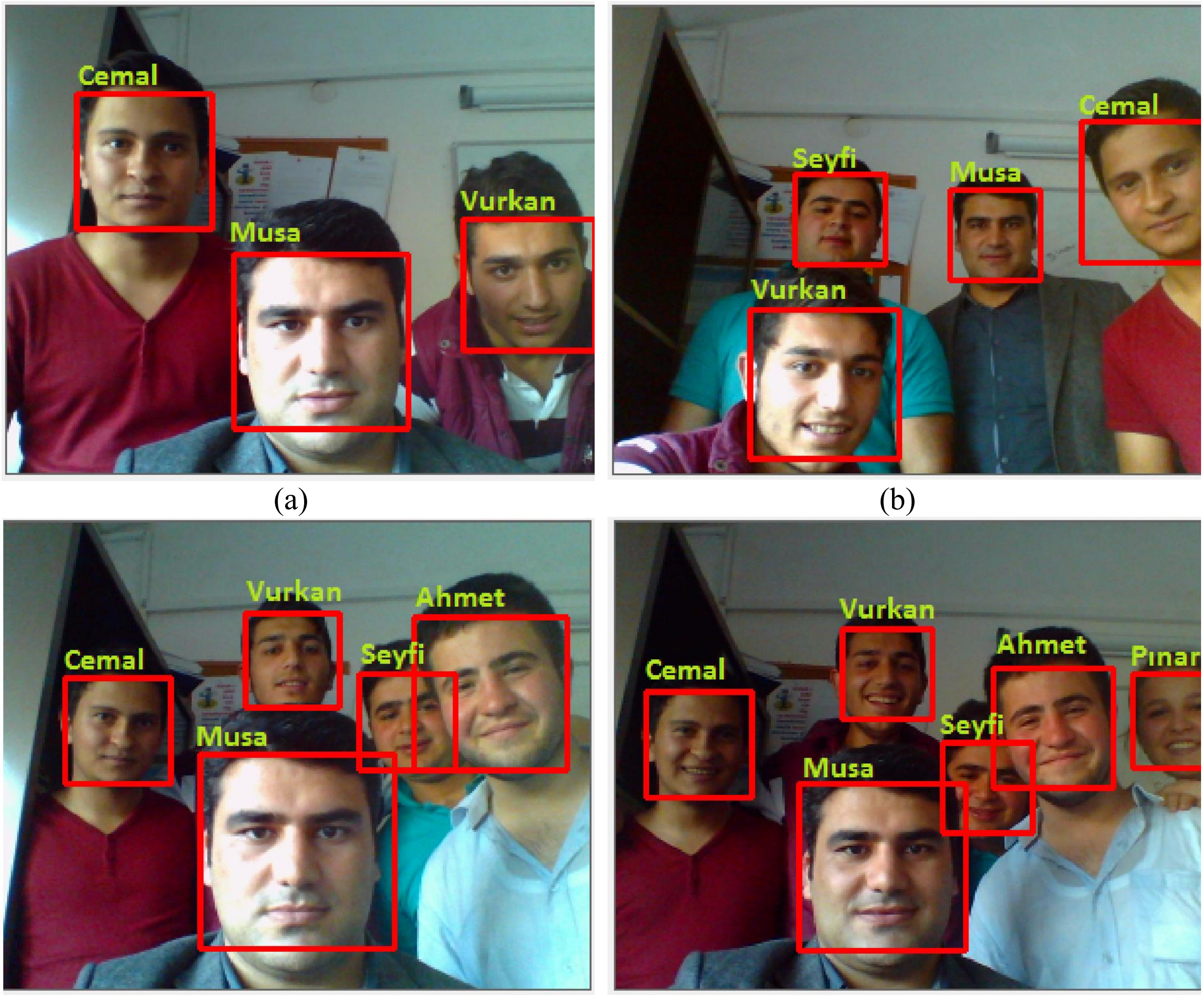
Face recognition experiments were conducted in two stages. In the first stage, experiments were performed real-timely. In the second stage, experiments were performed on ORL database (The Database of Faces, 2014).

Experiment 1: Experiments have been tried on 150 people. In experiments conducted by one person, number of faces recognized as incorrect was two. In this case, success rate was found as % 98.6. In multi-person applications, depending on the number of different person (2, 3, 4 and 5), number of faces recognized as incorrect was six. In this case, success rate was found as % 96. The results have shown in Table 2.

Table 2. Performance analysis for 150 people with the recommended method (real time application).

Number of people	The number of faces	Wrong number	Success rate
One person	150	2	98.6 %
Multi person (2, 3, 4 and 5)	150	6	96 %

Figure 6 presents an example of the real-time implementation of this application.



(a) (b) (c) (d)

Figure 6. Real-time face recognition and detection
a) Three people b) Four people c) Five people d) Six people

Experiment 2: Table 3 shows the analysis of the success of face recognition carried out in different conditions. ORL database were used in performance analysis. Analyzes of success were performed by changing the total number of images, size of the images and the number of images received from each person. An image set consisting of 300 people was created with 10 different pictures (taken from various fronts) of 30 different people. A database was created with 30 people by taking one picture of each person in the cluster. Using the remaining 270 pictures, face recognition was performed in different image sizes. Then, performance analyses were carried out by taking 2, 3 and 5, different image of each person, respectively.

Table 3. Performance analysis for 300 people with the recommended method.

Number of image	Number of images per person	Database image number	Image size	Accuracy rate (%)
270	1	30	16x14	76.5
270	1	30	27x24	77.3
270	1	30	40x30	78.4
240	2	60	16x14	86
240	2	60	27x24	86.7
240	2	60	40x30	88.5
210	3	90	16x14	90
210	3	90	27x24	91
210	3	90	40x30	94
150	5	150	16x14	97
150	5	150	27x24	97.5
150	5	150	40x30	98.5

In the obtained analyses, it was observed that as the number of samples taken from the same individual increased, performance of the system also increased. It was also observed that when the image size is increased, performance slightly increased but, face recognition process took a long time.

RESULTS

Face detection and analysis are often used in different fields. Face recognition is used in a wide range such as defense industry, security systems, robotics industry, for commercial purposes.

Face recognition problem is one of the up-to-date, important and difficult problems. Many scientists have been working on this issue for a long time. However, due to the difficulty of the problem, face recognition systems which have the success for solving real-life problems have not been developed yet.

In this study, software which performs real time face recognition and face detection has been developed. After face detection was performed in images, important features of the face were detected. Gabor wavelet transform was preferred for attribute determination. Obtained attributes were presented as input to the neural network trained with particle swarm algorithm. In performance tests carried out with face recognition system, it was observed that the system performance is in the acceptable quality.

REFERENCES

Acar, E., & Özerdem, M.S. (2012). Kızıltepe tarımsal alan imgelerinin gabor dalgacık dönüşümü ile sınıflandırılması, *IEEE 20. Sinyal İşleme ve İletişim Uygulamaları Kurultayı (SİU2012)*, Fethiye, 2012.

Bakbak P.O., & Peker, M. (2013). Particle swarm optimization design of optical directional coupler based on power loss analysis. *International Journal of Intelligent Systems and Applications in Engineering*, 1(2), 29-33.

Buciu, I., & Gacsadi, A. (2009). Gabor Wavelet Based Features for Medical Image Analysis and Classification”. *Applied Sciences in Biomedical and Communication Technologies (ISABEL 2009)*, (pp. 1-4).

Chen, L.F., Liao, H.Y.M, Ko, M.T., Lin, J.C., & Yu, G.J. (2000). A new LDA-based face recognition system which can solve the small sample size problem. *Pattern Recognition*, 33, 1713-1726.

Ikizler, N., & Duygulu, P. (2005). Haber videolari icin yüz bulma yöntemlerinin.iyilestirilmesi, *In Proceedings of IEEE 13. Sinyal Isleme ve Iletisim Uygulamalari Kurultayi (SIU 2005)*, Kayseri, (pp. 1-4).

Kurt, B., Nabiyev, V., & Bekiroglu, Y. (2007). Yüz ifadelerinin tanınması. *Elektrik-Elektronik-Bilgisayar Mühendisliği 12. Ulusal Kongresi*, (pp. 1-5).

Peker, M., & Zengin, A. (2010). Real-Time Motion-Sensitive Image Recognition System. *Scientific Research and Essays*, 5 (15), 2044-2050.

Peker, M, & Zengin, A. (2011). A Real-Time and Motion-Sensitive Security Application with Face Recognition. *6th International Advanced Technologies Symposium (IATS'11)*, Elazığ, Turkey, (pp. 92-97).

Sen, B., & Peker, M. (2013). Novel approaches for automated epileptic diagnosis using FCBF feature selection and classification algorithms. *Turkish Journal of Electrical Engineering & Computer Sciences*, 21, 2092-2109.

Swets, D. L., & Weng, J. (1996). Using discriminant eigenfeatures for image retrieval. *IEEE Transaction on Pattern Analysis and Machine Intelligence*, 18(8), 831-836.

The Database of Faces, "The ORL database of face", http://www.cl.cam.ac.uk/ research/dtg/attarchive/facedatabase.html, Access time: 01.01.2014.

Turk, M., & Pentland, A.P. (1991). Eigenfaces for recognition. *Journal of Cognitive Neuroscience*, 3(1), 71-86.

Umbaugh, S. (1998). Computer vision and image processing fundamentals. *Computer Vision and Image Processing*. Prentice Hall PTR, Bernard Goodwin.

Yalçın, N. (2012). Heuristic algorithm basis artificial neural networks for epilepsy detection. The Graduate School of Natural and Applied Science of Selçuk University, *MS Thesis*, Konya.

Yalçın, N., Tezel, G., & Karakuzu, C. (2013). Epilepsy Diagnosis Using Artificial Neural Network Learned by PSO. *Turkish Journal of Electrical Engineering & Computer Sciences*, online, DOI: 10.3906/elk-1212-151.

Yu, H., & Yang, J. (2001). A direct LDA algorithm for high-dimensional data with application to face recognition. *Pattern Recognition*, 34, 2067-2070.

Musa Peker graduated from Zonguldak Karaelmas University in 2007. Received his master's degree in 2009 from Sakarya University. He received his Ph.D. in Computer Engineering from Karabuk University in 2014. Currently working in Istanbul Samandira Vocational High School as IT teacher. His interests are biomedical signal processing, image processing and artificial intelligence applications.

Huseyin Guruler is academic staff in the Department of Information Systems Engineering in Mugla Sitki Kocman University, Turkey. His bachelor's degree is in the field of electronics and computer from Marmara University. MSc is in the field of statistics and computer in Mugla Sitki Kocman University. PhD is in the field of electronics and computer in Sakarya University. PhD thesis is about diagnosing sleep apnea using ECG signals. His research interests merge in data mining and knowledge discovery besides dealing with computational biology and multi-user computers architecture.