



Contents list available at JMCS

Journal of Mathematics and Computer Science

Journal Homepage: [www.tjmcs.com](http://www.tjmcs.com)



## A New Neural Network Approach for Face Recognition based on Conjugate Gradient Algorithms and Principal Component Analysis

Hamed Azami

*Department of Electrical Engineering, Iran University of Science and Technology*  
*hamed\_azami@ieee.org*

Milad Malekzadeh

*Department of Electrical and Computer Engineering, Babol University of Technology*  
*m.malekzade@stu.nit.ac.ir*

Saeid Sanei

*Faculty of Engineering and Physical Sciences, University of Surrey, UK,*  
*s.sanei@surrey.ac.uk*

### Article history:

Received January 2013

Accepted February 2013

Available online February 2013

### Abstract

This paper presents a new approach based on conjugate gradient algorithms (CGAs) and principal component analysis (PCA) for face recognition. First, images are decomposed into a set of time-frequency coefficients using discrete wavelet transform (DWT). Basic back propagation (BP) is a well established technique in training a neural network. However, since in this algorithm the steepest descent direction is not the quickest convergence, it is slow for many practical problems and in many cases including face recognition, its performance is not satisfactory. To overcome this problem, four algorithms, namely, Fletcher-Reeves CGA, Polak-Ribikre CGA, Powell-Beale CGA, and scaled CGA have been proposed. Also, in this paper the PCA as a pre-processing step to create the uncorrelated and distinct features of the DWT of images is used. The simulation results show that all of the proposed methods, compared with the basic BP, have greater accuracies.

**Keywords:** Face recognition, discrete wavelet transform, conjugate gradient algorithm, and principal component analysis

## 1. Introduction

A method that can verify or indentify a person from a digital image is named face recognition. As a special study of pattern recognition, face recognition has proved to be very useful in daily life such as for security access control systems, content-based indexing, and bank teller machines [1-5]. Generally, there are two kinds of approaches to face recognition, namely, feature-based and template matching (holistic approaches). The feature-based approach is based on the shapes and relationships between the individual facial features containing mouth, nose, eye and chin. On the other hand, the holistic approach handles the face images globally and extracts important facial features based on high-dimensional intensity values of face images automatically [6]. Although feature-based approaches are more robust against rotation, scale, and illumination variations, they significantly depend on the accuracy of facial feature detection methods. It has been argued that the existing feature-based methods are not reliable enough for extracting individual facial features [6]. On the other hand, due to the well-known statistical methods in holistic face recognition, it has attracted more attention compared with the feature-based approach [6].

In previous methods, the basic back propagation (BP) was used to train a neural network (NN) for classifying the images [7]. Although the basic BP is a widely used method for training an NN, in many cases, the BP has many deficiencies such as too slow convergence speed, easy to fall into local minimum, and easily affected by sudden peaks in the signal trend during the learning process. To overcome these problems, in this paper, four kinds of conjugate gradient algorithms (CGAs), including Fletcher-Reeves CGA, Polak-Ribikre CGA, Powell-Beale CGA, and scaled CGA have been proposed for training a feed-forward NN.

In the basic BP, the weights in the steepest descent direction (negative of the gradient), the direction in which the performance function is lessening most quickly, are adjusted. It should be mentioned that although the error function decreases most quickly along the negative slope of the gradient, it does not unavoidably generate the quickest convergence. In the CGAs a search is done along conjugate directions, which normally generates faster convergence than that of the steepest descent directions [1].

In order to enhance the performance of the best training algorithm for training an NN, principal component analysis (PCA) is used after extracting the features by DWT. PCA is a useful statistical technique that is widely used in many applications such as face recognition and image compression, and it is a common technique to find patterns in data of high dimension. PCA performs a covariance analysis between coefficients and find the projection directions corresponding to the largest data variation. These directions are determined by the eigenvectors of the covariance matrix corresponding to the largest eigenvalues. The obtained uncorrelated features of the PCA are optimal to classify an NN [8]. In this paper we use the PCA only for the best CGAS algorithm, i.e. Powell-Beale CGA.

The rest of this paper is organized as follows. The DWT as a feature extraction method is introduced in Section 2. Then, four types of the CGAs named before are briefly explained in Section 3. In Section 4 we provide our experimental results of applying our proposed method on the ORL dataset. Finally, conclusions are given in Section 5.

## 2. Discrete Wavelet Transform

DWT plays a significant role in reducing the dimension of an image and extract the features by decomposing an image in frequency domain into sub-bands at different scales. The DWT of an image is

created as follows: In the first level of decomposition, the image is split into four sub-bands, namely HH1, HL1, LH1, and LL1, as shown in Figure 1. The HH1, HL1 and LH1 sub-bands represent the diagonal details, horizontal features and vertical structures of the image, respectively. The LL1 sub-band is the low resolution residual consisting of low frequency components and it is this sub-band which is further split at higher levels of decomposition [1]. In Figure 2 an image from ORL face database with decomposed one-level wavelet and after three-level wavelet transform are shown, respectively.

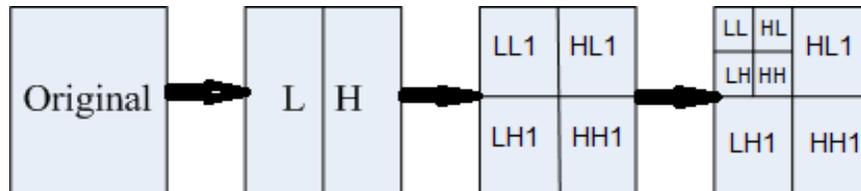


Figure 1. The process of decomposing an image

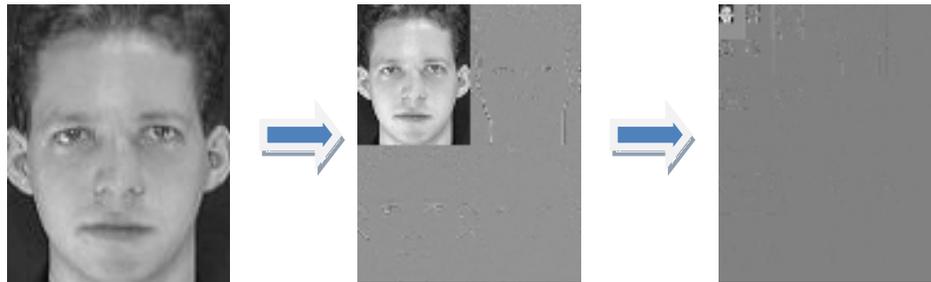


Figure 2. Original image and the DWT after one-level and three-level DWT respectively.

### 3. Conjugate Gradient Algorithms

In the basic BP, the weights in the steepest descent direction are adjusted. Although the performance function decreases most quickly along the negative slope of the gradient, it does not necessarily generate the fastest convergence [1,9]. Unlike the basic BP, in CGA the weights and biases are not adjusted in the steepest descent direction. In fact, the CGA makes an NN converge to the desired responses faster than basic BP along the conjugate directions.

Generally, there are four types of CGAs for training a feed-forward NN, namely, Fletcher-Reeves CGA, Polak-Ribikre CGA, Powell-Beale CGA, and scaled CGA. All the CGAs include the following steps [1, 10]:

**Step 1:** The algorithm is started by searching in the steepest descent direction as follows:

$$p_0 = -g_0 \tag{1}$$

where  $p_0$  and  $g_0$  are the search vector and gradient in the first iteration, respectively.

**Step 2:** The CGA takes a step and selects the learning rate to minimize the function along the search direction as follows:

$$x_{k+1} = x_k + \alpha_k p_k \tag{2}$$

where  $x_k$  demonstrates the estimate of the minimum at the start of the  $k$ -th iteration. Equation (2) expresses that new estimate  $x_{k+1}$  is computed using search direction ( $p_k$ ) and the length of the step preformed at iteration  $k$  ( $\alpha_k$ ).

**Step 3:** Then, CGA determines the next search direction. The general procedure for determining the new search direction is generated by combining the new steepest descent direction with the previous search direction according to:

$$p_k = -g_k + \beta_k p_{k-1} \tag{3}$$

where  $\beta_k$  is a positive scalar that varies in different versions of CGAs.  $\beta_k$  for Fletcher-Reeves CGA is defined as below:

$$\beta_k = \frac{g_k^T g_k}{g_{k-1}^T g_{k-1}} \tag{4}$$

where  $g_{k-1}$  and  $g_k$  are the previous and current gradients, respectively. Fletcher Reeves suggested restarting the CGA after every  $n$  or  $n+1$  iterations. This suggestion can be very helpful in practice [10].

The second version of the CGA, namely, Polak-Ribiere CGA, was proposed in 1977 [11]. In this algorithm, the search direction at each iteration is computed as [12]:

$$\beta_k = \frac{\Delta g_k^T g_k}{g_{k-1}^T g_{k-1}} \tag{5}$$

Powell has showed that Polak-Ribiere CGA has better performance compared with Fletcher-Reeves CGA in many numerical experiments [10].

The mentioned CGAs, Fletcher-Reeves and Polak-Ribire algorithms, periodically restart by using the steepest descent direction every  $n$  or  $n+1$  iterations. A disadvantage of the periodical restarting is that the immediate reduction in the objective function is usually less than it would be without restart. Therefore, Powell proposed a new method of restart based on an earlier version [10,12]. As stated by this method, when there is negligibility orthogonality left between the current gradient and the previous gradient (the following condition is satisfied), the restart occurs [10,12].

Another version of the CGA, scaled CGA, was proposed by Moller. This algorithm uses a step size scaling mechanism that avoids a time consuming line-search per learning iteration. It makes the algorithm faster than other recently proposed second order algorithms. Moller in [13] has showed that the scaled CGA has superlinear convergence for most problems.

#### 4. Performance Evaluation

The ORL face database contains 400 images of 40 individuals (10 different images from each person) with various facial expressions and lighting conditions. A number of images from this database are illustrated in Figure 3.



**Figure 3. Sample images from ORL face database.**

In order to extract features, at the first step we implemented the Haar wavelet with four-level decomposition and then features from the lower sub-band coefficients were chosen. For all these methods, we train a feed-forward NN with 50% of the images in the database (the odd numbers of images for each person) and then used the rest for testing (the even numbers of images for each person).

Deciding how many neurons to use in the hidden layer is one of the most important characteristics in an NN. When the number of neurons is too low, the NN cannot model complex data and the resulting may be unacceptable. If too many numbers of neurons are used for an NN, it not only increases the training time, but also may reduce the performance of the NN. Therefore, we change the number of neurons of the hidden layer from 10 to 200, then, test their performances.

Generally, there are three functions, namely, linear, tangent sigmoid and logarithm sigmoid that are used as the activation functions for an NN. These factors considerably affects in the performance of the approximation or classification of data. We select the linear activation function for the NN with training by all of the CGAs with trial and error. Also, in training by the basic BP, tangent sigmoid function is experimentally chosen. In fact, for each NN training method, the mentioned function had the best performance.

When we increase the number of iterations to more than 700, the accuracies of the proposed methods don't change significantly, but the training time increases considerably. Thus, we suggest using 700 iterations for training a feed-forward NN by using the improved BP training algorithms. Also, because of uncertain behavior of NNs, we run all algorithms 40 times, and the average of the results is presented.

Figures 4 to 8 represent recognition rates of the ORL dataset, the number of features and number of hidden layer neurons for basic BP and the four proposed methods, namely, Fletcher-Reeves CGA, Polak-Ribikre CGA, Powell-Beale CGA and scaled CGA, respectively. The achieved results indicate for each proposed training methods 70 neurons result an acceptable accuracy.

Table 1 shows a comparative analysis of the best obtained recognition rates of the mentioned trainings for the NN with 700 iterations. As illustrated in Table 1, all the proposed methods (four improved BP training algorithms) classify the ORL data better than the basic BP. Also, this table represents that the accuracy and efficiency of these algorithms increase as follows: simple BP, scaled CGA, Polak-Ribikre CGA, Fletcher-Reeves CGA, and Powell-Beale CGA, respectively.

Finally, it has been found that the fastest algorithm or the best classification for this application between the proposed methods is obtained by using the Powell-Beale CGA with one hidden layer and 110 neurons. It should be mentioned that training times for all of the proposed methods are approximately equal.

The advantages of the proposed methods are high accuracy and low CPU time. In this paper the ORL data classification has been improved about 8.53% in accuracy by the Powell-Beale CGA compared with that the basic BP.

Also, in order to increase the recognition rate we use the PCA before using the NN for the Powell-Beale CGA as the best CGA. The PCA can create independent and informative data and increase the accuracy from 96.03% to 96.5%.

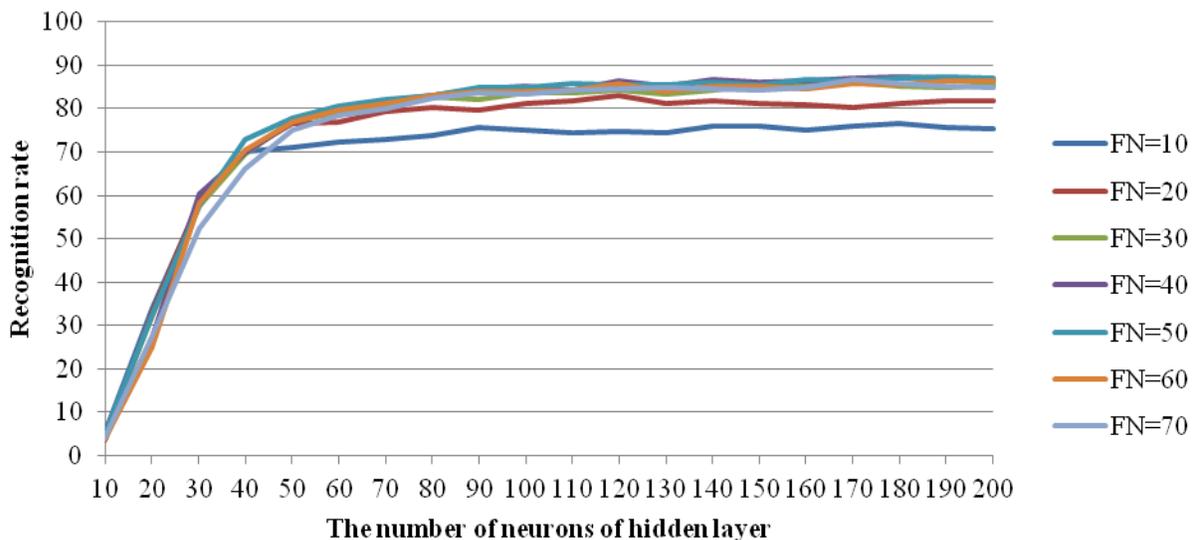


Figure 4. Accuracy of the ORL dataset classification using the basic BP for different the number of features and hidden neurons

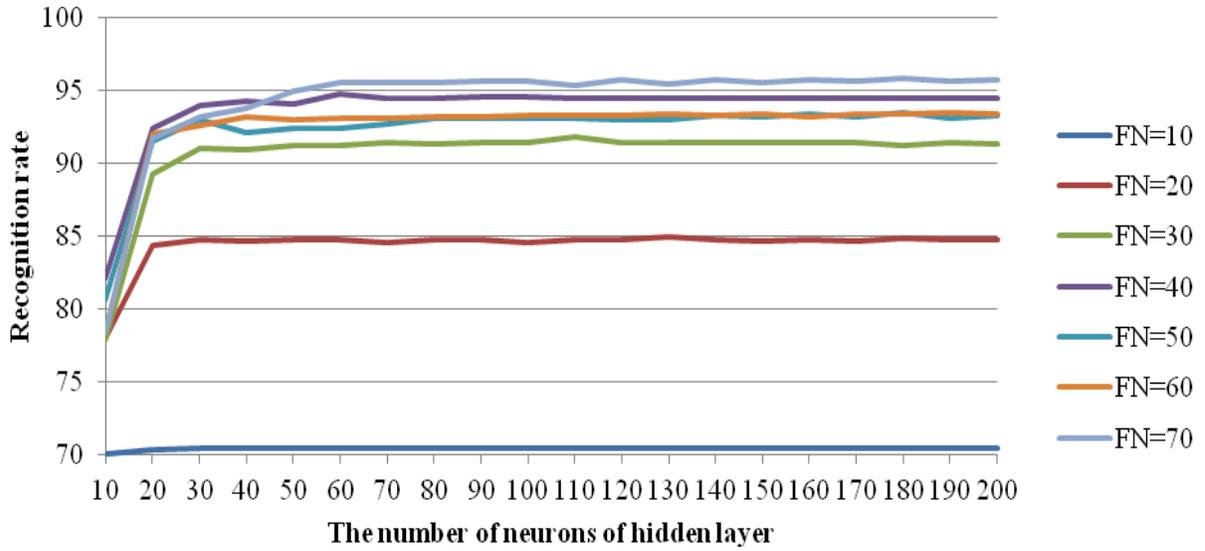


Figure 5. Accuracy of the ORL dataset classification using the Fletcher-Reeves CGA for different the number of features and hidden neurons

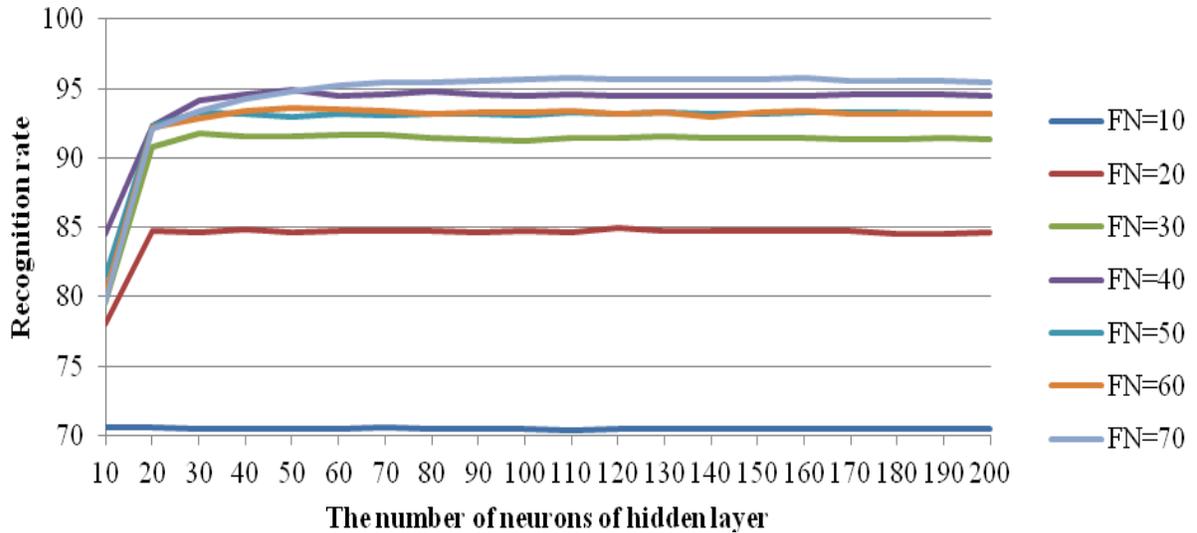


Figure 6. Accuracy of the ORL dataset classification using the Polak-Ribikre CGA for different the number of features and hidden neurons

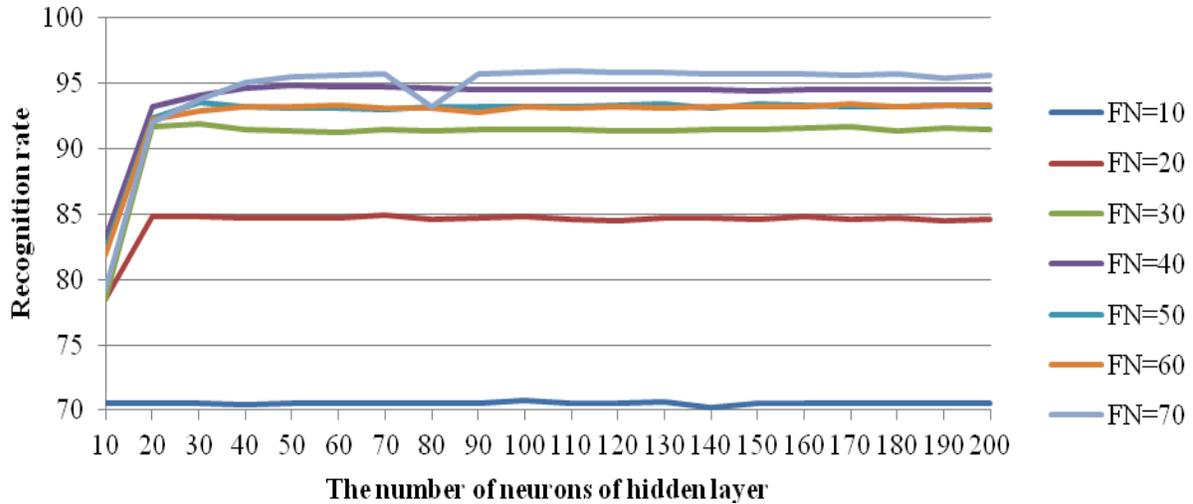


Figure 7. Accuracy of the ORL dataset classification using the Powell-Beale CGA for different the number of features and hidden neurons

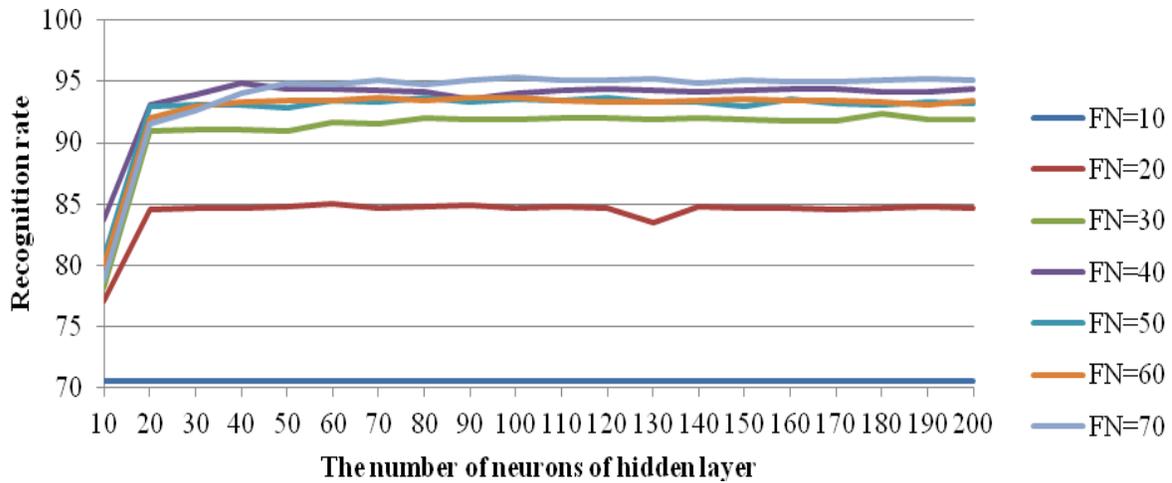


Figure 8. Accuracy of the ORL dataset classification using the scaled CGA for different the number of features and hidden neurons

Table 1. Comparison of the proposed methods and traditional method (simple BP)

Classification methods	simple BP [14]	Scaled CGA	Polak-Ribikre CGA	Fletcher-Reeves CGA	Powell-Beale CGA	Powell-Beale CGA and PCA
Correct classification rate	87.5%	95.37%	95.8%	95.85%	96.03%	96.5%

## 5. Conclusions

In order to enhance the performance of the NN-based face recognition systems, in this study four types of conjugate gradient algorithms (CGAs), namely, Fletcher-Reeves CGA, Polak-Ribikre CGA, Powell-Beale CGA, and scaled CGA have been proposed to classify the ORL database images. After decomposing the images and extracting their features, we trained the NN by applying the above methods. Among these methods, the Powell-Beale CGA gave the best results with 110 neurons in hidden layer and 70 features and could improve the recognition accuracy rated from 87.5% to 96.03% when applying the method to the ORL database and without any pre-processing step. Also, in order to improve the face recognition we used the PCA before using the NN for the Powell-Beale CGA as the best CGA. The PCA could create independence and informative data and increase the accuracy from 96.03% to 96.5%.

## 6. References

- [1] H. Azami, M. R. Mosavi and S. Sanei, "Classification of GPS satellites using improved back propagation training algorithms", *Wireless Personal Communications*, Springer-Verlog, DOI 10.1007/s11277-012-0844-7 (2012).
- [2] E. Gumus, N. Kilic, A. Sertbas and O. N. Ucan, "Evaluation of face recognition techniques using PCA, wavelets and SVM", *Expert Systems with Applications*, vol. 37, pp. 6404-6408 (2008).
- [3] M. R. M. Rizk and A. Taha, "Analysis of neural networks for face recognition systems with feature extraction to develop an eye localization based method", *International Conference on Electronics, Circuits and Systems*, pp. 847-850 (2002).
- [4] J. Daugman, "Face and gesture recognition: overview", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 19, no. 7, pp. 675-676 (1997).
- [5] L. Wiskott, J. Fellous, N. Kruger and C. Malsburg, "Face recognition by elastic bunch graph matching", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 19, no. 7, pp. 775-779 (1997).
- [6] M. J. Er, W. Chen and S. Wu, "High-speed face recognition based on discrete cosine transform and RBF neural networks", *IEEE Transactions on Neural Networks*, vol. 16, no. 3 (2005).
- [7] M. Firdaus, "Face recognition using neural networks", *International Conference on Intelligent System (ICIS)*, CD-ROM (2005).
- [8] H. Azami, S. Sanei and H. Alizadeh, "GPS GDOP Classification via Advanced Neural Network Training" *International Conference on Contemporary Issues in Computer and Information Sciences*, Brown Walker press, USA, pp. 315-320 (2012).
- [9] F. Paulin and A. Santhakumaran, "Classification of breast cancer by comparing back propagation training algorithms", *International Journal on Computer Science and Engineering*, vol. 3, no. 1, pp. 327-332 (2011).
- [10] M. H. Shaheed, "Performance analysis of 4 types of conjugate gradient algorithms in the nonlinear dynamic modelling of a TRMS using feedforward neural networks", *IEEE Conference on Systems, Man and Cybernetics*, pp. 5985-5990 (2004).
- [11] S. Sheel, T. Varshney and R. Varshney, "Accelerated learning in MLP using adaptive learning rate with momentum coefficient", *International Conference on Industrial and Information Systems*, pp. 307-310 (2007).

- [12] Z. Zakaria, N. A. M. Isa and S. A. Suandi, "A study on neural network training algorithm for multiface detection in static images", International Conference on Computer, Electrical, Systems Science, and Engineering, pp. 170-173 (2010).
- [13] M. F. Moller, "A scaled conjugate gradient algorithm for fast supervised learning", Neural Networks, vol. 6, pp. 525-533 (1993).
- [14] R. Bhati, S. Jain, D. K. Mishra and D. Bhati, "A comparative analysis of different neural networks for face recognition using principal component analysis and efficient variable learning rate", International Conference on Mathematical/Analytical Modelling and Computer Simulation, pp. 354-359 (2010).