

Face Recognition Based on Haar Wavelet Transform and Principal Component Analysis via Levenberg-Marquardt Backpropagation Neural Network

Mohammed Alwakeel

*Computer Science Department, Faculty of Computers and Information Technology
University of Tabuk, Tabuk, P.O. Box 1458, Tabuk 71341, Saudi Arabia
E-mail: malwakeel@ut.edu.sa*

Zyad Shaaban

*Information Technology Department, Faculty of Computers and Information Technology
University of Tabuk, Tabuk, P.O. Box 1458, Tabuk 71341, Saudi Arabia
E-mail: zshaaban@ut.edu.sa*

Abstract

In this paper, a new face recognition system based on Haar wavelet transform (HWT) and Principal Component Analysis (PCA) using Levenberg-Marquardt backpropagation (LMBP) neural network is presented. The image face is preprocessed and detected. The Haar wavelet is used to form the coefficient matrix for the detected face. The image feature vector is obtained by computing PCA for the coefficient matrix of DWT. A comparison between the proposed recognition system using DWT, PCA and Discrete Cosine Transform (DCT) is also made. The experimental results indicated that the image faces can be recognized by the proposed face recognition system effectively.

Keywords: Wavelet Transform, Principal Component Analysis, Neural Network, Levenberg-Marquardt backpropagation, Fast Fourier Transform, Discrete Cosine Transform

1. Introduction

Face recognition is one of most successful applications in computer vision and pattern recognition and the main objective of it is to recognize persons from pictures or video using a stored database of faces [1]. The building of face recognition system is a sophisticated problem because the faces has a lot of variations and may be located in a changed environment. Because of these reasons, the recognition of faces is a challenging problem due to the wide variety of illumination, facial expression and pose variations. In developing a face recognition system, we have to select suitable properties to represent a face under environmental changes. Face recognition is used in many applications such as human computer interaction, biometrics and security system [2].

In the recent years, wavelet analysis have generated a great interest in both theoretical and applied mathematics, and the wavelet transform in particular has proven to be an effective tool for e.g. data analysis, numerical analysis, and image processing [3].

The face recognition methods are categorized into holistic matching methods, feature-based matching methods, and hybrid methods [4]. Holistic matching methods use the whole face region as

the raw input to a recognition system. It is reported in [4] that one of the methods used in representations of the face region is Eigen faces, which are based on Principal Component Analysis (PCA). Using PCA, many face recognition techniques have been developed: Eigen faces, which use a nearest neighbor classifier; feature-line-based methods, which replace the point-to-point distance with the distance between a point and the feature line linking two stored sample points; Fisher faces which use linear/Fisher discriminant analysis (FLD/LDA); Bayesian methods, which use a probabilistic distance metric; and SVM methods, which use a support vector machine as the classifier. Utilizing higher order statistics, independent-component analysis (ICA) is argued to have more representative power than PCA, and hence may provide better recognition performance than PCA [4]. For the other types of recognition, it can be referred to reference [4].

It is reported in [5] that dimension reduction is as important as the class separation in applications like face recognition to make the face recognition system model based on the discrete cosine transform (DCT) computationally efficient. DCT helps to separate the image into parts (or spectral sub-bands) of differing importance (with respect to the image's visual quality). The DCT is similar to the discrete Fourier transform: it transforms a signal or image from the spatial domain to the frequency domain and represents an image as a sum of sinusoids of varying magnitude and frequencies. In proposed model [5] of face recognition dimension reduction is achieved firstly through decimation algorithm and then DCT is applied which exhibits large variance distribution in a small number of coefficients and much of the signal energy lies in low frequencies; these appear in the upper left corner of the DCT [5].

Neural networks (NN) have been applied in many applications such as: automotive, aerospace, banking, medical, robotics, electronic, and transportation [6]. NN is also used in classification of face images.

In this paper, a new face recognition system based on HWT and PCA using LMBP neural network is proposed.

This paper is arranged as follows. Section 2 gives an overview of HWT. Section 3 introduces PCA. The LMBP neural network classifier is given in sec. 4. The proposed face recognition system is depicted in Sec. 5. Section 6 discusses the results. Finally, a conclusion is given in Sec. 7.

2. Haar Wavelet Transform

In numerical analysis and functional analysis, a discrete wavelet transform (DWT) is any wavelet transform for which the wavelets are discretely sampled. As with other wavelet transforms, a key advantage it has over Fourier transforms is temporal resolution: it captures both frequency *and* location information. The first DWT was invented by the Hungarian mathematician Alfréd Haar. For an input represented by a list of 2^n numbers, the Haar wavelet transform may be considered to simply pair up input values, storing the difference and passing the sum. This process is repeated recursively, pairing up the sums to provide the next scale: finally resulting in $2^n - 1$ differences and one final sum. The Haar DWT illustrates the desirable properties of wavelets in general. First, it can be performed in $O(n)$ operations; second, it captures not only a notion of the frequency content of the input, by examining it at different scales, but also temporal content, i.e. the times at which these frequencies occur. Combined, these two properties make the Fast wavelet transform (FWT) an alternative to the conventional Fast Fourier Transform (FFT). The discrete wavelet transform has a huge number of applications in science, engineering, mathematics and computer science. Most notably, it is used for signal coding, to represent a discrete signal in a more redundant form, often as a preconditioning for data compression [7].

The Haar transformation technique is used [8] to form a wavelet since it is the simplest wavelet transformation method of all and can effectively serve our interests. In the Haar wavelet transformation method, low-pass filtering is conducted by averaging two adjacent pixel values, whereas the difference

between two adjacent pixel values is figured out for high-pass filtering. The Haar wavelet applies a pair of low-pass and high-pass filters to image decomposition first in image columns and then in image rows independently. As a result, it produces four sub-bands as the output of the first level Haar wavelet. The four sub-bands are LL1, HL1, LH1, and HH1. The low-frequency sub-band LL1 can be further decomposed into four sub-bands LL2, HL2, LH2, and HH2 at the next coarser scale. LL_i is a reduced resolution corresponding to the low frequency part of the image. The other three sub-bands HL_i , LH_i and HH_i are the high frequency parts in the vertical, horizontal, and diagonal directions, respectively [8].

The Haar transform cross-multiplies a function against the Haar wavelet with various shifts and stretches, like the Fourier transform cross-multiplies a function against a sine wave with two phases and many stretches. The Haar transform is derived from the Haar matrix. An example of a 4x4 Haar matrix is shown below [9]:

$$H_4 = \frac{1}{\sqrt{4}} \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & 1 & -1 & -1 \\ \sqrt{2} & -\sqrt{2} & 0 & 0 \\ 0 & 0 & \sqrt{2} & -\sqrt{2} \end{bmatrix} \quad (1)$$

3. Principal Component Analysis

The Principal Component Analysis (PCA) is one of the most successful methods that have been used in digital image processing and pattern recognition. The purpose of PCA is to reduce the large dimensionality of the data space for describing the data efficiently.

PCA is mathematically defined as an orthogonal linear transformation that transforms the data to a new coordinate system such that the greatest variance by any projection of the data comes to lie on the first coordinate (called the first principal component), the second greatest variance on the second coordinate, and so on. PCA is theoretically the optimum transform for given data in least square terms.

PCA can be used for dimensionality reduction in a data set by retaining those characteristics of the data set that contribute most to its variance, by keeping lower-order principal components and ignoring higher-order ones. Such low-order components often contain the "most important" aspects of the data. However, depending on the application this may not always be the case [10].

For a data matrix, X^T , with zero empirical mean (the empirical mean of the distribution has been subtracted from the data set), where each row represents a different repetition of the experiment, and each column gives the results from a particular probe, the PCA transformation is given by:

$$Y^T = X^T W = V \Sigma \quad (2)$$

where the matrix Σ is an m-by-n diagonal matrix with nonnegative real numbers on the diagonal and $W = \Sigma^{-1} V^T$ is the singular value decomposition (svd) of X. PCA has the distinction of being the optimal linear transformation for keeping the subspace that has largest variance. This advantage, however, comes at the price of greater computational requirement if compared, for example, to the discrete cosine transform [10]

4. LMBP Neural Network

The architecture of the back propagation algorithm is used [6]. Back propagation was created by generalizing the Widrow-Hoff learning rule to multiple layer network and non linear differentiable transfer function. Input vectors and corresponding target vectors are used to train a network until it can approximate a function, associate input vectors with specific output vectors, or classify input vectors in an appropriate way as defined in this study. Networks with biases, a sigmoid layer and a linear output layer are capable of approximating any function with a finite number of discontinuities. The back propagation algorithm consists of two paths; forward path and backward path. Forward path contain

creating a feed forward network, initializing weight, simulation and training the network. The network weights and biases are updated in backward path. For more details see reference [6].

The Levenberg-Marquardt algorithm was designed to approach second-order training speed without having to compute the Hessian matrix. When the performance function has the form of a sum of squares (as is typical in training feed forward networks), then the Hessian matrix can be approximated as

$$H = J^T J \quad (3)$$

and the gradient can be computed as

$$g = J^T e \quad (4)$$

Where J is the Jacobian matrix that contains first derivatives of the network errors with respect to the weights and biases, and e is a vector of network errors. The Jacobian matrix can be computed through a standard backpropagation technique (see [11]) that is much less complex than computing the Hessian matrix.

The Levenberg-Marquardt algorithm uses this approximation to the Hessian matrix in the following Newton-like update:

$$x_{k+1} = x_k - [J^T J + \mu I]^{-1} J^T e \quad (5)$$

When the scalar μ is zero, this is just Newton's method, using the approximate Hessian matrix. When μ is large, this becomes gradient descent with a small step size. Newton's method is faster and more accurate near an error minimum, so the aim is to shift towards Newton's method as quickly as possible. Thus, μ is decreased after each successful step (reduction in performance function) and is increased only when a tentative step would increase the performance function. In this way, the performance function will always be reduced at each iteration of the algorithm [6, 11].

5. The Proposed Face Recognition System

The proposed face recognition system is described as in the following stages:

Stage 1-Preprocessing: Preprocessing of faces images prior to face detection and classification is essential. The RGB face image is converted into grayscale image and then resized into 50 by 50 pixels.

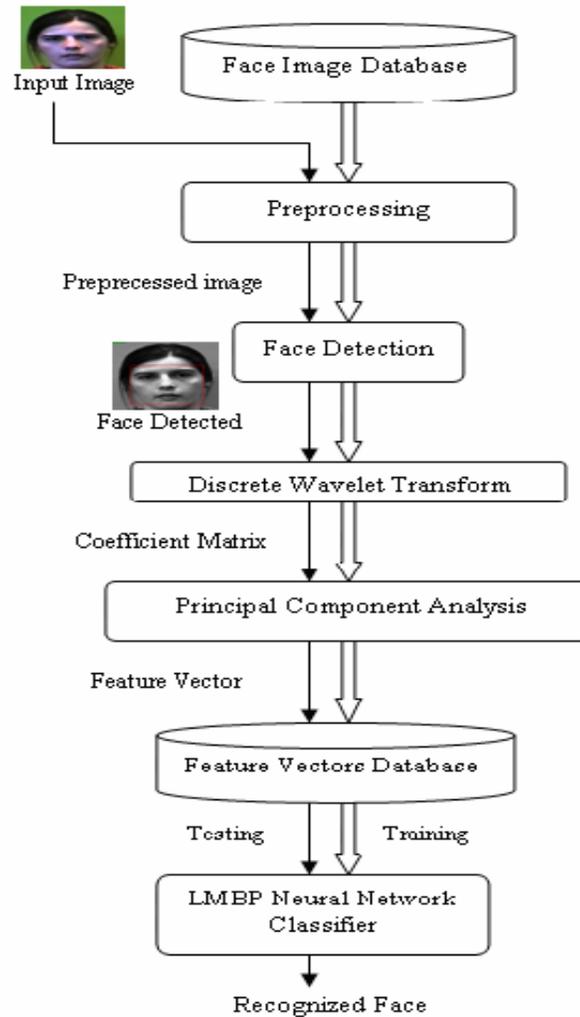
Stage 2- Face Detection: In this stage, the face is detected using local SMQT features and split up snow classifier proposed by Mikael Nilson et al.[12].

Stage 3- DWT Feature Extraction: This stage is extracting DWT features from the rectangular image obtained in stage 2 by applying Haar DWT on this image [7].

Stage 4- PCA Feature Extraction: In this stage, PCA is applied on DWT features obtained in the previous stage.

Stage 5- Recognition Process: This stage has two parts: the training part and the retrieving part. In the training part, LMBP neural network with seven input nodes and sixty output nodes is trained using the training features vectors. The optimal weights are obtained to be used in the recognition part. In this part, a new face is classified using the trained LMBP neural network.

All stages of the proposed system are illustrated in Fig. 1.

Figure 1: The proposed face recognition system

6. Experiments and Results

The face image database used in our experiments is Face 94 Directory [13], which consists of 153 subjects with 20 face images available for each subject. Some samples of images from this database is shown in Fig. 2. These face images varies in facial expression and illumination. In the experiments, all face images (180 x 200) in the database were resized to 80 x 60 and 60 x 40. The experiments were divided into two parts and have been done for 5 persons chosen from the dataset. In the first part, one sample per person is used in the training process and the number of testing samples per person is 19. The second part has five samples per person and 15 samples per subject in testing process.

Figure 2: Sample images of Face94/Dr Libor Spacek Faces Directories.

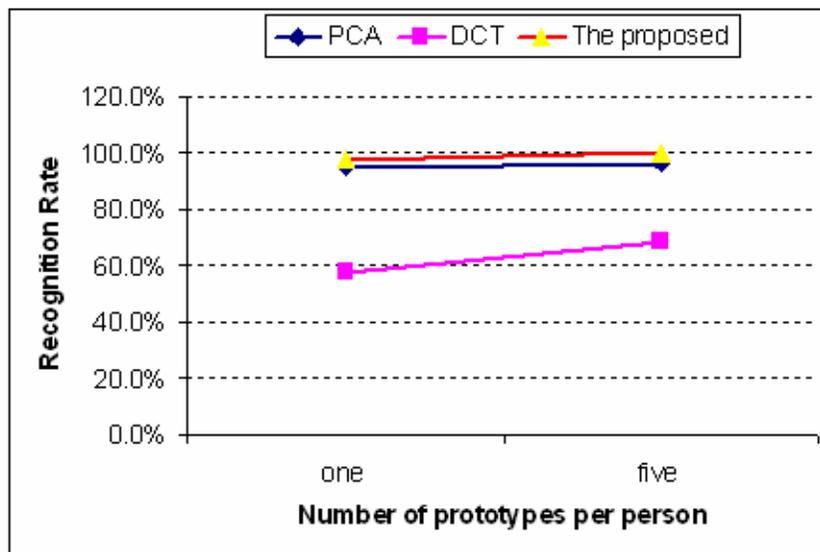


The proposed face recognition method shown in Fig. 1 is compared with, PCA and DCT using LMBP neural network. Three networks have been designed for the proposed, PCA and DCT feature extraction methods for the trained samples used in the experiment of the first part. And also three networks have been designed for the samples trained in the second part.

In the proposed method, the images are decomposed into 4 levels using Haar wavelets. The neural network for the proposed methods has 4 inputs and 5 outputs. In the neural network of PCA, the number of inputs is 40 and the number of outputs is 5. The neural network of DCT contains 50 inputs and 5 output nodes.

The recognition rate of the three methods is shown in Fig. 3. The results of the proposed method are higher than classical PCA and DCT as shown in Fig. 3.

Figure 3: The recognition rate of the proposed face recognition method, PCA and DCT.



7. Conclusion

The proposed face recognition system based on Haar wavelet and PCA using LMBP neural network has been introduced and evaluated. Using LMBP neural network, the proposed face recognition system gave the highest recognition rate in all the experiments, as compared with PCA and DCT.

References

- [1] Xiaoyang Tana, Songcan Chena, Zhi-Hua Zhou, Fuyan Zhang, 2006. "Face recognition from a single image per person: A survey, *Pattern Recognition*", Vol. 39, 2006, pp. 1725 – 1745.
- [2] Mohamed Rizon, Muhammad Firdaus Hashim, Puteh Saad, Sazali Yaacob, Mohd Rozailan Mamat, Ali Yeon Md Shakaff, Abdul Rahman Saad, Hazri Desa and M. Karthigayan, 2006. "Face Recognition using Eigenfaces and Neural Networks", *American Journal of Applied Sciences*, Vol. 2, No. 6, 2006, pp. 1872-1875.
- [3] R. Sanjeev Kunte, R. D. Sudhaker Samuel, 2007. "Wavelet Descriptors For Recognition Of Basic Symbols In Printed Kannada Text", *International Journal of Wavelets, Multiresolution and Information Processing*, Vol. 5, No. 2, 2007, pp. 351–367.
- [4] S. T. Gandhe, K. T. Talele, and A.G.Keskar, 2008, "Face Recognition Using Contour Matching", *IAENG International Journal of Computer Science*, Vol. 35, No. 2, 2008.
- [5] M. Almas Anjum, M. Younus Javed, and A. Basit, 2006, "A New Approach to Face Recognition Using Dual Dimension Reduction", *International Journal of Signal Processing* Vol. 2, No.1, 2006, pp1-6.
- [6] Hosseini Aria, E., J. Amini and M.R. Saradjian, 2003. "Back propagation neural network for classification of IRS-1D satellite images", *Proceedings of the Conference of High Resolution Mapping from Space*, 2003.
- [7] Stéphane G. Mallat, 1999. A wavelet tour of signal processing, *Academic Press*, 1999
- [8] Chin-Chen Chang, Jun-Chou Chuang and Yih-Shin Hu, 2004. "Similar Image Retrieval Based On Wavelet Transformation", *International Journal Of Wavelets, Multiresolution And Information Processing*, Vol. 2, No. 2, 2004, pp.111–120.
- [9] Charles K. Chui, 1992. An Introduction to Wavelets, Academic Press, 1992.
- [10] Jolliffe I.T., 2002. Principal Component Analysis, Series: Springer Series in Statistics, Springer, 2002.
- [11] Hagan, M.T., and M. Menhaj, 1999. "Training feed-forward networks with the Marquardt algorithm", *IEEE Transactions on Neural Networks*, Vol. 5, No. 6, 1999, pp. 989-993
- [12] Nilsson, M. and Nordberg, J. and Claesson, I., 2007. "Face Detection using Local SMQT Features and Split Up SNoW Classifier", *IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, Volume 2, No.2, 2007, pp. 589 - 592.
- [13] Dr Libor Spacek, Faces Directories, Faces 94 Directory, <http://cswww.essex.ac.uk/mv/allfaces/faces94.html>.