Face Recognition using DWT with HMM

Dr. Eyad I. Abbas
Department of Electrical Engineering, University of Technology/ Baghdad
Hameed R. Farhan
Department of Electrical Engineering, Engineering College, University of Kerbala/ Kerbala
Email: hrf64ha@yahoo.com

Received on: 19/6/2011 & Accepted on: 3/11/2011

ABSTRACT

This paper presents an efficient face recognition system based on Hidden Markov Model (HMM) and the simplest type “Haar” of the Discrete Wavelet Transform (DWT). The one dimensional ergodic HMM with Gaussian outputs, which represent the simplest and robust type of HMM, is used in the proposed work. A novel method is introduced for selecting the training images implemented by choosing the images that have the odd identifying numbers from the database. Some of these images are replaced according to the trial-and-error results. The proposed work achieves the maximum recognition rate (100%), where the experiments are carried out on the ORL face database.

Keywords: Discrete Wavelet Transform, Hidden Markov Model, Face recognition.

INTRODUCTION

Face recognition has received significant attention during the last 20 years. It is a biometric technique, which uses computer software to determine the identity of the individual, and is attractive for numerous applications including visual surveillance and security. Face recognition using Hidden Markov Model (HMM) is based on the overall probability of statistical methods. HMM is a parameter expressed as a probabilistic model which describes the statistical properties of stochastic processes [1]. Excellent surveys on HMM have been introduced to face recognition technique. Most of the differences between these works are restricted in the nature of feature

https://doi.org/10.30684/etj.30.1.12
2412-0758/University of Technology-Iraq, Baghdad, Iraq
This is an open access article under the CC BY 4.0 license http://creativecommons.org/licenses/by/4.0
Face Recognition using DWT with HMM

In this paper, the system composed of Haar wavelet transform and the ergodic HMM is described and a comparison is made with other works.

DISCRETE WAVELET TRANSFORM

The DWT decomposes a signal into the coarse approximation and detailed information. It employs two sets of functions, called scaling functions and wavelet functions, which are associated with low pass $h[n]$ and high-pass $g[n]$ filters, respectively. The decomposition process into different frequency bands is obtained by the filter banks composed of the successive high-pass and low-pass filters. At each level of the decomposition, the signal is split into high frequency and low frequency components, and the low-frequency components can be further decomposed until the desired resolution is reached [10].

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 [11]. Figure 1 shows an example of Haar wavelet.

The Haar wavelet’s mother function $\psi(t)$ can be described as:

$$
\psi(t) = \begin{cases} 
1 & 0 \leq t < 1/2, \\
-1 & 1/2 \leq t < 1, \\
0 & \text{otherwise.}
\end{cases}
$$

and its scaling function $\varphi(t)$ can be described as

$$
\varphi(t) = \begin{cases} 
1 & 0 \leq t < 1, \\
0 & \text{otherwise.}
\end{cases}
$$

The procedure of 1D-DWT starts with passing $x[n]$ through a half band digital low-pass filter with impulse response $h[n]$. Filtering a signal corresponds to the mathematical operation of convolution of the signal with the impulse response of the filter [12].

The convolution operation in discrete time is defined as follows [12]:

$$
[n] = x[n] * h[n] 
$$

The signal is also decomposed simultaneously using a high-pass filter $g[n]$.

$$
[n] = x[n] * g[n] 
$$
The outputs give the detail coefficients (from the high-pass filter) and approximation coefficients (from the low-pass). The filter outputs are then subsampled by 2 and this procedure can mathematically be expressed as follows:

\[ y_l[n] = \sum_{k=-\infty}^{\infty} x[k] h[2n - k] \]  
\[ y_h[n] = \sum_{k=-\infty}^{\infty} x[k] g[2n - k] \]

where \( y_l[n] \) and \( y_h[n] \) are the outputs of the low-pass and high-pass filters, respectively, after subsampling by 2. The decomposition process is repeated to further increase the frequency resolution and the approximation coefficients decomposed with high and low pass filters and then down-sampled [13]. Figure 2 shows 1-level of filter analysis.

Images are treated as two-dimensional signals, they change horizontally and vertically, thus a 2D wavelet analysis must be used for images. A 2D transform can be accomplished by performing two separate one-dimensional transforms. First the image is filtered along horizontal-direction using low pass and high pass analysis filters and decimated by two. Because of decimation the total size of the transformed image is same as the original image. Then it is followed by filtering the sub-image along vertical-direction and decimated by two. Finally split the image into four sub-bands denoted by LL, LH, HL and HH, after one level of decomposition [14].

At every level, four sub-images are obtained; the approximation (LL), the vertical detail (LH), the horizontal detail (HL) and the diagonal detail (HH) [15]. Figure 3 illustrates a 3-level structure of 2D-DWT.

**HIDDEN MARKOV MODEL**

Hidden Markov Model (HMM) is a stochastic model which provides a high level of flexibility for modeling the structure of an observation sequence. It consists of a number of non-observable (hidden) states and an observable sequence, generated by the individual hidden states.

An HMM is characterized by the following five elements [16]:

- \( N \), the number of states in the model. The individual states are denoted as \( S = \{ S_1, S_2, \ldots, S_N \} \), and the state at time \( t \) as \( q_t \).
- \( M \), the number of distinct observation symbols per state, i.e., the discrete alphabet size (also called the codebook of the model). The individual symbols are denoted as \( V = \{ v_1, v_2, \ldots, v_M \} \).
- \( A = \{ a_{ij} \} \), \( N \) by \( N \) state transition probability matrix, where:
  \[ a_{ij} = P(q_{t+1} = s_j | q_t = s_i) , \ 1 \leq i, j \leq N \]
  with the constraint, \( 0 \leq a_{ij} \leq 1 \), and \( \sum_{j=1}^{N} a_{ij} = 1 \), \( 1 \leq i \leq N \)

State transition probabilities do not change over time and these probabilities are independent of the observations.

- \( B = \{ b_j(k) \} \), \( N \) by \( M \) emission (observation) probability matrix, indicating the probability of a specified symbol being emitted given that the system is in a particular state \( j \), i.e.,
Face Recognition using DWT with HMM

Where, \( 1 \leq j \leq N \), \( 1 \leq k \leq M \) and \( O_t \) is the observation symbol at time \( t \) and the observations are chosen from a finite alphabet with the condition:

\[
\sum_{j=1}^{N} b_j(k) = 1, \quad 1 \leq k \leq M
\]

- \( \pi = \{ \pi_i \}, \) N by 1 initial state probability vector, i.e.,

\[
\pi_i = p(q_1 = s_i), \quad 1 \leq i \leq N
\]

Where \( \pi_i \geq 0 \) and

\[
\sum_{i=1}^{N} \pi_i = 1
\]

Given appropriate values of \( N, M, A, B, \) and \( \pi \), HMM can be used as a generator to give an observation sequence \( O = O_1O_2 \ldots O_T \), where each observation \( O_t \) is one of the symbols from \( \{v_1, v_2, \ldots, v_M\} \), and \( T \) is the number of observations in the sequence, i.e. \( T = \text{length of the observation sequence} \). The compact notation to indicate the complete parameter set of the model is represented by:

\[
\lambda = (A, B, \pi)
\]

For a continuous HMM, \( b_i(O_t)\)’s are defined by a mixture of some parametric probability density functions. The most common parametric pdf used in continuous HMM, is the mixture Gaussian density, where

\[
b_i(O_t) = \sum_{j=1}^{M} c_{ij} b_{ij}(O_t), \quad i = 1, 2, \ldots, N
\]

where \( M \) is the number of mixture components in state \( i \), \( c_{ij} \) is the mixture coefficient for the \( j^{th} \) mixture component in state \( i \), and satisfies the constraints \( c_{ij} \geq 0 \), and

\[
\sum_{j=1}^{M} c_{ij} = 1, \quad i = 1, 2, \ldots, N
\]

\( b_{ij}(O_t) \) is a \( D \) dimensional multivariate Gaussian density with mean vector \( \mu_{ij} \) and covariance matrix \( \Sigma_{ij} \).

The probability density function (pdf) usually written in the form:

\[
b_{ij}(O_t) = \mathcal{G}(o_t, \mu_{ij}, \Sigma_{ij})
\]

Where \( \mathcal{G} \) is a Gaussian pdf, and

\[
b_{ij}(o_t) = \frac{1}{(2\pi)^{D/2} |\Sigma_{ij}|^{1/2}} e^{-\frac{1}{2}(o_t-\mu_{ij})^T \Sigma_{ij}^{-1}(o_t-\mu_{ij})}
\]

The training of the HMM parameters \( \lambda = (A, B, \pi) \), given an observation sequence \( O = O_1O_2 \ldots O_T \), is usually performed using the standard Baum-Welch algorithm [18] to determine the model parameters that maximize the probability \( P(O|\lambda) \).

**THE DWT-HMM APPROACH**

**Feature Extraction**

The entire face image is decomposed using one of the Daubechies family “db1” (Haar wavelet) to the approximation and detail coefficients. Again the approximation

\[ b_j(k) = p(O_t = v_k | q_t = s_j) \]
coefficient is decomposed to produce a new approximation and detail coefficients. Figure 4 shows an example of DWT image decomposition.

In order to generate the observation vectors, the matrix of the approximation part (LL) at the second level of DWT is coded using regular grid between the minimum and the maximum values (1 and 256) of the input matrix. The coded matrix is divided into overlapping vertical strips of height $H$ and width $w_s$, where $H$ equals the height of the decomposed image. The amount of overlap between consecutive strips is $p_s$, and the number of overlapping strips extracted from each image is $N_s$. Where $N_s$ is calculated as follows:

$$N_s = \frac{W - w_s}{w_s - p_s} + 1$$  \hspace{1cm} (20)

Each vertical strip is subsequently segmented horizontally into overlapping blocks of width $w_g$ and height $h_b$. The amount of overlap between consecutive blocks is $p_b$ and the number of overlapping blocks extracted from each strip is $N_b$.

$$N_b = \frac{H - h_b}{h_b - p_b} + 1$$  \hspace{1cm} (21)

Figure 5 shows a process of generation sub-images (blocks).

The extracted blocks are arranged consecutively and converted to vector form (column-wise), such vectors are called the observation vectors. Each image is represented by one sequence of observation vectors starting from the top left block and ending in the down right block.

**HMM IMPLEMENTATION**

A simple and robust type implemented by one dimensional (1D) ergodic HMM with Gaussian outputs is used in the proposed work. Figure 6 shows a state diagram of 5-state 1D ergodic HMM. The state transition probability matrix is formed as follows:

$$A = \{a_{ij}\} = \begin{bmatrix}
a_{11} & a_{12} & a_{13} & a_{14} & a_{15} \\
a_{21} & a_{22} & a_{23} & a_{24} & a_{25} \\
a_{31} & a_{32} & a_{33} & a_{34} & a_{35} \\
a_{41} & a_{42} & a_{43} & a_{44} & a_{45} \\
a_{51} & a_{52} & a_{53} & a_{54} & a_{55} 
\end{bmatrix}$$

For the training scheme, number of images is selected from each subject, the images that have the odd identifying numbers are chosen to form the training set. Some of these images are replaced according to the trial-and-error results. Each of these images is processed to extract features to form observation vectors. A preliminary model $\lambda$ is constructed specifying the initial probabilities of the states, their allowed transitions and the dimension of the observation vector. This model is re-estimated using the Baum Welch algorithm until it converges to a maximum likelihood value. The flowchart of the training process is shown in Figure 7.

The test image is processed in the same manner as in the training process, so that a resultant series of observation vectors is generated. The Forward algorithm [18] is used to get the likelihood of each subject’s model with the observation vectors of the test image. The subject with the highest likelihood is selected as the most likely person. Figure 8 shows a schematic diagram of the recognition process.
EXPERIMENTAL RESULTS

AT&T database [19] (already known as ORL) which contains 10 face images for each of 40 persons is used in the proposed work. The size of each image is 112x92 and the corresponding reduced size \((HxW)\) after the second level of DWT is 28x23.

The system was trained by \(K\) facial images per person, while the remaining \((10–K)\) images were used for testing. The maximum recognition rate (100\%) is achieved when the following are used:

- Five images per person for training.
- Block size of \((5\times3), (5\times4)\) and \((5\times5)\).
- The overlap between consecutive blocks is one pixel less than the width or height of the block.

Table 1 illustrates the effect of varying the number of training images, while Table 2 shows the effect of varying the block size.

Table 3 shows a comparison with other works which use the wavelet transform and 5-state of HMM, in addition to five training images for each person in ORL database. Figure 9 shows a sample of ORL database.

The experiments are done in Matlab (matrix laboratory) environment using the wavelet toolbox [20] for wavelet transform, and BNT toolbox [21] for Hidden Markov Model. The version of Matlab software is 7.9, and it is used on Windows 7 (32-bit) operating system.

DISCUSSION

It is clear from table (1,2), the recognition rate gets better when increasing the number of training images and the block size of the proposed work. Although the work introduced in [5] presents the lowest recognition time, but it occupies a comparatively more memory size due to the method used in the recognition process, where the Euclidean distances between the unknown vectors and the observation vectors of the training images were computed. Therefore, all training images after training have to be kept. In contrast, in the proposed work, only the parameters of the models are kept and there is no need to the training images after training and therefore, a lower memory size is used.

CONCLUSIONS

The segmentation of face images into vertical strips gives a good result. Therefore, the feature extraction method is independent of the facial regions (forehead, eyes, nose, mouth and chin), because the features are converted to a sequence of vectors in the training process and the same manner is used in the recognition process.

The 1D ergodic HMM has no constraints for the transition between any two states in the model. In other words, the state transition probability matrix \(A\) has no zeros in its structure. This is very useful in avoiding divide by zero in the algorithms used in training and testing the HMMs. The testing algorithm used in this work is the Forward algorithm only, which is the simplest one and the results appear that this algorithm has high recognition efficiency.
REFERENCES


Face Recognition using DWT with HMM


Table 1 Effect of varying the number of training images per person.

<table>
<thead>
<tr>
<th>Block size</th>
<th>Length of block (bytes)</th>
<th>No. of images</th>
<th>Recognition rate (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>32X3</td>
<td>546</td>
<td>10</td>
<td>95</td>
</tr>
<tr>
<td>4X5</td>
<td>576</td>
<td>6</td>
<td>97</td>
</tr>
<tr>
<td>4X4</td>
<td>576</td>
<td>7</td>
<td>96.5</td>
</tr>
<tr>
<td>5X2</td>
<td>540</td>
<td>9</td>
<td>100</td>
</tr>
</tbody>
</table>

Table 2 Effect of varying block size

<table>
<thead>
<tr>
<th>No. of training images</th>
<th>No. of classes</th>
<th>Recognition rate (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>92</td>
<td>74.44</td>
</tr>
<tr>
<td>2</td>
<td>61</td>
<td>80.93</td>
</tr>
<tr>
<td>3</td>
<td>9</td>
<td>95.78</td>
</tr>
<tr>
<td>4</td>
<td>6</td>
<td>97.5</td>
</tr>
<tr>
<td>5</td>
<td>0</td>
<td>100</td>
</tr>
</tbody>
</table>
Table (3) Comparative results with other works

<table>
<thead>
<tr>
<th>Reference number</th>
<th>Year</th>
<th>Training time per model (second)</th>
<th>Recognition time (second)</th>
<th>Recognition rate (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>4</td>
<td>2003</td>
<td>N/A</td>
<td>N/A</td>
<td>100</td>
</tr>
<tr>
<td>5</td>
<td>2003</td>
<td>1.13</td>
<td>0.3</td>
<td>100</td>
</tr>
<tr>
<td>6</td>
<td>2003</td>
<td>0.68</td>
<td>1.55</td>
<td>97.5</td>
</tr>
<tr>
<td>7</td>
<td>2007</td>
<td>N/A</td>
<td>N/A</td>
<td>98.5</td>
</tr>
<tr>
<td>8</td>
<td>2008</td>
<td>4.31</td>
<td>3.45</td>
<td>97</td>
</tr>
<tr>
<td>9</td>
<td>2009</td>
<td>N/A</td>
<td>N/A</td>
<td>96.5</td>
</tr>
<tr>
<td>The proposed work</td>
<td>2011</td>
<td>1.35</td>
<td>0.8</td>
<td>100</td>
</tr>
</tbody>
</table>

Figure (1) Haar wavelet example
Figure 2 1-level filter analyses

Figure (3) a structure of 3-level 2D-DWT

Figure (4) An example of 2-level DWT image decomposition.
Figure (5) Sub-images generation method

Figure (6) State diagram of 5-state ergodic HMM
Figure (7) Flowchart of the training process

Figure (8) Block diagram of the recognition process
Figure (9) Sample of ORL database