Research Article

A Statistical Multiresolution Approach for Face Recognition Using Structural Hidden Markov Models

P. Nicholl,[1] A. Amira,[2] D. Bouchaffra,[3] and R. H. Perrott[1]

[1] *School of Electronics, Electrical Engineering and Computer Science, Queens University, Belfast BT7 1NN, UK*

[2] *Electrical and Computer Engineering, School of Engineering and Design, Brunel University, London UB8 3PH, UK*

[3] *Department of Mathematics and Computer Science, Grambling State University, Carver Hall, Room 281-C, P.O. Box 1191, LA, USA*

Correspondence should be addressed to P. Nicholl, p.nicholl@qub.ac.uk

This paper introduces a novel methodology that combines the multiresolution feature of the discrete wavelet transform (DWT) with the local interactions of the facial structures expressed through the structural hidden Markov model (SHMM). A range of wavelet filters such as Haar, biorthogonal 9/7, and Coiflet, as well as Gabor, have been implemented in order to search for the best performance. SHMMs perform a thorough probabilistic analysis of any sequential pattern by revealing both its inner and outer structures simultaneously. Unlike traditional HMMs, the SHMMs do not perform the state conditional independence of the visible observation sequence assumption. This is achieved via the concept of local structures introduced by the SHMMs. Therefore, the long-range dependency problem inherent to traditional HMMs has been drastically reduced. SHMMs have not previously been applied to the problem of face identification. The results reported in this application have shown that SHMM outperforms the traditional hidden Markov model with a 73% increase in accuracy.

1. INTRODUCTION

With the current perceived world security situation, governments, as well as businesses, require reliable methods to accurately identify individuals, without overly infringing on rights to privacy or requiring significant compliance on the part of the individual being recognized. Person recognition systems based on biometrics have been used for a significant period for law enforcement and secure access. Both fingerprint and iris recognition systems are proven as reliable techniques; however, the method of capture for both limits their versatility [1]. Although face recognition technology is not as mature as other biometric verification methods, it is the subject of intensive research and may provide an acceptable solution to some of the problems mentioned. As it is the primary method used by humans to recognize each other, and because an individual's face image is already stored in numerous locations, it is seen as a more acceptable method of automatic recognition [2]. A robust face recognition solution has many potential applications. Business organizations are aware of the ever-increasing need for security,

this is mandated not only by both their own desire to protect property and processes, but also by their workforce's increasing demands for workplace safety and security [3]. Local law enforcement agencies have been using face recognition for rapid identification of individuals suspected of committing crimes. They have also used the technology to control access at large public gatherings such as sports events, where there are often watchlists of known trouble-makers. Similarly, face recognition has been deployed in national ports-of-entry, making it easier to prevent terrorists from entering a country.

However, face recognition is a more complicated task than fingerprint or iris recognition. This is mostly due to the increased variability of acquired face images. Whilst controls can sometimes be placed on face image acquisition, for example, in the case of passport photographs, in many cases this is not possible. Variation in pose, expression, illumination, and partial occlusion of the face therefore become nontrivial issues that have to be addressed. Even when strict controls are placed on image capture, variation over time of an individual's appearance is unavoidable, both in the short

term (e.g., hairstyle change) and in the long term (aging process). These issues all increase the complexity of the recognition task [4].

A multitude of techniques have been applied to face recognition and they can be separated into two categories: geometric feature matching and template matching. Geometric feature matching involves segmenting the distinctive features of the face, eyes, nose, mouth, and so on, and extracting descriptive information about them such as their widths and heights. Ratios between these measures can then be stored for each person and compared with those from known individuals [5]. Template matching is a holistic approach to face recognition. Each face is treated as a two-dimensional array of intensity values, which is compared with other facial arrays. Techniques of this type include principal component analysis (PCA) [6], where the variance among a set of face images is represented by a number of eigenfaces. The face images, encoded as weight vectors of the eigenfaces, can be compared using a suitable distance measure [7, 8]. In independent component analysis (ICA), faces are assumed to be linear mixtures of some unknown latent variables. The latent variables are assumed non-Gaussian and mutually independent, and they are called the independent components of the observed data [9]. In neural network models (NNMs), the system is supplied with a set of training images along with correct classification, thus allowing the neural network to ascertain a weighting system to determine which areas of an image are deemed most important [10].

Hidden Markov models (HMMs) [11], which have been used successfully in speech recognition for a number of decades, are now being applied to face recognition. Samaria and Young used image pixel values to build a top-down model of a face using HMMs. Nefian and Hayes [12] modified the approach by using discrete cosine transform (DCT) coefficients to form observation vectors. Bai and Shen [13] used discrete wavelet transform (DWT) [14] coefficients taken from overlapping image subwindows taken from the entire face image, whereas Bicego et al. [15] used DWT coefficients of subwindows generated by a raster scan of the image.

As HMMs are one dimensional in nature, a variety of approaches have been adopted to try to represent the two-dimensional structure of face images. These include the 1D discrete HMM (1D-DHMM) approach [16], which models a face image using two standard HMMs, one for observations in the vertical direction and one for the horizontal direction. Another approach is the pseudo-2D HMM (2D-PHMM) [17], which is a 1D HMM, composed of super states to model the sequence of columns in the image, in which each super state is a 1D-HMM, itself modeling the blocks within the columns. An alternative approach is the low-complexity 2D-HMM (LC 2D-HMM) [18], which consists of a rectangular constellation of states, where both vertical and horizontal transitions are supported. The complexity of the LC 2D-HMM is considerably lower than that of the 2D-PHMM and the two-dimensional HMM (2D-HMM), however, recognition accuracy is lower as a result. The hierarchical hidden Markov models (HHMMs) introduced in [19] and applied in video-content analysis [20] are capable of modeling the complex multiscale structure which appears in many natural sequences. However, the original HHMM algorithm is rather complicated since it takes $O(T^3)$ time, where T is the length of the sequence, making it impractical for many domains.

Although HMMs are effective in modeling statistical information [21], they are not suited to unfold the sequence of local structures that constitutes the entire pattern. In other words, the state conditional independence assumption inherent to traditional HMMs makes these models unable to capture long-range dependencies. They are therefore not optimal for handling structural patterns such as the human face. Humans distinguish facial regions in part due to our ability to cluster the entire face with respect to some features such as colors, textures, and shapes. These well-organized clusters sensed by the human's brain are the facial regions such as lips, hair, forehead, eyes, and so on. They are all composed of similar symbols that unfold their global appearances. One recently developed model for pattern recognition is the structural hidden Markov model (SHMM) [22, 23]. To avoid the complexity problem inherent to the determination of the higher level states, the SHMM provides a way to explicitly control them via an unsupervised clustering process. This capability is offered through an equivalence relation built in the visible observation sequence space. The SHMMs approach allows both the structural and the statistical properties of a pattern to be represented within the same probabilistic framework. This approach also allows the user to weight substantially the local structures within a pattern that are difficult to disguise. This provides an SHMM recognizer with a higher degree of robustness. Indeed, SHMMs have been shown to outperform HMMs in a number of applications including handwriting recognition [22], but have yet to be applied to face recognition. However, SHMMs are well-suited to model the inner and outer structures of any sequential pattern (such as a face) simultaneously.

As well as being used in conjunction with HMMs for face recognition, DWT has been coupled with other techniques. Its ability to localize information in terms of both frequency and space (when applied to images) makes it an invaluable tool for image processing. In [24], the authors use it to extract low frequency features, reinforced using linear discriminant analysis (LDA). In [25], wavelet packet analysis is used to extract rotation invariant features and in [5], the authors use it to identify and extract the significant structures of the face, enabling statistical measures to be calculated as a result. DWT has also been used for feature extraction in PCA-based approaches [26, 27]. The Gabor wavelet in particular has been used extensively for face recognition applications. In [28], it is used along with kernel PCA to recognize faces where a large degree of rotation is present, whereas in [29], AdaBoost is employed to select the most discriminant Gabor features.

The objective of the work presented in this paper is to develop a hybrid approach for face identification using SHMMs for the first time. The effect of using DWT for feature extraction is also investigated, and the influence of wavelet type is analyzed.

The rest of this paper is organized as follows. Section 2 describes face recognition using an HMM/DWT approach.

Section 3 proposes the use of SHMM for face recognition. Section 4 describes the experiments that were carried out and presents and analyzes the results obtained. Section 5 contains concluding remarks.

2. RECOGNITION USING WAVELET/HMM

2.1. Mathematical background

(1) Discrete wavelet transform

In the last decade, DWT has been recognized as a powerful tool in a wide range of applications, including image/video processing, numerical analysis, and telecommunication. The advantage of DWT over existing transforms such as discrete Fourier transform (DFT) and DCT is that DWT performs a multiresolution analysis of a signal with localization in both time and frequency [14, 30]. In addition to this, functions with discontinuities and functions with sharp spikes require fewer wavelet basis vectors in the wavelet domain than sinecosine basis vectors to achieve a comparable approximation. DWT operates by convolving the target function with wavelet kernels to obtain wavelet coefficients representing the contributions of wavelets in the function at different scales and orientations.

DWT can be implemented as a set of filter banks, comprising a high-pass and low-pass filters. In standard wavelet decomposition, the output from the low-pass filter can then be decomposed further, with the process continuing recursively in this manner. DWT can be mathematically expressed by

$$\mathrm{DWT}_{\mathrm{x}(n)} = \begin{cases} d_{j,k} = \sum x(n) h_j^* (n - 2^j k), \\ a_{j,k} = \sum x(n) g_j^* (n - 2^j k). \end{cases} \quad (1)$$

The coefficients $d_{j,k}$ refer to the detail components in signal $x(n)$ and correspond to the wavelet function, whereas $a_{j,k}$ refer to the approximation components in the signal. The functions $h(n)$ and $g(n)$ in the equation represent the coefficients of the high-pass and low-pass filters, respectively, whilst parameters j and k refer to wavelet scale and translation factors. Figure 1 illustrates DWT schematically.

For the case of images, the one-dimensional DWT can be readily extended to two dimensions. In standard twodimensional wavelet decomposition, the image rows are fully decomposed, with the output being fully decomposed columnwise. In nonstandard wavelet decomposition, all the rows are decomposed by one decomposition level followed by one decomposition level of the columns.

The decomposition continues by decomposing the low resolution output from each step, until the image is fully decomposed. Figure 2 illustrates the effect of applying the nonstandard wavelet transform to an image from the AT&T Database of Faces [31]. The wavelet filter used, number of levels of decomposition applied, and quadrants chosen for feature extraction are dependent upon the particular application. For the experiments described in this paper, the nonstandard DWT is used, which allows for the selection of areas with similar resolutions in both horizontal and vertical directions to take place for feature extraction. For further information on DWT, see [32].

(2) Gabor wavelets

Gabor wavelets are similar to DWT, but their usage is different. A Gabor wavelet is convolved with an image either locally at selected points in the image, or globally. The output reveals the contribution that a frequency is making to the image at each location. A Gabor wavelet $\psi_{u,v}(z)$ is defined as [28]

$$\psi_{u,v}(z) = \frac{||k_{u,v}||^2}{\sigma^2} e^{-(||k_{u,v}||^2 ||z||^2)/2\sigma^2} [e^{ik_{u,v}z} - e^{-\sigma^2/2}], \quad (2)$$

where $z = (x, y)$ is the point with the horizontal coordinate x and the vertical coordinate y. The parameters u and v define the orientation and scale of the Gabor kernel, $||\cdot||$ defines the norm operator, and σ is related to the standard deviation of the Gaussian window in the kernel and determines the ratio of the Gaussian window width to the wavelength. The wave vector k_u, v is defined as follows:

$$k_{u,v} = k_v e^{i\phi_u}, \quad (3)$$

where $k_v = k_{\max}/f^v$ and $\phi_u = \pi \mu/n$ if n different orientations have been chosen. $k_{\max}$ is the maximum frequency, and f^v is the spatial frequency between kernels in the frequency domain.

(3) Hidden markov models

HMMs are used to characterize the statistical properties of a signal [11]. They have been used in speech recognition applications for many years and are now being applied to face recognition. An HMM consists of a number of nonobservable states and an observable sequence, generated by the individual hidden states. Figure 3 illustrates the structure of a simple HMM.

HMMs are defined by the following elements.

(i) N is the number of hidden states in the model.
(ii) M is the number of different observation symbols.
(iii) $S = \{S_1, S_2, \ldots, S_N\}$ is the finite set of possible hidden states. The state of the model at time t is given by $q_t \in S, 1 \le t \le T$, where T is the length of the observation sequence.
(iv) $A = \{a_{ij}\}$ is the state transition probability matrix, where

$$a_{ij} = P[q_{t+1} = S_j \mid q_t = S_i], \quad 1 \le i, j \le N, \quad (4)$$

with

$$0 \le a_{i,j} \le 1,$$
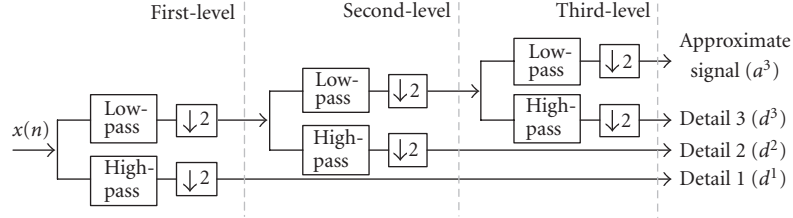$$\sum_{j=1}^{N} a_{ij} = 1, \quad 1 \le i \le N. \quad (5)$$

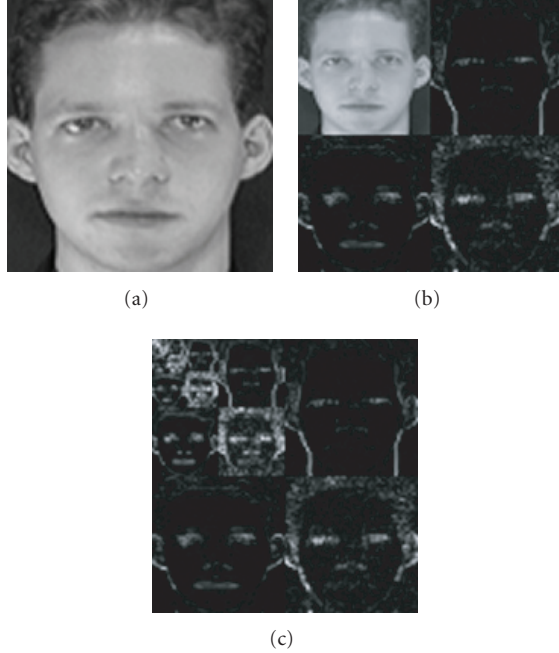FIGURE 1: A three-level wavelet decomposition system.

FIGURE 2: Wavelet transform of image: (a) original image, (b) 1-level Haar decomposition, (c) complete decomposition.
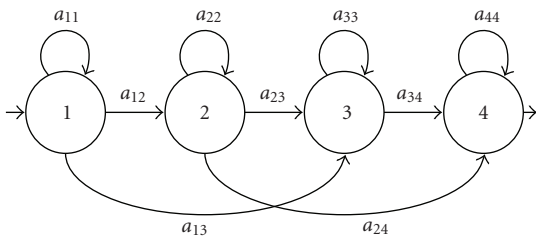

FIGURE 3: A simple left-right HMM.

(i) $B = \{b_j(k)\}$ is the emission probability matrix, indicating the probability of a specified symbol being emitted given that the system is in a particular state, that is,

$$b_j(k) = P[O_t = k \mid q_t = S_j] \qquad (6)$$

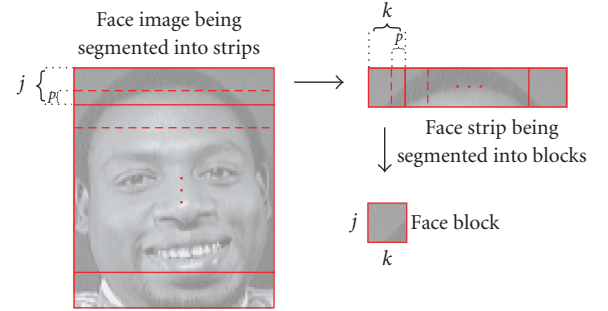with $1 \leq j \leq N$ and O_t is the observation symbol at time t.

FIGURE 4: An illustration showing the creation of the block sequence.

(ii) $\Pi = \{\pi_i\}$ is the initial state probability distribution, that is,

$$\pi_i = P[q_1 = S_i], \quad 1 \leq i \leq N, \qquad (7)$$

with $\pi_i \geq 0$ and $\sum_{i=1}^{N} \pi_i = 1$.

An HMM can therefore be succinctly defined by the triplet

$$\lambda = (A, B, \Pi). \qquad (8)$$

HMMs are typically used to address three unique problems [11].

 (i) *Evaluation.* Given a model λ and a sequence of observations O, what is the probability that O was generated by model λ, that is, $P(O \mid \lambda)$.

(ii) *Decoding.* Given a model λ and a sequence of observations O, what is the hidden state sequence q^* most likely to have produced O, that is, $q^* = \arg\max_q [P(q \mid \lambda, O)]$.

(iii) *Parameter estimation.* Given an observation sequence O, what model λ is most likely to have produced O.

For further information on HMMs, see [11].

2.2. Recognition process

(1) Training

The first phase of identification is feature extraction. In the cases where DWT is used, each face image is divided into overlapping horizontal strips of height j pixels where the strips overlap by p pixels. Each horizontal strip is subsequently segmented vertically into blocks of width k pixels,

with overlap of p. This is illustrated in **Figure 4**. For an image of width w and height h, there will be approximately $(((h/(j-p))+1)*(w/(k-p))+1)$ blocks.

Each block then undergoes wavelet decomposition, producing an average image and a sequence of detail images. This can be shown as $[a_J, \{d_j^1, d_j^2, d_j^3\}_{j=1,...,J}]$ where a_J refers to the approximation image at the Jth scale and d_j^k is the detail image at scale j and orientation k. For the work described, 4-level wavelet decomposition is employed, producing a vector with one average image and twelve detail images. The L2 norms of the wavelet detail images are subsequently calculated and it is these that are used to form the observation vector for that block. The L2 norm of an image is simply the square root of the sum of all the pixel values squared. As three detail images are produced at each decomposition level, the dimension of a block's observation vector will be three times the level of wavelet decomposition carried out. The image norms from all the image blocks are collected from all image blocks, in the order the blocks appear in the image, from left to right and from top to bottom, this forms the image's observation vector [13].

In the case of Gabor being used for feature extraction, the image is convolved with a number of Gabor filters, with 4 orientations and 6 scales being used. The output images are split into blocks in the same manner as that used for DWT. For each block, the L2 norm is calculated. Therefore, each block from the original image can be represented by a feature vector with 24 values (4 orientations $\times$ 6 scales). The image's observation vector is then constructed in the same manner as for DWT, with the features being collected from each block in the image, from left to right and from top to bottom.

This vector, along with the observation vectors from all other training images of the same individual, is used to train the HMM for this individual using maximum likelihood (ML) estimation. As the detail image norms are real values, a continuous observation HMM is employed. One HMM is trained for each identity in the database.

(2) Testing

A number of images are used to test the accuracy of the face recognition system. In order to ascertain the identity of an image, a feature vector for that image is created in the same way as for those images used to train the system. For each trained HMM, the likelihood of that HMM producing the observation vector is calculated. As the identification process assumes that all probe images belong to known individuals, the image is classified as the identity of the HMM that produces the highest likelihood value.

3. STRUCTURAL HIDDEN MARKOV MODELS

3.1. Mathematical background

One of the major problems of HMMs is due to the state conditional independence assumption that prevents them from capturing long-range dependencies. These dependencies often exhibit structural information that constitute the entire pattern. Therefore, in this section, the mathematical expression of SHMMs is introduced. The entire description of the SHMM can be found in [22, 23].

Let $O = (O_1, O_2, \ldots, O_s)$ be the time series sequence (the entire pattern) made of s subsequences (also called subpatterns). The entire pattern can be expressed as: $O = (o_{11}o_{12}\ldots o_{1r_1}, \ldots, o_{s1}, o_{s2}, \ldots, o_{sr_s})$, where r_1 is the number of observations in subsequence O_1 and r_2 is the number of observations in subsequence O_2, and so forth, such that $\sum_{i=1}^{i=s} r_i = T$. A local structure C_j is assigned to each subsequence O_i. Therefore, a sequence of local structures $C = (C_1, C_2, \ldots, C_s)$ is generated from the entire pattern O. The probability of a complex pattern O given a model λ can be written as

$$P(O \mid \lambda) = \sum_C P(O, C \mid \lambda). \tag{9}$$

Therefore, we need to evaluate $P(O, C \mid \lambda)$. The model λ is implicitly present during the evaluation of this joint probability, so it is omitted. We can write

$$\begin{aligned} P(O, C) &= P(C, O) = P(C \mid O) \times P(O) \\ &= P(C_s \mid C_{s-1} \cdots C_2 C_1 O_s \cdots O_1) \\ &\quad \times P(C_{s-1} \cdots C_2 C_1 \mid O_s \cdots O_1) \times P(O). \end{aligned} \tag{10}$$

It is assumed that C_i depends only on O_i and C_{i-1}, and the structure probability distribution is a Markov chain of order 1. It has been proven in [22] that the likelihood function of the observation sequence can be expressed as

$$P(O \mid \lambda) \approx \sum_C \left[\prod_{i=1}^{s} \frac{P(C_i \mid O_i) P(C_i \mid C_{i-1})}{P(C_i)} \times P(O) \right]. \tag{11}$$

The organization (or syntax) of the symbols $o_i = o_{uv}$ is introduced mainly through the term $P(C_i \mid O_i)$ since the transition probability $P(C_i \mid C_{i-1})$ does not involve the interrelationship of the symbols o_i. Besides, the term $P(O)$ of (11) is viewed as a traditional HMM.

Finally, an SHMM can be defined as follows.

Definition 1. A structural hidden Markov model is a quintuple $\lambda = [\pi, \mathcal{A}, \mathcal{B}, \mathcal{C}, \mathcal{D}]$, where

  (i) π is the initial state probability vector;
  (ii) $\mathcal{A}$ is the state transition probability matrix;
  (iii) $\mathcal{B}$ is the state conditional probability matrix of the visible observations,
  (iv) $\mathcal{C}$ is the posterior probability matrix of a structure given a sequence of observations;
  (v) $\mathcal{D}$ is the structure transition probability matrix.

An SHMM is characterized by the following elements.

  (i) **N** is the number of hidden states in the model. The individual states are labeled as $1, 2, \ldots, N$, and denote the state at time t as q_t.
  (ii) **M** is the number of distinct observations o_i.
  (iii) π is the initial state distribution, where $\pi_i = P(q_1 = i)$ and $1 \leq i \leq N$, $\sum_i \pi_i = 1$.
  (iv) $\mathcal{A}$ is the state transition probability distribution matrix: $\mathcal{A} = \{a_{ij}\}$, where $a_{ij} = P(q_{t+1} = j \mid q_t = i)$ and $1 \leq i, j \leq N$, $\sum_j a_{ij} = 1$.
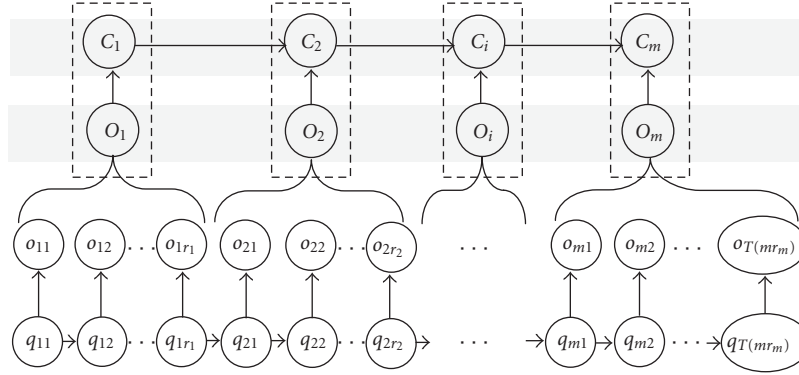
FIGURE 5: A graphical representation of a first-order structural hidden Markov model.

(v) $\mathcal{B}$ is the state conditional probability matrix of the observations, $\mathcal{B} = \{b_j(k)\}$, in which $b_j(k) = P(o_k \mid q_j)$, $1 \leq k \leq M$, and $1 \leq j \leq N$, $\sum_k b_j(k) = 1$. In the continuous case, this probability is a density function expressed as a finite weighted sum of Gaussian distributions (mixtures).

(vi) $\mathbf{F}$ is the number of distinct local structures.

(vii) $\mathcal{C}$ is the posterior probability matrix of a structure given its corresponding observation sequence: $\mathcal{C} = c_i(j)$, where $c_i(j) = P(C_j \mid O_i)$. For each particular input string O_i, we have $\sum_j c_i(j) = 1$.

(viii) $\mathcal{D}$ is the structure transition probability matrix: $\mathcal{D} = \{d_{ij}\}$, where $d_{ij} = P(C_{t+1} = j \mid C_t = i)$, $\sum_j d_{ij} = 1$, $1 \leq i, j \leq F$.

Figure 5 depicts a graphical representation of an SHMM of order 1. The problems that are involved in an SHMM can now be defined.

3.2. Problems assigned to a structural HMM

There are four problems that are assigned to an SHMM: (i) probability evaluation, (ii) statistical decoding, (iii) structural decoding, and (iv) parameter estimation (or training).

(i) *Probability evaluation.* Given a model λ and an observation sequence $O = (O_1, \ldots, O_s)$, the goal is to evaluate how well does the model λ match O.

(ii) *Statistical decoding.* In this problem, an attempt is made to find the best state sequence. This problem is similar to problem 2 of the traditional HMM and can be solved using Viterbi algorithm as well.

(iii) Structural decoding. This is the most important problem. The goal is to determine the "optimal local structures of the model." For example, the shape of an object captured through its external contour can be fully described by the local structures sequence: ⟨round, curved, straight,..., slanted, concave, convex,...⟩. Similarly, a primary structure of a protein (sequence of amino acids) can be described by its secondary structures such as "Alpha-Helix," "Beta-Sheet," and so forth. Finally, an autonomous robot can be trained to recognize the components of a human face described as a sequence of shapes such as ⟨round (human head), vertical line in the middle of the face (nose), round (eyes), ellipse (mouth),...,⟩.

(iv) *Parameter estimation (Training).* This problem consists of optimizing the model parameters $\lambda = [\pi, \mathcal{A}, \mathcal{B}, \mathcal{C}, \mathcal{D}]$ to maximize $P(O \mid \lambda)$. We now define each problem involved in an SHMM in more details.

(1) Probability evaluation

The evaluation problem in a structural HMM consists of determining the probability for the model $\lambda = [\pi, \mathcal{A}, \mathcal{B}, \mathcal{C}, \mathcal{D}]$ to produce the sequence O. From (11), this probability can be expressed as

$$P(O \mid \lambda) = \sum_C P(O, C \mid \lambda) = \sum_C \left\{ \prod_{i=1}^s \frac{c_i(i) \times d_{i-1,i}}{P(C_i)} \right\}$$
$$\times \sum_q \pi_{q_1} b_{q_1}(o_1) a_{q_1 q_2} b_{q_2}(o_2) \cdots a_{q_{(T-1)} q_T} b_{q_T}(o_T).$$

$$(12)$$

(2) Statistical decoding

The statistical decoding problem consists of determining the optimal state sequence $q^* = \arg\max_q(P(O_i, q \mid \lambda))$ that best "explains" the sequence of symbols within O_i. It is computed using Viterbi algorithm as in traditional HMM's.

(3) Structural decoding

The structural decoding problem consists of determining the optimal structure sequence $C^* = \langle C_1^*, C_2^*, \ldots, C_t^* \rangle$ such that

$$C^* = \arg\max_C P(O, C \mid \lambda). \tag{13}$$

We define

$$\delta_t(i) = \max_{\mathcal{C}} \left[P(O_1, O_2, \ldots, O_t, C_1, C_2, \ldots, C_t = i \mid \lambda) \right] \tag{14}$$

that is, $\delta_t(i)$ is the highest probability along a single path, at time t, which accounts for the first t strings and ends in structure i. Then, by induction we have

$$\delta_{t+1}(j) = [\max_i \delta_t(i)d_{ij}]c_{t+1}(j)\frac{P(O_{t+1})}{P(C_j)}. \qquad (15)$$

Similarly, this latter expression can be computed using *Viterbi* algorithm. However, δ is estimated in each step *through the structure transition probability matrix*. This optimal sequence of structures describes the structural pattern piecewise.

(4) Parameter estimation (training): the estimation of the density function

$P(C_j \mid O_i) \propto P(O_i \mid C_j)$ is established through a weighted sum of Gaussian mixtures. The mathematical expression of this estimation is

$$P(O_i \mid C_j) \approx \sum_{r=1}^{m=R} \alpha_{j,r} N(\mu_{j,r}, \Sigma_{j,r}, O_i), \qquad (16)$$

where $N(\mu_{j,r}, \Sigma_{j,r}, O_i)$ is a Gaussian distribution with mean $\mu_{j,r}$ and covariance matrix $\Sigma_{j,r}$. The mixing terms are subject to the constraint $\sum_{r=1}^{m=R} \alpha_{j,r} = 1$.

This *Gaussian mixture* posterior probability estimation technique obeys the exhaustivity and exclusivity constraint $\sum_j c_i(j) = 1$. This estimation enables the entire matrix $\mathcal{C}$ to be built. The Baum-Welch optimization technique is used to estimate the matrix $\mathcal{D}$. The other parameters, $\pi = \{\pi_i\}$, $\mathcal{A} = \{a_{ij}\}$, $\mathcal{B} = \{b_j(k)\}$, were estimated like in traditional HMM's [33].

(5) Parameter reestimation

Many algorithms have been proposed to re-estimate the parameters for traditional HMM's. For example, Djurić and chun [34] used "Monte Carlo Markov chain" sampling scheme. In the structural HMM paradigm, we have used a "forward-backward maximization" algorithm to re-estimate the parameters contained in the model λ. We used a bottom-up strategy that consists of re-estimating $\{\pi_i\}$, $\{a_{ij}\}$, $\{b_j(k)\}$ in the first phase and then re-estimating $\{c_j(k)\}$ and $\{d_{ij}\}$ in the second phase. Let us define

(i) $\xi_r(u, v)$ as the probability of being at structure u at time r and structure v at time $(r + 1)$ given the model λ and the observation sequence O. We can write

$$\xi_r(u, v) = P(q_r = u, q_{r+1} = v \mid \lambda, O)$$

$$= \frac{P(q_r = u, q_{r+1} = v, O \mid \lambda)}{P(O \mid \lambda)}. \qquad (17)$$

Using Bayes formula, we can write

$$\xi_r(u, v) = \frac{P(O_1 O_2 \cdots O_r, q_r = u \mid \lambda)d_{uv}P_v(O_{r+1})}{P(O_1 O_2 \ldots O_T \mid \lambda)}$$

$$\times \frac{P(O_{r+2}O_{r+3}\cdots O_T \mid q_r = v, \lambda)}{P(O_1 O_2 \cdots O_T \mid \lambda)}. \qquad (18)$$

Then we define the following probabilities:

(i) $\alpha_r(u) = P(O_1 O_2 \cdots O_r, q_r = u \mid \lambda)$,
(ii) $\beta_r(u) = P(O_{r+1}O_{r+2}\cdots O_T \mid q_r = u, \lambda)$,
(iii) $P_v(O_{r+1}) = P(q_{r+1} = v \mid O_{r+1}) \times P(O_{r+1})/P(q_{r+1} = v)$,

therefore,

$$\xi_r(u, v) = \frac{\alpha_r(u)d_{uv}s_{r+1}(v)P(O_{r+1})\beta_{r+1}(v)}{P(O_1 O_2 \cdots O_T \mid \lambda)P(q_{r+1} = v)}. \qquad (19)$$

We need to compute the following:

(i) $P(O_{r+1}) = P(o_{r+1}^1 \cdots o_{r+1}^k \mid \lambda) = \sum_{\text{all } q} P(O_{r+1} \mid q, \lambda)P(q \mid \lambda) = \sum_{q_1,\ldots,q_T} \pi_{q_1} b_{q_1}(o_1)a_{q_1 q_2} \cdots b_{q_k}(o_k)$,
(ii) $P(q_{r+1} = v) = \sum_j P(q_{r+1} = v \mid q_r = j)$,
(iii) The term $P(O_1 O_2 \cdots O_T \mid \lambda)$ requires π, $\mathcal{A}$, $\mathcal{B}$, $\mathcal{C}$, $\mathcal{D}$. However, the parameters π, $\mathcal{A}$, and $\mathcal{B}$ can be re-estimated as in traditional HMM. In order to re-estimate $\mathcal{C}$ and $\mathcal{D}$, we define

$$\gamma_r(u) = \sum_{v=1}^{N} \xi_r(u, v). \qquad (20)$$

Then we compute the improved estimates of $c_v(r)$ and d_{uv} as

$$\hat{d}_{uv} = \frac{\sum_{r=1}^{T-1}\xi_r(u, v)}{\sum_{r=1}^{T-1}\gamma_r(u)}, \qquad (21)$$

$$\hat{c}_v(r) = \frac{\sum_{r=1,O_r=v_r}^{T-1}\gamma_r(v)}{\sum_{r=1}^{T}\gamma_r(v)}. \qquad (22)$$

From (22), we derive

$$\hat{c}_r(v) = \hat{c}_v(r) \times \frac{\hat{P}(q_r = v)}{\hat{P}(O_r)}. \qquad (23)$$

We calculate improved $\xi_r(u, v)$, $\gamma_r(u)$, $\hat{d}_{uv}$, and $\hat{c}_r(v)$ repeatedly until some convergence criterion is achieved. We have used the Baum-Welch algorithm also known as forward-backward (an example of a generalized expectation-maximization algorithm) to iteratively compute the estimates $\hat{d}_{uv}$ and $\hat{c}_r(v)$.

The stopping or convergence criterion that we have selected in line 8 halts learning when no estimated transition probability changes more than a predetermined positive amount ε. Other popular stopping criteria (e.g., as the one based on overall probability that the learned model could have produced the entire training data) can also be used. However, these two criteria can produce only a local optimum of the likelihood function, they are far from reaching a global optimum.

3.3. Novel SHMM modeling for human face recognition

(1) Feature extraction

SHMM modeling of the human face has never been undertaken by any researchers or practitioners in the biometric

(1) *Begin initialize* $\hat{d}_{uv}$, $\hat{c}_r(v)$, training sequence, convergence criterion ε
(2) repeat
(3) $z \leftarrow z + 1$
(4) compute $\hat{d}(z)$ from $d(z-1)$ and $c(z-1)$ using (21)
(5) compute $\hat{c}(z)$ from $d(z-1)$ and $c(z-1)$ using (22)
(6) $d_{uv}(z) \leftarrow \hat{d}_{uv}(z-1)$
(7) $c_{rv}(z) \leftarrow \hat{c}_{rv}(z-1)$
(8) *until* $\max_{u,r,v}[d_{uv}(z) - d_{uv}(z-1), c_{rv}(z) - c_{rv}(z-1)] < \varepsilon$ (convergence achieved)
(9) *return* $d_{uv} \leftarrow \hat{d}_{uv}(z)$; $c_{rv} \leftarrow \hat{c}_{rv}(z)$
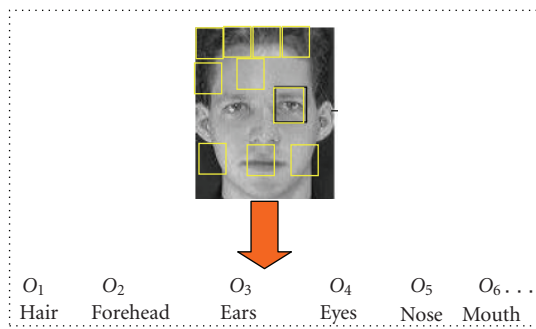(10) *End*

ALGORITHM 1

FIGURE 6: A face O is viewed as an ordered sequence of observations O_i. Each O_i captures a significant facial region such as "hair," "forehead," "eyes," "nose," "mouth," and so on. These regions come in a natural order from top to bottom and left to right.
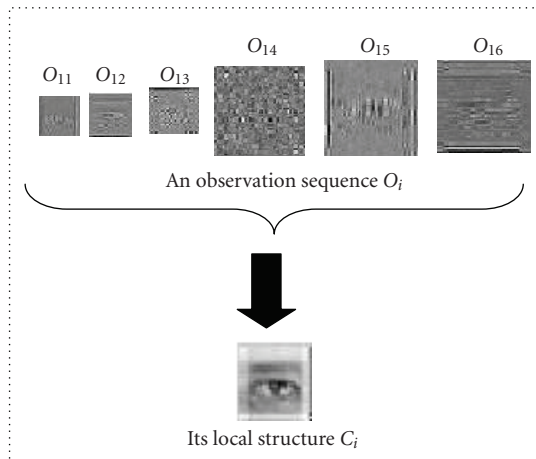

FIGURE 7: A block O_i of the whole face O is a time-series of norms assigned to the multiresolution detail images. This block belongs to the local structure "eyes."

community. Our approach of adapting the SHMM's machine learning to recognize human faces is novel. The SHMM approach to face recognition consists of viewing a face as a sequence of blocks of information O_i which is a fixed-size two-dimensional window. Each block O_i belongs to some predefined facial regions as depicted in Figure 6. This phase involves extracting observation vector sequences from subimages of the entire face image. As with recognition using standard HMMs, DWT is used for this purpose. The observation vectors are obtained by scanning the image from left to right and top to bottom using the fixed-size two-dimensional window and performing DWT analysis at each subimage. The subimage is decomposed to a certain level and the energies of the subbands are selected to form the observation sequence O_i for the SHMM. If Gabor filters are used, the original image is convolved with a number of Gabor kernels, producing 24 output images. These images are then divided into blocks using the same fixed-size two-dimensional window as for DWT. The energies of these blocks are calculated and form the observation sequence O_i for the SHMM. *The local structures C_i of the SHMM include the facial regions of the face. These regions are hair, forehead, ears, eyes, nose, mouth, and so on.* However, the observation sequence O_i corresponds to the different resolutions of the block images of the face. *The sequence of norms of the detail images d_j^k represents the observation sequence O_i.* Therefore, each observation sequence O_i is a multidimensional vector. Each block is assigned one and only one facial region. Formally, a local structure C_j is simply an equivalence class that gathers all "similar" O_i. *Two vectors $O_i's$ (two sets of detail images) are equivalent if they share the same facial region of the human face.* In other words, the facial regions are all clusters of vectors $O_i's$ that are formed when using the k-means algorithm. Figure 7 depicts an example of a local structure and its sequence of observations. This modeling enables the SHMM to be trained efficiently since several sets of detail images are assigned to the same facial region.

(2) Face recognition using SHMM

The training phase of the SHMM consists of building a model $\lambda = [\pi, \mathcal{A}, \mathcal{B}, \mathcal{C}, \mathcal{D}]$ for each human face during a training phase. Each parameter of this model will be trained through the wavelet multiresolution analysis applied to each face image of a person. The testing phase consists of decomposing each test image into blocks and automatically assigning a facial region to each one of them. As the structure of a face is significantly more complex than other applications for which SHMM has been employed [22, 23], this phase is
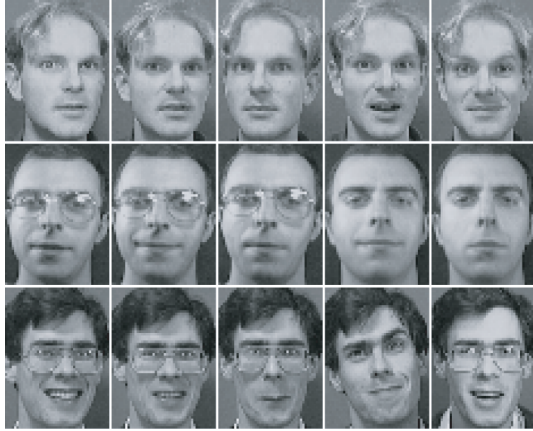
(a)

(b)

FIGURE 8: Samples of faces from (a) the AT&T Database of Faces [17] and (b) the Essex Faces95 database [35]. The images contain variation in pose, expression, scale, and illumination, as well as presence/absence of glasses.

conducted via the k-means clustering algorithm. The value of k corresponds to the number of facial regions (or local structures) selected a priori. The selection of this value was based in part upon visual inspection of the output of the clustering process for various values of k. When k equalled 6, the clustering process appeared to perform well, segmenting the face image into regions such as forehead, mouth, and so on. Each face is expressed as a sequence of blocks O_i with their facial regions C_i. The recognition phase will be performed by computing the model λ^* in the training set (database) that maximizes the likelihood of a test face image.

4. EXPERIMENTS

4.1. Data collection

Experiments were carried out using three different training sets. The AT&T (formerly ORL) Database of Faces [17] contains ten grayscale images each of forty individuals. The images contain variation in lighting, expression, and facial details (e.g., glasses/no glasses). Figure 8(a) shows some im-
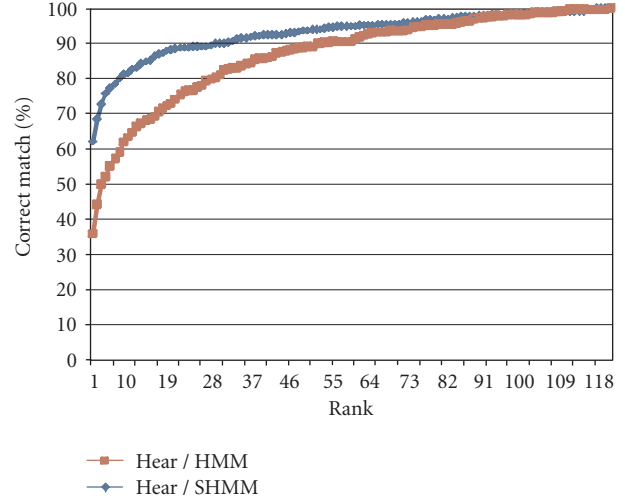

FIGURE 9: Cumulative match scores for FERET database using Haar wavelet.

ages taken from the AT&T Database. The second database used was the Essex Faces95 database [35], which contains twenty color images each of seventy-two individuals. These images contain variation in lighting, expression, position, and scale. Figure 8(b) shows some images taken from the Essex database. For the purposes of the experiments carried out, the Essex faces were converted to grayscale prior to training. The third database used was the Facial Recognition Technology (FERET) grayscale database [36, 37]. Images used for experimentation were taken from the fa (regular facial expression), fb (alternative facial expression), ba (frontal "b" series), bj (alternative expression to ba), and bk (different illumination to ba) images sets. Those individuals with at least five images (taken from the specified sets) were used for experimentation. This resulted in a test set of 119 individuals. These images were rotated and cropped based on the known eye coordinate positions, followed by histogram equalization. Experimentation was carried out using Matlab on a 2.4 Ghz Pentium 4 PC with 512 Mb of memory.

4.2. Face identification results using wavelet/HMM

The aim of the initial experiments was to investigate the efficacy of using wavelet filters (DWT/Gabor) for feature extraction with HMM-based face identification. A variety of DWT filters were used, including Haar, biorthogonal9/7, and Coiflet(3). The observation vectors were produced as described in Section 2, with both height j and width k of observation blocks equalling 16, with overlap of 4 pixels. The size of the blocks was chosen so that significant structures/textures could be adequately represented within the block. The overlap value of 4 was deemed large enough to allow structures (e.g., edges) that straddled the edge of one block to be better contained within the next block. Wavelet decomposition was carried out to the fourth decomposition level (to allow a complete decomposition of the image). In the case of Gabor filters, 6 scales and 4 orientations were used, producing an observation blocks of size 24.
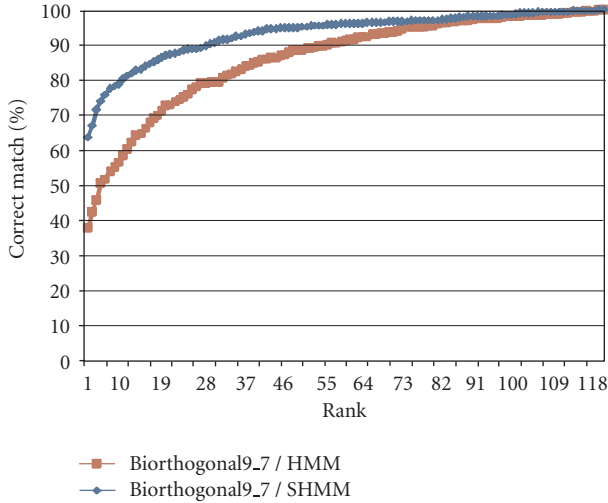
Figure 10: Cumulative match scores for FERET database using Biorthogonal9/7 wavelet.
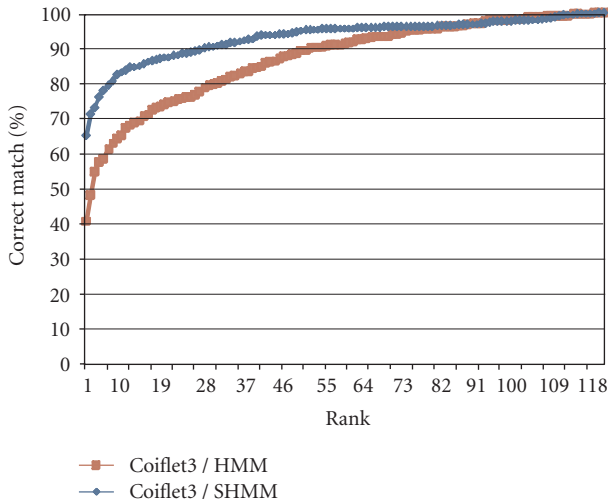
Table 1: Comparison of HMM face identification accuracy when performed in the spatial domain and with selected wavelet filters (%).

	AT&T	Essex95	FERET
Spatial	87.5	71.9	31.1
Haar	95.75	84.2	35.8
Biorthogonal 9/7	93.5	78.0	37.5
Coiflet(3)	96.5	85.6	40.5
Gabor	96.8	85.9	42.9

Figure 11: Cumulative match scores for FERET database using Coiflet(3) wavelet.
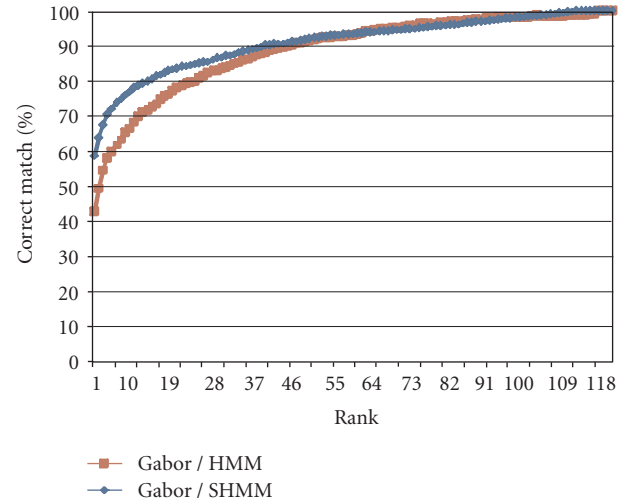
Figure 12: Cumulative match scores for FERET database using Gabor features.

The experiments were carried out using five-fold cross validation. This involved splitting the set of training images for each person into five equally sized sets and using four of the sets for system training with the remainder being used for testing. The experiments were repeated five times with a different set being used for testing each time to provide a more accurate recognition figure. Therefore, with the AT&T database, eight images were used for training and two for testing during each run. When using the Essex95 database, sixteen images were used for training and four for testing during each run. For the FERET database, four images per individual were used for training, with the remaining image being used for testing.

One HMM was trained for each individual in the database. During testing, an image was assigned an identity according to the HMM that produced the highest likelihood value. As the task being performed was face identification, it was assumed that all testing individuals were known individuals. Accuracy of an individual run is thus defined as the ratio of correct matches to the total number of face images tested, with final accuracy equalling the average accuracy figures from each of the five cross-validation runs. The accuracy figures for HMM face recognition performed in both the spatial domain and using selected wavelet filters are presented in Table 1.

As can be seen from Table 1, the use of DWT for feature extraction improves recognition accuracy. With the AT&T database, accuracy increased from 87.5%, when the observation vector was constructed in the spatial domain, to 96.5% when the Coiflet(3) wavelet was used. This is a very substantial 72% decrease in the rate of false classification. The increase in recognition rate is also evident for the larger Essex95 database. Recognition rate increased from 71.9% in the spatial domain to 84.6% in the wavelet domain. As before, the Coiflet(3) wavelet produced the best results. Recognition rate also increased for the FERET database, with the recognition rate increasing from 31.1% in the spatial domain to 40.5% in the wavelet domain. DWT has been shown to improve recognition accuracy when used in a variety of face recognition approaches, and clearly this benefit extends to HMM-based face recognition. Using Gabor filters increased recognition results even further. The identification rate for the AT&T database rose to 96.8% and the Essex figure became 85.9%.

TABLE 2: Comparison of face identification accuracy when performed using wavelet/HMM and wavelet/SHMM (%).

	AT&T		Essex		FERET	
	DTW/HMM	DWT/SHMM	DWT/HMM	DWT/SHMM	DWT/HMM	DWT/SHMM
Haar	95.75	97.5	84.2	89.4	35.8	62.0
Biorthogonal 9/7	93.5	95.1	78.0	84.6	37.5	63.9
Coiflet(3)	96.5	97.8	85.6	90.7	40.5	65.2
Gabor	96.8	97.3	85.9	88.7	42.9	58.7

TABLE 3: Comparative results on AT&T database.

Method	Accuracy (%)	Ref.
DCT/HMM	84	[12]
ICA	85	[38]
Weighted PCA	88	[39]
Gabor filters and rank correlation	91.5	[40]
2D-PHMM	94.5	[17]
NMF	96	[41]
LFA	97	[42]
DWT/SHMM	97	(Proposed)

TABLE 4: Comparison of training and classification times for AT&T database images (s).

	Training time per image	Classification time per image
Spatial/HMM	7.24	22.5
DWT/HMM	1.09	1.19
DWT/SHMM	4.31	3.45

4.3. Face identification results using wavelet/SHMM

The next set of experiments was designed to establish if SHMM provided a benefit over HMM for face recognition. Where appropriate, the same parameters were used for SHMM as for HMM (such as block size). The experiments were carried out solely in the wavelet domain, due to the benefits identified by the previous results. The recognition accuracy for SHMM face recognition is presented in Table 2. In addition, Figures 9 to 12 present the cumulative match score graphs for the FERET database.

As can be seen from the results, the use of SHMM instead of HMM increases recognition accuracy in all cases tested. Indeed, the incorrect match rate for Haar/SHMM is 40% lower than the equivalent figure for Haar/HMM when tested using the AT&T database. This is a significant increase in accuracy.

The most significant increases in performance, however, were for the FERET dataset. The use of 5-fold cross-validation constrained options when it came to choosing images for experimentation. As the system was not designed to handle images with any significant degree of rotation, they were selected from those subsets which were deemed suitable—Fa, Fb, ba, bj, and bk. Within these subsets, how-

ever, there was variation in illumination, pose, scale, and expression. Most significantly, the "b" set images were captured in different sessions from the images in the "F" sets. Coupled with the number of identities in the FERET dataset that were used (119), the variation among the images made this a difficult task for a face identification system. It is for this reason that the recognition rates for wavelet/HMM are rather low for this database, ranging from 35.8% when Haar was used to 42.9% for Gabor. The recognition rates increase dramatically though when SHMM is used. 62.9% of images are correctly identified when Haar is used, with a more modest increase to 58.7% for Gabor filters. The Coiflet(3) wavelet produces the best results, with 65.2% correctly identified, as opposed to 40.5% for wavelet/HMM. In many face recognition applications, it is less important that an individual is recognized correctly than it is that an individual's identity appears within the top x matches, where x could be, perhaps, 10. The cumulative match score graphs allow for this information to be retrieved. SHMM provides a substantial benefit in cases where the top x matches can be considered. For example, using the Biorthogonal9/7 wavelet, the correct identity appears within the top 10 matches 60.2% of the time. This increases to 81.3% with SHMM. If the Haar wavelet is used, the figure increases from 65.0% to 82.9%.

Experiments were also carried out to enable comparison of the results with those reported in the literature. Although the ability to compare works was an important consideration in the creation of the FERET database, many authors use subsets from it that match their particular requirements. There are, however, many studies employing the AT&T database that use 50% of the database images for training and the remaining 50% for testing. With this in mind, an experiment was performed with these characteristics. Table 3 shows that the DWT/SHMM approach performs well when compared with other techniques that have used this data set.

In addition to recognition accuracy, an important factor in a face recognition system is the time required for both system training and classification. As can be seen from Table 4, this is reduced substantially by the use of DWT. Feature extraction and HMM training took approximately 7.24 seconds per training image when this was performed in the spatial domain using the AT&T database, as opposed to 1.09 seconds in the wavelet domain, even though an extra step was required (transformation to wavelet domain). This is a very substantial time difference and is due to the fact that the number of observations used to train the HMM is reduced by a factor of almost 30 in the wavelet domain. The time benefit realized by using DWT is even more obvious during the

recognition stage, as the time required is reduced from 22.5 seconds to 1.19 seconds.

SHMM does increase the time taken for both training and classification, although this is offset by the improvement in recognition accuracy. Fortunately, the increase in time taken for classification is still a vast improvement on the time taken for HMM recognition in the spatial domain. The time taken for classification is particularly important, as it is this stage where real-time performance is often mandated.

5. CONCLUSION

In this paper, we have carried out an analysis of the benefits of using DWT along with HMM for face recognition. In addition, a novel approach to this problem has been proposed, based on the fusion of the DWT and, for the first time in the field of face recognition, the SHMM. It is worth noting that the SHMM allows both the statistical and the structural information of a pattern to be modeled within the same probabilistic framework. The combination of the DWT and the SHMM has been shown to outperform the combination of DWT and HMM for face identification, as well as techniques such as PCA and ICA. Our future work is twofold: we plan to

(i) study the effect of window size (block dimension) on the SHMM model parameters and therefore on the accuracy;

(ii) adapt the SHMM modeling to account for prior information such as morphological differences of human faces with respect to their geographical environment, this external information will enhance the power of generalization of the SHMMs.

REFERENCES

[1] W. Zhao, R. Chellappa, P. J. Phillips, and A. Rosenfeld, "Face recognition: a literature survey," *ACM Computing Surveys*, vol. 35, no. 4, pp. 399–458, 2003.

[2] R. Gross, S. Baker, I. Matthews, and T. Kanade, "Face recognition across pose and illumination," in *Handbook of Face Recognition*, S. Z. Li and A. K. Jain, Eds., Springer, New York, NY, USA, June 2004.

[3] G. Lawton, "Biometrics: a new era in security," *Computer*, vol. 31, no. 8, pp. 16–18, 1998.

[4] L. Torres, "Is there any hope for face recognition?" in *Proceedings of the 5th Internationl Workshop on Image Analysis for Multimedia Interactive Services (WIAMIS '04)*, pp. 2709–2712, Lisbon, Portugal, April 2004.

[5] A. Amira and P. Farrell, "An automatic face recognition system based on wavelet transforms," in *Proceedings of International Symposium on Circuits and Systems (ISCAS '05)*, vol. 6, pp. 6252–6255, Kobe, Japan, May 2005.

[6] M. Turk and A. Pentland, "Eigenfaces for recognition," *Journal of Cognitive Neuroscience*, vol. 3, no. 1, pp. 71–86, 1991.

[7] H. Moon and P. J. Phillips, "Computational and performance aspects of PCA-based face-recognition algorithms," *Perception*, vol. 30, no. 3, pp. 303–320, 2001.

[8] P. Nicholl, A. Amira, and R. Perrott, "An automated grid-enabled face recognition system using hybrid approaches," in *Proceedings of the 5th IEE/IEEE Postgraduate Research Conference on Electronics, Photonics, Communications and Networks (PREP '05)*, pp. 144–146, Lancester, UK, March 2005.

[9] P. C. Yuen and J.-H. Lai, "Face representation using independent component analysis," *Pattern Recognition*, vol. 35, no. 6, pp. 1247–1257, 2002.

[10] E. Kussul, T. Baidyk, and M. Kussul, "Neural network system for face recognition," in *Proceedings of International Society for Computer Aided Surgery (ISCAS '04)*, vol. 5, pp. 768–771, Vancouver, Canada, May 2004.

[11] L. R. Rabiner, "A tutorial on hidden markov models and selected applications in speech recognition," in *Readings in Speech Recognition*, pp. 267–296, Morgan Kaufmann, San Francisco, Calif, USA, 1990.

[12] A. V. Nefian and M. H. Hayes, "Hidden markov models for face recognition," in *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP '98)*, pp. 2721–2724, Seattle, Wash, USA, May 1998.

[13] L. Bai and L. Shen, "Combining wavelets with hmm for face recognition," in *Proceedings of the 23rd International Conference on Innovative Techniques and Applications of Artificial Intelligence (SGAI '03)*, Cambridge, UK, December 2003.

[14] I. Daubechies, "Wavelet transforms and orthonormal wavelet bases," in *Different Perspectives on Wavelets (San Antonio, Tex, 1993)*, vol. 47 of *Proceedings of Symposia in Applied Mathematics*, pp. 1–33, American Mathematical Society, Providence, RI, USA, 1993.

[15] M. Bicego, U. Castellani, and V. Murino, "Using hidden markov models and wavelets for face recognition," in *Proceedings of the12th International Conference on Image Analysis and Processing (ICIAP '03)*, pp. 52–56, Mantova, Italy, September 2003.

[16] H.-S. Le and H. Li, "Recognizing frontal face images using hidden Markov models with one training image per person," in *Proceedings of International Conference on Pattern Recognition (ICPR '04)*, vol. 1, pp. 318–321, Cambridge, UK, August 2004.

[17] F. Samaria, *Face recognition using hidden markov models*, Ph.D. thesis, Department of Engineering, Cambridge University, Cambridge, UK, 1994.

[18] H. Othman and T. Aboulnasr, "A separable low complexity 2D HMM with application to face recognition," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 25, no. 10, pp. 1229–1238, 2003.

[19] S. Fine, Y. Singer, and N. Tishby, "The hierarchical hidden markov model: analysis and applications," *Machine Learning*, vol. 32, no. 1, pp. 41–62, 1998.

[20] G. Jin, L. Tao, and G. Xu, "Cues extraction and hierarchical hmm based events inference in soccer video," in *proceedings of the 2nd European Workshop on the Integration of Knowledge, Semantics and Digital Media Technology*, pp. 73–76, London, UK, November-December 2005.

[21] A. V. Nefian and M. H. Hayes, "Maximum likelihood training of the embedded HMM for face detection and recognition," in *IEEE International Conference on Image Processing (CIP '00)*, vol. 1, pp. 33–36, Vancouver, Canada, September 2000.

[22] D. Bouchaffra and J. Tan, "Introduction to structural hidden markov models: application to handwritten numeral recognition," *Intelligent Data Analysis Journal*, vol. 10, no. 1, 2006.

[23] D. Bouchaffra and J. Tan, "Structural hidden markov models using a relation of equivalence: application to automotive designs," *Data Mining and Knowledge Discovery*, vol. 12, no. 1, pp. 79–96, 2006.

[24] J.-T. Chien and C.-C. Wu, "Discriminant waveletfaces and nearest feature classifiers for face recognition," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 24, no. 12, pp. 1644–1649, 2002.

[25] S. Gundimada and V. Asari, "Face detection technique based on rotation invariant wavelet features," in *Proceedings of International Conference on Information Technology: Coding Computing (ITCC '04)*, vol. 2, pp. 157–158, Las Vegas, Nev, USA, April 2004.

[26] G. C. Feng, P. C. Yuen, and D. Q. Dai, "Human face recognition using PCA on wavelet subband," *Journal of Electronic Imaging*, vol. 9, no. 2, pp. 226–233, 2000.

[27] M. T. Harandi, M. N. Ahmadabadi, and B. N. Araabi, "Face recognition using reinforcement learning," in *Proceedings of International Conference on Image Processing (ICIP '04)*, vol. 4, pp. 2709–2712, Singapore, October 2004.

[28] C. Liu, "Gabor-based kernel PCA with fractional power polynomial models for face recognition," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 26, no. 5, pp. 572–581, 2004.

[29] M. Zhou and H. Wei, "Face verification using gaborwavelets and adaboost," in *Proceedings of the 18th International Conference on Pattern Recognition (ICPR '06)*, vol. 1, pp. 404–407, Hong Kong, August 2006.

[30] S. Mallat, "A theory for multiresolution signal decomposition: the wavelet representation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 11, no. 7, pp. 674–693, 1989.

[31] F. S. Samaria and A. C. Harter, "Parameterisation of a stochastic model for human face identification," in *Proceedings of the 2nd IEEE Workshop on Applications of Computer Vision (ACV '94)*, pp. 138–142, Sarasota, Fla, USA, December 1994.

[32] E. J. Stollnitz, T. D. DeRose, and D. H. Salestin, "Wavelets for computer graphics: a primer.1," *IEEE Computer Graphics and Applications*, vol. 15, no. 3, pp. 76–84, 1995.

[33] L. Rabiner and B. H. Juang, *Fundamentals of Speech Recognition*, Prentice-Hall, Upper Saddle River, NJ, USA, 1993.

[34] P. M. Djurić and J.-H. Chun, "An MCMC sampling approach to estimation of nonstationary hidden Markov models," *IEEE Transactions on Signal Processing*, vol. 50, no. 5, pp. 1113–1123, 2002.

[35] D. Hond and L. Spacek, "Distinctive descriptions for face processing," in *Proceedings of the 8th British Machine Vision Conference (BMVC '97)*, pp. 320–329, Essex, UK, September 1997.

[36] P. J. Phillips, H. Wechsler, J. Huang, and P. J. Rauss, "The FERET database and evaluation procedure for face-recognition algorithms," *Image and Vision Computing*, vol. 16, no. 5, pp. 295–306, 1998.

[37] P. J. Phillips, H. Moon, S. A. Rizvi, and P. J. Rauss, "The FERET evaluation methodology for face-recognition algorithms," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 22, no. 10, pp. 1090–1104, 2000.

[38] J. Kim, J. Choi, J. Yi, and M. Turk, "Effective representation using ICA for face recognition robust to local distortion and partial occlusion," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 27, no. 12, pp. 1977–1981, 2005.

[39] H.-Y. Wang and X.-J. Wu, "Weighted PCA space and its application in face recognition," in *Proceedings of International Conference on Machine Learning and Cybernetics (ICMLC '05)*, vol. 7, pp. 4522–4527, Guangzhou, China, August 2005.

[40] O. Ayinde and Y.-H. Yang, "Face recognition approach based on rank correlation of Gabor-filtered images," *Pattern Recognition*, vol. 35, no. 6, pp. 1275–1289, 2002.

[41] Y. Xue, C. S. Tong, W.-S. Chen, W. Zhang, and Z. He, "A modified non-negative matrix factorization algorithm for face recognition," in *Proceedings of International Conference on Pattern Recognition (ICPR '06)*, vol. 3, pp. 495–498, Hong Kong, August 2006.

[42] E. F. Ersi and J. S. Zelek, "Local feature matching for face recognition," in *Proceedings of the 3rd Canadian Conference on Computer and Robot Vision (CRV '06)*, p. 4, Quebec City, Canada, June 2006.