Available online at www.sciencedirect.com
ScienceDirect
Pattern Recognition 41 (2008) 852–867
PATTERN RECOGNITION
THE JOURNAL OF THE PATTERN RECOGNITION SOCIETY
www.elsevier.com/locate/pr

Structural hidden Markov models for biometrics: Fusion of face and fingerprint

Djamel Bouchaffra[a,*], Abbes Amira[b]

[a]*Department of Mathematics and Computer Science, Grambling State University, Grambling, LA 71245, USA*
[b]*Electronic & Computer Engineering School of Engineering & Design, Brunel University, London, UK*

Received 9 May 2007; accepted 26 June 2007

Abstract

The goal of this paper is threefold: (i) propose a novel face and fingerprint feature modeling using the structural hidden Markov models (SHMMs) paradigm, (ii) explore the use of some feature extraction techniques such as ridgelet transform, discrete wavelet transform with various classifiers for biometric identification, and (iii) determine the best method for classifier combination. The experimental results reported in both fingerprint and face recognition reveal that the SHMMs concept is promising since it has outperformed several state-of-the-arts classifiers when combined with the discrete wavelet transform. Besides, this study has shown that the ridgelet transform without principal components analysis (PCA) dimension reduction fits better with the support vector machines (SVMs) classifier than it does with the SHMMs in the fingerprint recognition task. Finally, these results also reveal a small improvement of the bimodal biometric system over unimodal systems; which suggest that a most effective fusion scheme is necessary.

Keywords: Multimodal biometrics; Discrete wavelet transform; Ridgelet transform; Structural hidden Markov models; Support vector machines; Classifier combination

1. Introduction

Multimodal biometrics systems are gaining acceptance among designers and practitioners due to: (i) their performance superiority over unimodal systems, and (ii) their system speed improvement which has become admissible and satisfactory. However, in an identification system, the choice of the first modality is usually based on the matching speed rather than on the error rate criterion. This modality is often combined in a serial mode with subsequent modalities that are slower but more accurate in the average [1]. Nevertheless, even if theoretically, under some hypotheses, the combination scheme has proven to be superior in terms of accuracy, practically, some precautions need to be taken to satisfy these hypotheses.

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 identification 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 data sensing methods currently in use in both modalities limit their versatility [2]. 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 [3]. A robust face recognition solution has many potential applications. Business organizations are aware of the ever-increasing need for security—this is mandated both by their own desire to protect property and processes, but also by their workforce's increasing demands for workplace safety and security [4]. Local law enforcement

* Corresponding author. Tel.: +1 586 744 3184.
E-mail addresses: dbouchaffra@ieee.org (D. Bouchaffra), Abbes-Amira@brunel.ac.uk (A. Amira).

0031-3203/$30.00
doi:10.1016/j.patcog.2007.06.033

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 due mostly 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 become therefore non-trivial 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 the long-term (aging process). These issues all increase the complexity of the recognition task [5] significantly.

Hidden Markov models (HMMs) [6], 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 [7] modified the approach by using discrete cosine transform (DCT) coefficients to form observation vectors. Bai and Shen [8] used discrete wavelet transform (DWT) [9] coefficients taken from overlapping image sub-windows taken from the entire face image, whereas Bicego and Murino [10] used DWT coefficients of sub-windows 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 [11], 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) [12], 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) [13], 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 2-D HMM (2D-HMM), however, recognition accuracy is lower as a result. Although HMMs are effective in modeling sequential information statistically [14], they are not suited to discern local structures that constitutes the entire pattern. In other words, the state conditional independence assumption inherent to the traditional HMMs makes these models unable to capture long-range dependencies. They are therefore not optimal for handling a structural patterns such as a person's 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 often identified as facial regions (lips, hair, forehead, eyes, etc.).

One recently developed model for pattern recognition is the structural hidden Markov models (SHMMs) [15,16]. To avoid the complexity problem inherent to the determination of the higher level states, the SHMMs provide 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. This approach also allows the user to weight substantially the local structures within a pattern that are difficult to disguise. This provides a SHMM recognizer with a higher degree of robustness. Therefore, the SHMM is well suited to simultaneously model the within and between structures information of any sequential pattern (such as a face or a fingerprint). Indeed, the concept of SHMMs have been shown to outperform HMMs in a number of applications including handwriting recognition [15], but have yet to be applied to face recognition.

Besides, fingerprint identification has a long history of utilization in forensic, it is one of the most well known and publicized biometrics. Its identification is based primarily on the minutiae points, or the location and direction of the ridge endings and bifurcations along a ridge path. Template vectors are computed and stored in the database for verification or identification purposes. However, prior to extracting minutiae points, a bank of Gabor filters, orientated to different angles are often applied to the fingerprint image to remove possible noise [17,18]. There are several types of fingerprint matching techniques such as SVM, PCA, neural networks (NNs), and HMMs [19] that have been proposed to model the fingerprint features. One of our ultimate goals in this paper is to compare these traditional classifiers with the proposed SHMM.

However, the fusion of both modalities remains still a challenge in the biometric community [20–22]. Two major reasons for this challenge are due to (i) the lack of interpretation of the confidence values assigned the output classes, and (ii) the difficulty of finding mathematical models that relate the optimization criterion for features selection with the classifier accuracy (power of generalization). We have adopted a fusion at the matching score level, and different methods for combining these scores are explored.

The objective of the work presented in this paper is threefold: (i) explore the use of discrete wavelets [9,23] transform, to extract face features known as "facial regions" and fingerprint features known as "minutiae types". The ridgelets [24] transform will be also be investigated, (ii) introduce SHMMs to the biometric community as a strong model for these two physiological features, and (iii) determine the best fusion method of classifier combination for a person identification task. Fig. 1 depicts an image of the proposed bimodal biometric system.

This paper is organized as follows. Section 2 covers some feature extraction techniques in the transform domain. Section 3 describes previous work in face recognition that is based on traditional HMMs. Section 4 introduces a novel approach to model physiological traits using the SHMMs. Experimental

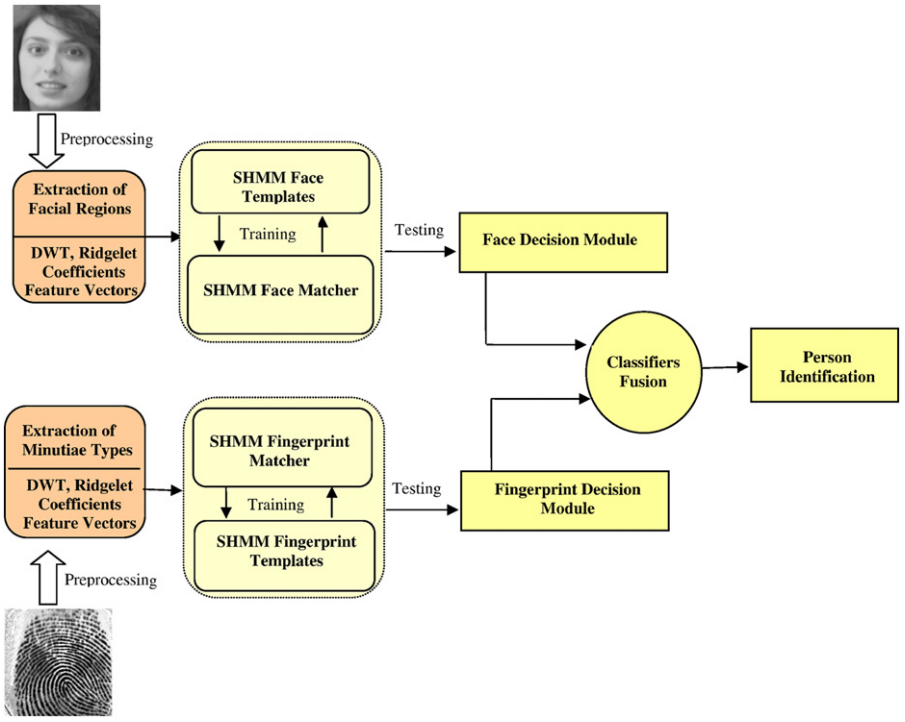

Fig. 1. A bimodal biometric system that uses wavelet and ridgelet transforms within a SHMM. This system is capable of extracting and learning facial regions and fingerprint minutiae types of a person.

results and analysis are reported in Section 5 and the conclusion and future work is laid in Section 6.

2. Feature extraction in the transform domain

This section describes some major feature extraction techniques that can be embedded within classifiers in order to build a biometric system.

2.1. 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 multi-resolution analysis of a signal with localization in both time and frequency. In addition to this, functions with discontinuities and functions with sharp spikes require fewer wavelet basis vectors in the wavelet domain than sine–cosine 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 filter. 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 expressed as

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

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. Fig. 2 illustrates DWT schematically.

The simplest wavelet basis is the Haar wavelet [25], which operates by calculating averages and differences for each pair of coefficients, with the process continuing recursively on the output of each step. The total number of steps involved is $\log_2(c)$, where c refers to the number of coefficients in the signal. An example of the process is given in Table 1. After decomposition, the original signal can be represented as a combination of one average coefficient and all the detail coefficients—in the example given, the signal becomes $[621 - 1]$.

For the case of images, the 1D-DWT can be readily extended to two dimensions. In standard 2D wavelet decomposition, the image rows are fully decomposed, with the output being fully decomposed columnwise. In non-standard wavelet decomposition, all the rows are decomposed by one decomposition level followed by one decomposition level of the columns. At this stage, the image is composed of quadrants, with each corresponding to a set of coefficients. The "low–low" (LL) quadrant contains low frequency information in both directions. The

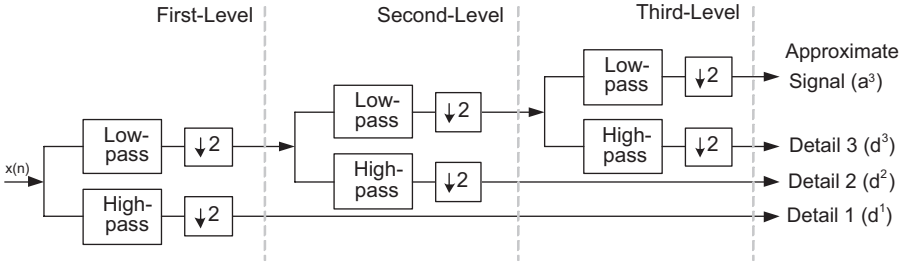

Fig. 2. A three-level wavelet decomposition system.

Table 1
Process of Haar decomposition

Resolution	Averages	Detail coefficients
4	[9 7 3 5]	
2	[8 4]	[1 −1]
1	[6]	[2]

"high–high" (HH) quadrant contains high frequency in both directions. Both the "high–low" (HL) and "low–high" (LH) quadrants contain low frequency information along one direction and high frequency information along the other. The process of non-standard decomposition is shown in Algorithm 1.

Algorithm 1. DWT non-standard decomposition (2DArray x)

Require : $x = 2^n * 2^n$
  begin
  $i = \mathrm{Log}_2(n)$
  while $i > 0$ **do**
    for all $r, r \in \{\text{rows of } x(1 \ldots 2^i)\}$ **do**
      Decompose r
    end for
    for all c, $c \in \{\text{columns of } x(1 \ldots 2^i)\}$ **do**
      Decompose c
    end for
    $i - -$
  end while
  end.

The decomposition of the image continues by decomposing the low resolution output from each step, until the image is fully decomposed. Fig. 3 illustrates the effect of applying the non-standard wavelet transform to an image. Fig. 3a shows an image taken from the AT&T Database of faces [26]. Fig. 3b shows the effect of applying a one-level wavelet decomposition to this image (using the Haar wavelet). Fig. 3c shows the effect of completely decomposing the image. 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 non-standard 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.

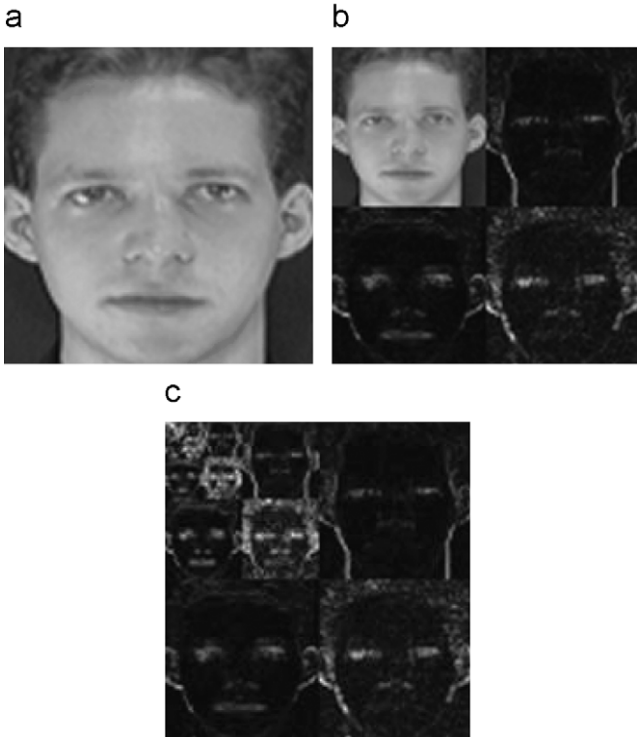

Fig. 3. Wavelet transform of image: (a) original image; (b) one-level Haar decomposition; (c) complete decomposition.

2.2. Finite ridgelet transform

Recently, the curvelet and ridgelet transforms [27,28,24] have been generating a lot of interest due to their superior performance over wavelets. While wavelets have been very successful in applications such as denoising and compact approximations of images containing zero dimensional (point) singularities, they do not isolate the smoothness along edges that occurs in images because they lack flexible directionality. Wavelets are thus more appropriate for the reconstruction of sharp point-like singularities than lines or edges. These shortcomings of wavelets are well addressed by the ridgelet and curvelet transforms, as they extend the functionality of wavelets to higher dimensional singularities, and are effective

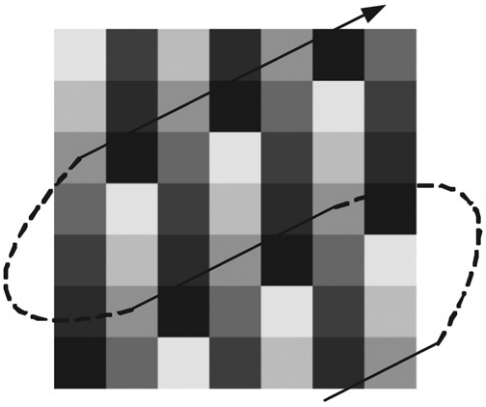

Fig. 4. FRAT basis function (projection kernel) for blocksize $p=7$ and $k=4$. The digital line superimposed over the kernel corresponds to $l=4$.

tools to perform sparse directional analysis. The basic building block of these transforms is the finite Radon transform (FRAT).

2.2.1. The finite radon transform

The FRAT was first introduced in Ref. [29] as the finite analogue of integration in the continuous Radon transform (CRT), with origins in the field of combinatorics. The mathematical representation of an injective form of the FRAT to ensure invertibility when applied on finite Euclidian planes has been presented in Ref. [30]. It is worth mentioning that the FRAT is not a discredited version of the RT, but a discrete finite version. Consider a cyclic group Z_p denoted by $Z_p=(0,1,\ldots,p-1)$ such that p is a prime number. Let the finite grid Z_p^2 be defined as the Cartesian product of $Z_p \times Z_p$. This finite grid has $(p+1)$ non-trivial subgroups, given by

$$L_{k,l}=\{(i,j): j=(ki+l)\ (\mathrm{mod}\ p), i\in Z_p\}, \quad k<p \tag{2}$$

and

$$L_{p,l}=\{(l,j): j\in Z_p\}, \tag{3}$$

where each subgroup $L_{k,l}$, is the set of points that define a line on the lattice Z_p. The radon projection of the function f on the finite grid Z_p^2 is given by

$$r_k[l]=FRAT_f(k,l)=\frac{1}{\sqrt{p}}\left(\sum_{(i,j)\in L_{k,l}} f[i,j]\right). \tag{4}$$

From Eqs. (2)–(4), it can be seen that the function f is treated as a periodic function, and hence the digital representation of the line displays a wrap around effect, as illustrated in Fig. 4.

Fig. 4 represents all the set of lines corresponding to the FRAT function for blocksize $p=7$ and $k=4$. There are $p+1$ vectors in the FRAT domain, each corresponding to one "direction". The line corresponding to $k=4$, $l=4$ corresponds to the 4th digital line in the 4th rotational direction. The concept of "periodicity" is highlighted for the specific case where $l=4$, and is superimposed over the kernel in Fig. 4. We can see that there are three (Euclidian) parallel lines that actually

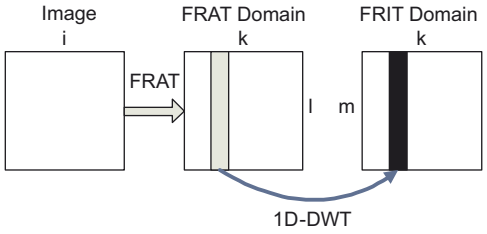

Fig. 5. Finite Ridgelet transform obtained by performing DWT on the FRAT vectors.

constitute a *single* digital line, connected by the dots. This is known as the wrap-around effect.

Analogous to the continuous case, as in Euclidian geometry, any two lines intersect at only one point in the finite grid Z_p^2. Hence, the inverse transform i.e. FBP is given by

$$FBP_r(i,j)=\frac{1}{\sqrt{p}}\left(\sum_{(k,l)\in P_{i,j}} r_k[l]\right), \quad (i,j)\in Z_p^2, \tag{5}$$

where

$$P_{i,j}=\{(k,l): l=(j-ki)\ (\mathrm{mod}\ p), k\in Z_p\}\cup\{(p,i)\} \tag{6}$$

Substituting Eq. (4) in (5), we get

$$FBP_r(i,j)=\frac{1}{p}\left(\sum_{(k,l)\in P_{i,j}}\ \sum_{(k,l)\in P_{i,j}} f[i',j']\right) \tag{7}$$

$$=f[i,j]. \tag{8}$$

2.2.2. Haar wavelet

The Haar wavelet is one of the simplest wavelet transforms. Although its simplicity makes it unsuitable for a number of applications, particularly due to the fact that it is not continuous and hence not differentiable, it is highly suitable for constructing the FRIT due to the following reasons. Firstly, the Haar wavelet is very simple in structure, and is less expensive to compute compared to other wavelets. Secondly, due to the averaging characteristics of the FRIT, it has been observed that the Haar wavelets display superior performance in energy compaction when compared to other wavelets.

2.2.3. Building the FRIT from FRAT and DWT

The continuous ridgelet transform [24] of a bivariate function $f(x)$ is given by

$$RT=\int_{R^2}\psi_{a,b,\theta}(x)\,f(x)\,\mathrm{d}x. \tag{9}$$

However, the use of digital images necessitates the development of suitable variations of the ridgelets to deal with images in the digital domain. An orthonormal, invertible and discrete form of the ridgelet, called finite ridgelet transform was first proposed in Ref. [27]. The FRIT is obtained by performing DWT on each FRAT projection sequence with fixed value of k. This process is pictorially represented in Fig. 5. Mathematically,

Fig. 6. Transformed face image, $p = 7$.

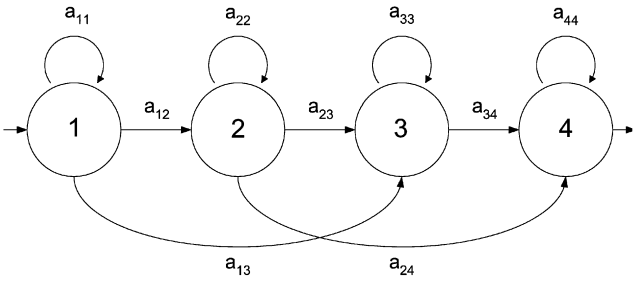

Fig. 7. A simple left-right HMM.

this is given by

$$FRIT_{i,i+1}(k,l) = \lfloor FRIT_{i,i+1}(k,l) \times H_2 \rfloor_2, \quad \forall i = 1, \ldots, p. \tag{10}$$

The finite and discredited variant of curvelets can be obtained by repeated application of the FRIT. The directional attributes of these higher dimensional generalizations of wavelets make them ideal for a number of applications such as alternate image representation, compression, denoising, etc.

The FRIT domain visualization of commonly used images is shown in Fig. 6.

3. Face identification using HMMs

Although there has been several modeling techniques for face identification, in this section, we focus only on approaches that view a human face as a time series sequence represented by a Markovian process.

3.1. Hidden Markov models formalism

HMMs are used to characterize the statistical properties of a signal [6]. 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 non-observable states and an observable sequence, generated by the individual hidden states. Fig. 7 illustrates the structure of a simple HMM. HMMs are defined by the following elements:

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

$$a_{ij} = P[q_{t+1} = S_j | q_t = S_i], \quad 1 \leqslant i, j \leqslant N \tag{11}$$

with

$$0 \leqslant a_{i,j} \leqslant 1$$

and

$$\sum_{j=1}^{N} a_{ij} = 1, \quad 1 \leqslant i \leqslant N.$$

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

$$b_j(k) = P[O_t = k | q_t = S_j] \tag{12}$$

with

$$1 \leqslant j \leqslant N$$

and O_t is the observation symbol at time t.
- $\Pi = \{\pi_i\}$, the initial state probability distribution, i.e.,

$$\pi_i = P[q_1 = S_i], \quad 1 \leqslant i \leqslant N \tag{13}$$

with

$$\pi_i \geqslant 0 \quad \text{and} \quad \sum_{i=1}^{N} \pi_i = 1.$$

An HMM can therefore be succinctly defined by the triplet $\lambda = [A, B, \Pi]$. HMMs are typically used to address three unique problems [6]:

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

This problem is typically used for pattern recognition tasks; the probability of an observation sequence being generated is evaluated against a number of distinct HMMs, each of which correspond to a class of pattern. The pattern is classified as belonging to the same class as the HMM which produces the highest probability of producing the observation sequence.

Decoding: Given a model λ and a sequence of observations O, what is the hidden state sequence q^* most likely to have produced O, i.e., $q^* = \text{argmax}_q[P(q|\lambda, O)]$. This problem can be used to learn about the structure of a model, or to find the optimal state sequence in some application, such as speech recognition.

Parameter estimation: Given an observation sequence O, what model λ is most likely to have produced O. This problem is referred to as training, as the model's parameters are adjusted until some convergence criterion is reached. Typically, a number of observation sequences are used to train a model [6].

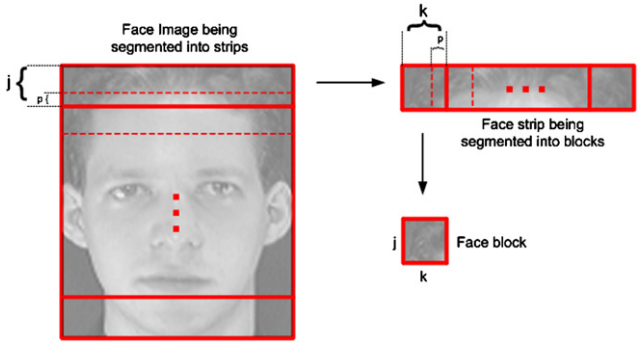

Fig. 8. An illustration showing the creation of the block sequence.

3.2. Face identification process

3.2.1. Feature extraction

In the context of face identification problem, 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 Fig. 8. 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 $[a_J, \{d_j^1, d_j^2, d_j^3\}_{j=1\ldots J}]$ 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, four-level wavelet decomposition was employed, producing a vector with one average image and 12 detail images. The L2 norms of the wavelet detail images were subsequently calculated and it is these that were 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 (or feature) vector.

3.2.2. Training and testing HMMs

In this section, the goal is to build the HMMs assigned to the different faces contained in the training set (database). One HMM is built for each face; each HMM is trained with different face templates (different level of illumination) of the same individual. This training process is conducted using the Baum–Welch algorithm (see Problem 3 in HMM section). As the detail image norms are real values, a continuous observation HMM is employed.

A number of images are used to test the accuracy of the face identification system. In order to ascertain the identity of a test image, a feature vector for that image is created in the same way as for the images used to train the system. For each trained HMM, the likelihood of that HMM producing the observation vector is calculated (see Problem 1 in HMM section). The image is classified as the identity of the HMM λ^* that produces the highest likelihood value.

4. Structural hidden Markov models for unimodal biometrics

4.1. Mathematical background

One of the major problems that is inherent to the traditional HMMs is the state conditional independence assumption that prevents them from capturing long-range dependencies. These dependencies often exhibit structural information that constitute the entire pattern. We introduce in this section a brief description of the SHMMs. The entire description of the SHMM concept can be found in Refs. [15,16].

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: $O=(o_{11}\ldots o_{1r_1}, \ldots, o_{s1}, \ldots, o_{sr_s})$, (where r_1 is the number of observations in subsequence O_1 and r_2 is the number of observations in subsequence O_2, etc., 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

$$P(O|\lambda)=\sum_C P(O, C|\lambda). \tag{14}$$

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

$$P(O, C)=P(C, O)=P(C|O)\times P(O) \tag{15a}$$

$$=P(C_s|C_{s-1}\ldots C_2C_1O_s\ldots O_1) \tag{15b}$$

$$\times P(C_{s-1}\ldots C_2C_1|O_s\ldots O_1)\times P(O). \tag{15c}$$

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 Ref. [15] that the likelihood function of the observation sequence can be expressed as

$$P(O|\lambda)\approx\sum_C\left[\prod_{i=1}^{s}\frac{P(C_i|O_i)P(C_i|C_{i-1})}{P(C_i)}\times P(O)\right]. \tag{16}$$

The organization (or syntax) of the symbols $o_i=o_{uv}$ is introduced mainly through the term $P(C_i|O_i)$ since the transition probability $P(C_i|C_{i-1})$ does not involve the inter-relationship of the symbols o_i. Besides, the term $P(O)$ of Eq. (16) is viewed as a traditional HMM.

Finally, a SHMM can be defined as follows:

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

- *π is the initial state probability vector,*
- *$\mathscr{A}$ is the state transition probability matrix,*

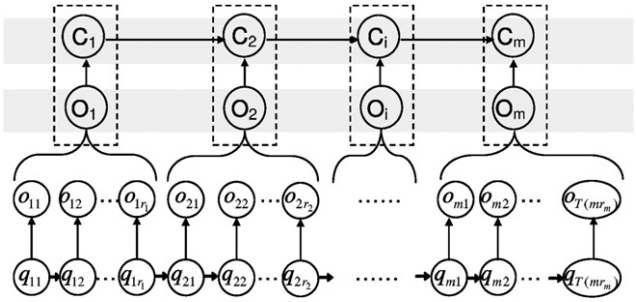

Fig. 9. A graphical representation of a first-order structural hidden Markov model.

- $\mathscr{B}$ *is the state conditional probability matrix of the visible observations*,
- $\mathscr{C}$ *is the posterior probability matrix of a structure given a sequence of observations*,
- $\mathscr{D}$ *is the structure transition probability matrix.*

A SHMM is characterized by the following elements:

- **N**, 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.
- **M**, the number of distinct observations o_i.
- π, the initial state distribution, where $\pi_i = P(q_1 = i)$ and $1 \leqslant i \leqslant N$, $\sum_i \pi_i = 1$.
- $\mathscr{A}$, the state transition probability distribution matrix: $\mathscr{A} = \{a_{ij}\}$, where: $a_{ij} = P(q_{t+1} = j | q_t = i)$ and $1 \leqslant i, j \leqslant N$, $\sum_j a_{ij} = 1$.
- $\mathscr{B}$, the state conditional probability matrix of the observations, $\mathscr{B} = \{b_j(k)\}$, in which: $b_j(k) = P(o_k | q_j)$, $1 \leqslant k \leqslant M$ and $1 \leqslant j \leqslant N$, $\sum_k b_j(k) = 1$. In the continuous case, this probability is a density function expressed as a finite weighted sum of Gaussian mixtures.
- **F**, the number of distinct local structures.
- $\mathscr{C}$ is the posterior probability matrix of a structure given its corresponding observation sequence: $\mathscr{C} = c_i(j)$, where: $c_i(j) = P(C_j | O_i)$. For each particular input string O_i, we have $\sum_j c_i(j) = 1$.
- $\mathscr{D}$, the structure transition probability matrix: $\mathscr{D} = \{d_{ij}\}$, where $d_{ij} = P(C_{t+1} = j | C_t = i)$, $\sum_j d_{ij} = 1$, $1 \leqslant i, j \leqslant F$.

Fig. 9 depicts a graphical representation of a SHMM of order 1. The problems that are involved in a SHMM can now be defined.

4.2. Problems assigned to a SHMM

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

Probability evaluation (Scoring): 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.

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.

Structural decoding: The goal in this phase 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", etc. 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),…,⟩.

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

4.2.1. Probability evaluation

The evaluation problem in a SHMM consists of determining the probability for the model $\lambda = [\pi, \mathscr{A}, \mathscr{B}, \mathscr{C}, \mathscr{D}]$ to produce the sequence O. From Eq. (16), this probability can be expressed as

$$P(O|\lambda) = \sum_C P(O, C|\lambda) = \sum_C \left\{ \prod_{i=1}^{s} \frac{c_i(i) \times d_{i-1,i}}{P(C_i)} \right\} \tag{17}$$

$$\times \sum_q \pi_{q_1} b_{q_1}(o_1) a_{q_1 q_2} b_{q_2}(o_2) \ldots a_{q_{(T-1)} q_T} b_{q_T}(o_T). \tag{18}$$

4.2.2. Statistical decoding

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

4.2.3. 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

$$C^* = \arg \max_C P(O, C|\lambda). \tag{19}$$

We define $\delta_t(i) = \max_{\mathscr{C}} [P(O_1, O_2, \ldots, O_t, C_1, C_2, \ldots, C_t = i|\lambda)]$ 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

$$\delta_{t+1}(j) = \left[\max_i \delta_t(i) d_{ij} \right] c_{t+1}(j) \frac{P(O_{t+1})}{P(C_j)}. \tag{20}$$

Similarly, this latter expression can be computed using *Viterbi* algorithm. However, δ is estimated in each step *through the*

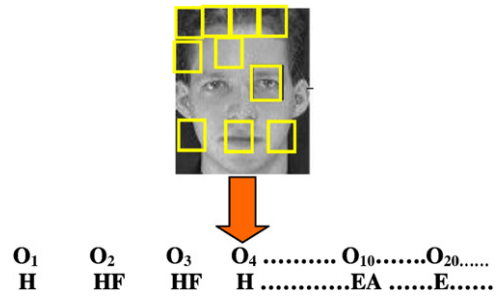

Fig. 10. A face O is viewed as an ordered observation sequence O_i. Each O_i captures a facial region such as: "hair" (H), "between hair and forehead" (HF), "forehead" (F), "ears" (EA), "eyes" (E), "nose" (N), "mouth" (M), "chin" (C), etc. These regions come in a natural order from top-to-bottom and left-to-right.

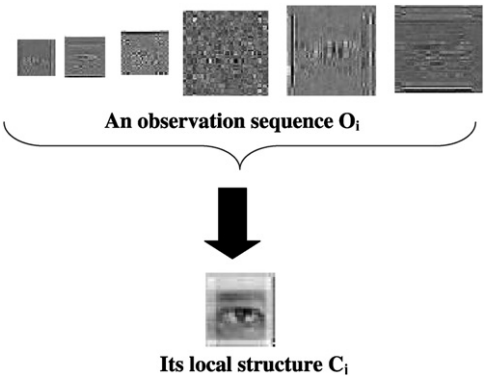

Fig. 11. A block O_i of the whole face O is a time-series of norms assigned to the multi-resolution detail images. This block belongs to the local structure "eyes".

structure transition probability matrix. This optimal sequence of structures describes the structural pattern piecewise.

4.2.4. Parameter estimation (training)

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

$$P(O_i|C_j) \approx \sum_{r=1}^{m=R} \alpha_{j,r} N(\mu_{j,r}, \Sigma_{j,r}, O_i), \tag{21}$$

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 $\mathscr{C}$ to be built. The Baum–Welch optimization technique is used to estimate the matrix $\mathscr{D}$. The other parameters, $\pi = \{\pi_i\}$, $\mathscr{A} = \{a_{ij}\}$, $\mathscr{B} = \{b_j(k)\}$, were estimated like in traditional HMM's [31].

In the next section, we show how the SHMMs can model physiological traits for a person identification task. It is worth to outline that the SHMM modeling of the human face and fingerprint has never been undertaken by any researchers or practitioners in the biometric community.

4.3. Face identification

4.3.1. Feature extraction

The SHMM approach to face recognition consists of viewing a face as a sequence of blocks of information O_i. Each block O_i is a fixed-size 2D window that belongs to some predefined facial regions as depicted in Fig. 10. 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 2D window and performing DWT analysis at each subimage. The subimage is decomposed into a certain level and the energies of the subbands are selected to 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, etc.* 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" $O_i's$. *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. Fig. 11 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 detailed images are assigned to the same facial region.

4.3.2. Training and testing the SHMMs

The training phase of the SHMM consists of building a model $\lambda = [\pi, \mathscr{A}, \mathscr{B}, \mathscr{C}, \mathscr{D}]$ for each human face template in the database. Each SHMM is trained with different face images of the same individual. Each parameter of this model will be trained through the wavelet multi-resolution analysis applied to each set of face images of a person. This task is performed using problem (iv) of SHMMs defined in Section 4.2.

The testing phase consists of decomposing each test image into blocks and automatically assigning a facial region to each one of them. This phase is conducted via the k-means clustering algorithm. The value of k corresponds to the number of facial regions (or local structures) selected a priori. 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. This problem

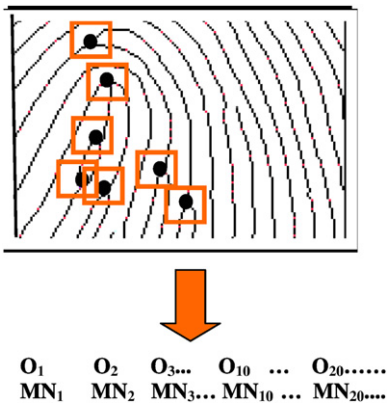

Fig. 12. A fingerprint O is viewed as an ordered observation sequence O_i. Each O_i captures a finger minutiae region MN_i.

is known as the probability evaluation: problem (i) discussed in Section 4.2.

4.4. Fingerprint identification

4.4.1. Feature extraction

Unlike in the face identification task, in the fingerprint identification task, the observation vectors $O_i's$ are obtained by scanning the finger image only around the minutiae points. A fixed size 2D rectangular window (region) is built around all minutiae points in the fingerprint image. Fig. 12 depicts this minutiae points in black and the regions of interest which are the windows. This process assumes that the minutiae points are predetermined and each window subimage reveals the type of the minutiae points. We then perform a DWT analysis at each subimage. The subimage is decomposed into a certain level and the energies of the subbands are selected to form the observation sequence O_i for the SHMM. Similar to the face recognition process, the vectors $O_i's$ which represent the minutiae regions are partitioned into classes of equivalence. The vectors $O_i's$ that are equivalent are gathered into a class of equivalence that represents a *minutiae type*. These minutiae types are the local structures $C_i's$ defined in the SHMM framework.

It is worth to underscore that the minutiae types convey more information than the minutiae points since it reveals the shape around the minutiae points. This allows the possibility of discriminating minutiae points by assigning different weights on the minutiae types.

5. Experimental results and analysis

5.1. Face identification

5.1.1. Data collection

Experiments were carried out using two different training sets. The AT&T (formerly ORL) Database of faces [12] contains 10 grayscale images each of 40 individuals. The images contain variation in lighting, expression and facial details (for example, glasses/no glasses). Fig. 13(a) shows some images taken from the AT&T database. The other database used was the Essex Faces 95 database [32], which contains 20 color images each of 72 individuals. These images contain variation in lighting, expression, position and scale. Fig. 13(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. Experiments were carried out using Matlab on a 2.4 Ghz Pentium 4 PC with 512 Mb of memory.

The aim of the initial experiments was to investigate the efficacy of using DWT for feature extraction with HMM-based face recognition. A variety of wavelet filters were used, including Haar, biorthogonal 9/7 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. Wavelet decomposition was carried out to the fourth decomposition level.

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 Essex 95 database, 16 images were used for training and four for testing during each run. 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. 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 2. As can be seen from this table, 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 Essex 95 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. 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.

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 Tables 3 and 4. As can be seen from the tables, the use of SHMM instead of HMM increases

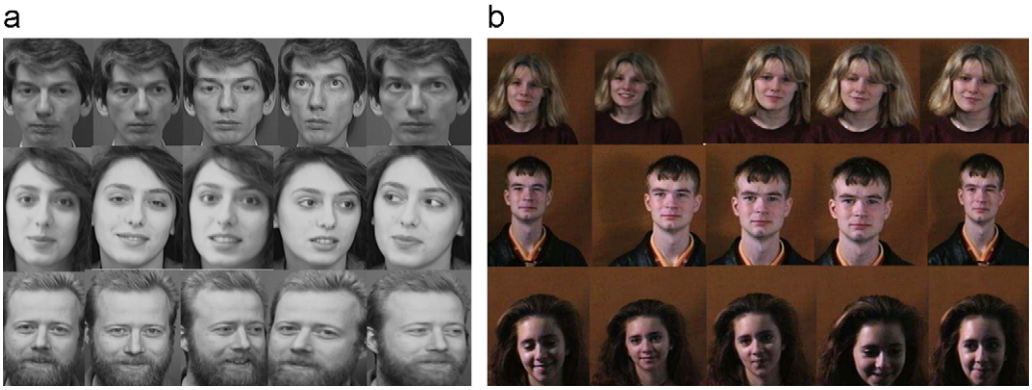

Fig. 13. Samples of faces from (a) the AT&T database of faces [12] and (b) the Essex Faces 95 database [32]. The images contain variation in pose, expression, scale and illumination, as well as presence/absence of glasses.

Table 2
Comparison of HMM identification accuracy (%) in the spatial domain with selected wavelet filters

	AT&T	Essex 95
Spatial	87.5	71.9
Haar	95.75	84.2
Biorthogonal 9/7	93.5	78.0
Coiflet(3)	96.5	85.6

Table 3
Comparison of face identification accuracy (%) using DWT/HMM and DWT/SHMM on the AT&T database

	DWT/HMM	DWT/SHMM
Haar	95.75	97.5
Biorthogonal 9/7	93.5	95.1
Coiflet(3)	96.5	97.8

Table 4
Comparison of face identification accuracy (%) using DWT/HMM and DWT/SHMM on the Essex database

	DWT/HMM	DWT/SHMM
Haar	84.2	89.4
Biorthogonal 9/7	78.0	84.6
Coiflet(3)	85.6	90.7

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. For comparison purposes, an experiment was performed to find the accuracy for DWT/SHMM when using five images from the AT&T database for training and five images for testing. As Table 5 shows, the DWT/SHMM approach to face recognition compares well with other techniques from the literature that have used this training set.

In addition to identification accuracy, an important factor in a face identification system is the time required for both system training and classification. As can be seen from Table 6, this is reduced substantially by the use of DWT. Feature extraction and HMM training took approximately 7.24 s per training image when this was performed in the spatial domain using the AT&T database, as opposed to 1.09 s 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 to 1.19 s.

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.

Table 5
Comparative results on AT&T database

Method	Accuracy (%)	Ref.
DCT/HMM	84	[7]
Independent component analysis (ICA)	85	[33]
PCA	91	[34]
Gabor filters & rank correlation	91.5	[35]
2D-PHMM	94.5	[12]
DWT/SHMM	97	(Proposed)
Ridgelet/SHMM	94.7	

Table 6
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

Fig. 14. Noise caused by image intensity.

Fig. 15. Noise in the images caused by distance.

Fig. 16. Image variations due to tilt.

Fig. 17. Noise in the images due to off center location.

Fig. 18. Noise due to incomplete images.

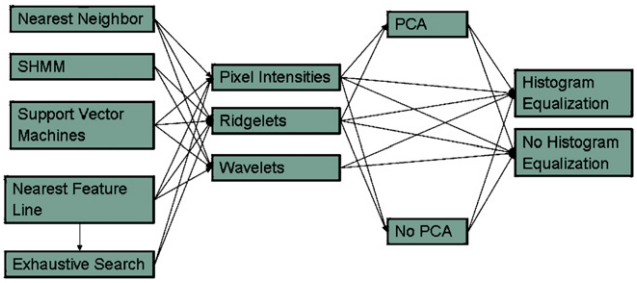

Fig. 19. Combinations of all the classifiers tested.

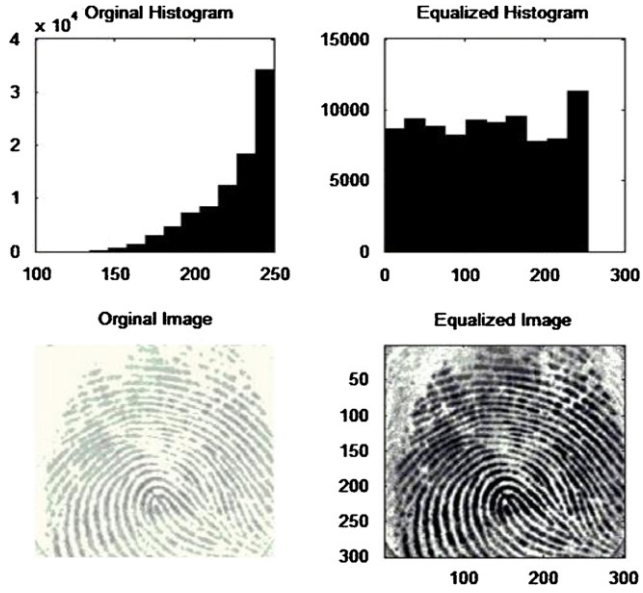

Fig. 20. Sample of a fingerprint before and after histogram equalization.

5.2. Fingerprint identification

5.2.1. Different types of noise in fingerprints

Fingerprints were much harder to recognize than the face images due to the lack of many distinct features. While, everyone has unique fingerprints, it is more challenging for a computer to distinguish them since they all hold similar lines. Furthermore, with even a slightest amount of noise, a fingerprint of one person may become more similar to a fingerprint of another person due to overlap of the lines in the matched areas. Figs. 14–18 show the types of noise observed in the fingerprint database. Fig. 14 shows two images on the same person but with different intensity, where one image is much darker than the other. The noise in Fig. 15 is caused by the distance of the image, where one image is closer and "zoomed in", whereas, the other image is farther. Fig. 16 shows the same image at different tilt angles. The images in Fig. 17 vary due to translation, where images are shifted either up or sideways. Lastly, Fig. 18 shows noise caused by incomplete images.

5.2.2. Combinations of feature extraction techniques with classifiers

Multiple combinations of various preprocessing, feature extraction, and classification algorithms were used for fingerprint identification. Fig. 19 shows all the possible combinations. Histogram equalization [36] was used for preprocessing to reduce the affects of noise caused by image intensity. Histogram equalization works by creating a histogram of pixel intensities and then changing some intensities to equalize the amount of pixels in each intensity range. This results in improving the contrast of the image. Fig. 20 depicts an image before and after histogram equalization with its corresponding pixel intensities histogram.

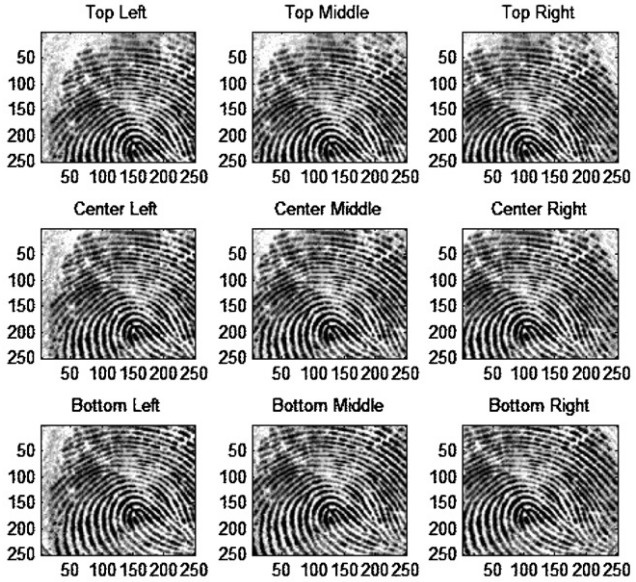

Fig. 21. Sample crops from exhaustive search algorithm.

Fig. 22. Actual full image.

Multiple feature extraction techniques were utilized for the purpose of comparison. These include pixel intensities, ridgelets, wavelets, and PCA. Furthermore, PCA was also used as a method of dimension reduction for large dimension of ridgelet coefficients. All the algorithms were tested with and without applying PCA for dimension reduction in ridgelet coefficients.

The classification techniques compared include nearest neighbor (NN), nearest feature line (NFL) [37], SVMs, and SHMMs. Furthermore, a shift optimization technique called "Exhaustive search" was introduced to account for shifting in the images. This technique crops the test image in various locations and compares it to the center crop of the database image. Fig. 21 shows sample crops of Fig. 22 for exhaustive search algorithm.

5.2.3. Performance comparison between modalities

It was observed that the combination of wavelet feature extraction and SHMMs produced the best results for both the fingerprint and the face data. In the face identification algorithm, the wavelet method produced approximately 2.8% better results than the ridgelets. This result is depicted in Table 5. In fingerprint identification, the SHMM results were approximately 3% better than the SVMs as shown in Table 7. Furthermore, it was observed that whether the ridgelets or the wavelets performed better was also classifier dependent. In the final algorithm, SHMM, wavelets performed approximately 1.25% better than ridgelets. However, in other algorithms such as nearest feature line and SVM, the performance of ridgelets was better than wavelets. Similarly, the results using images with or without histogram equalization were also classifier dependent. Lastly, it was observed that reducing the dimension of the ridgelet coefficients using PCA worsened the accuracy most, but not all the time.

5.3. Fusion of face and fingerprint modalities

The desired goal of the data fusion step is to increase the accuracy of the identification process by combining the results of the face and fingerprint classifiers [38]. The fusion is conducted between the top five confidence values from each classifier. The following methods were used to fuse the confidence values from the two classifiers: "Maximum Z-score", "Sum of Z-scores", "Sum of raw Scores" and "Borda count" [39].

5.3.1. Normalization

In order to be able to compare the confidence values between the two classifiers, we first need to put them into the same scale. Therefore, we have first performed a normalization of these values using the Z-score normalization method [40]. Other techniques that convert confidence values to a posteriori probabilities can also be applied [41]. We have assumed a Gaussian distribution of the confidence value random variable, and we used the following formula for normalization:

$$\textit{Normalized score} = \frac{\textit{Raw score} - \textit{Mean}}{\textit{Standard deviation}}. \tag{22}$$

The results of the data fusion of the final face and fingerprint algorithm using wavelets and SHMM are shown in Table 8. This data highlights results when using fingerprint images with and without histogram equalization. The classification from various fusion techniques is also shown.

It is observed that using histogram equalization better results are produced when using Max Z-score and Raw score. The Borda count method returned the highest overall accuracy of 98% in the AT&T database without using histogram equalization. In cases where the correct person did not have the highest normalized confidence value, the Borda count was able to take into account the correct persons representation between the two classifiers.

6. Conclusion and future work

We have introduced a novel classifier known as SHMMs to the biometric community. We have also explored different feature extraction techniques such as DWT (with different kernels), and ridgelet for face and fingerprint data. Finally, we devised a bimodal biometric system that merges evidences from both modalities.

Our experiments highlight the benefits of using DWT along with SHMM for both face and fingerprint identification. In fact, the combination DWT/SHMM has been proven to outperform the combination ridgelet/SHMM, as well as techniques such as

Table 7
Fingerprint classifier accuracy (%) using different pairs of feature extraction/classifier methods

Preprocessing	Shift optimization	Feature extraction	PCA for dimension reduction	Classifier	Accuracy
		Wavelet		SHMM	80.94
Histogram equalization		Wavelet		SHMM	76.45
–	–	Ridgelet	–	SHMM	79.66
Histogram equalization		Ridgelet		SHMM	73.52
		Ridgelet	Yes	SHMM	74.75
Histogram equalization	–	Ridgelet	Yes	SHMM	69.26
		Wavelet		SVM	78.20
Histogram equalization		Wavelet		SVM	75.17
		Ridgelet		SVM	79.71
Histogram equalization		Ridgelet		SVM	77.76
		Ridgelet	Yes	SVM	77.28
Histogram equalization		Ridgelet	Yes	SVM	76.21
		Pixel intensities		SVM	64.35
Histogram equalization	–	Pixel intensities	–	SVM	69.85
		Wavelet		NFL	72.19
Histogram equalization		Wavelet		NFL	69.80
		Ridgelet		NFL	74.79
Histogram equalization		Ridgelet	–	NFL	75.15
		Ridgelet	Yes	NFL	76.95
Histogram equalization		Ridgelet	Yes	NFL	75.25
		Pixel intensities		NFL	67.45
Histogram equalization		Pixel intensities		NFL	69.50
	Exhaustive search	Ridgelet		NFL	78.15
Histogram equalization	Exhaustive search	Ridgelet		NFL	77.35
	Exhaustive search	Pixel intensities	–	NFL	76.45
Histogram equalization	Exhaustive search	Pixel intensities	–	NFL	76.84
		Yes		NFL	69.93
Histogram equalization	–	Yes	–	NFL	70.21
–	Exhaustive search	Yes	–	NFL	72.65
Histogram equalization	Exhaustive search	Yes	–	NFL	74.55
–	–	Pixel intensities	–	NN	20.34
Histogram equalization	–	Pixel intensities	–	NN	23.14
–	Exhaustive search	Pixel intensities	–	NN	32.12
Histogram equalization	Exhaustive search	Pixel intensities	–	NN	37.89
–	–	Ridgelet	–	NN	43.23
Histogram equalization	–	Ridgelet	–	NN	45.97
–	–	Ridgelet	Yes	NN	33.25
Histogram equalization	–	Ridgelet	Yes	NN	33.46
–	–	Wavelet	–	NN	37.67
Histogram equalization	–	Wavelet	–	NN	39.52

Table 8
Accuracy (%) of the combination of SHMM/wavelet face and fingerprint identifiers using different fusion methods with and without histogram equalization

Fusion methods	No hist. equalization	Hist. equalization
Max Z-score	94.00	95.625
Sum of Z-scores	80.00	57.186
Raw score	3.00	10.625
Borda count	98.00	95

PCA and ICA. However, in the fingerprint identification task, it appears that the accuracy of the ridgelet/SVM pair isvery close to the DWT/SHMM pair. Furthermore, these experimental results revealed a slight improvement of the bimodal biometric system (using the Borda count fusion technique) over the best classifiers using a single modality. Our future work consists of: (i) improving the classifier combination techniques by weighting both classifiers differently, and (ii) studying the impact of the fusion accuracy with respect to the database quality for robustness.

References

[1] D. Maltoni, D. Maio, A.K. Jain, S. Prabhakar, Handbook of Fingerprint Recognition, Springer Press, New York, 2003.
[2] W. Zhao, R. Chellappa, P.J. Phillips, A. Rosenfeld, Face recognition: a literature survey, ACM Comput. Surv. 35 (4) (2003) 399–458.
[3] R. Gross, S. Baker, I. Matthews, T. Kanade, Face recognition across pose and illumination, in: S.Z. Li, A.K. Jain (Eds.), Handbook of Face Recognition, Springer-Verlag, Berlin, 2004.
[4] G. Lawton, Biometrics: a new era in security, IEEE Comput. (1998) 16–18.
[5] L. Torres, Is there any hope for face recognition?, in: WIAMIS, 2004, pp. 2709–2712.

[6] L.R. Rabiner, A tutorial on hidden markov models and selected applications in speech recognition, 1990, pp. 267–296.
[7] A. Nefian, M. Hayes, Hidden markov models for face recognition, in: ICASSP98, 1998, pp. 2721–2724.
[8] L. Bai, L. Shen, Combining wavelets with hmm for face recognition, in: Proceedings of SGAI, 23rd International Conference on Innovative Techniques and Applications of Artificial Intelligence, 2003.
[9] I. Daubechies, Wavelet transforms and orthonormal wavelet bases, in: Different Perspectives on Wavelets, San Antonio, TX, 1993, Proceedings of the Symposium Applications Mathematics, vol. 47, American Mathematical Society, Providence, RI, 1993, pp. 1–33.
[10] U.M. Bicego, V. Murino, Using hidden markov models and wavelets for face recognition, in: ICIAP03, 2003, pp. 52–56.
[11] H.S. Le, H. Li, Recognizing frontal face images using hidden markov models with one training image per person, ICPR, vol. 1, 2004, pp. 318–321.
[12] F. Samaria, Face recognition using hidden markov models, Ph.D. Thesis, Cambridge University Engineering Department, 1994.
[13] H. Othman, T. Aboulnasr, A separable low complexity 2d hmm with application to face recognition, IEEE Trans. Pattern Anal. Mach. Intell. 25 (10) (2003) 1229–1238.
[14] A. Nefian, M. Hayes, Maximum likelihood training of the embedded hmm for face detection and recognition, in: IEEE International Conference on Image Processing, 2000.
[15] D. Bouchaffra, J. Tan, Introduction to structural hidden markov models: application to handwritten numeral recognition, Intell. Data Anal. J. 10 (1) (2006).
[16] D. Bouchaffra, J. Tan, Structural hidden markov models using a relation of equivalence: application to automotive designs, Data Min. Knowl. Discovery J. 12 (1) (2006).
[17] A.K. Jain, S. Parbhakar, L. Hong, S. Pankanti, Filterbank-based fingerprint matching, IEEE Trans. Image Process. 9 (5) (2000) 846–859.
[18] A. Jain, Y. Chen, M. Demirkus, Pores and ridges: high-resolution fingerprint matching using level 3 features, IEEE Trans. Pattern Anal. Mach. Intell. 29 (1).
[19] S. Andrew, A hidden markov model fingerprint classifier, in: Proceedings of the 31st Asilomar Conference on Signals, Systems and Computers, 1997, pp. 306–310.
[20] J. Kittler, M. Hatef, R. Duin, J.G. Mattas, On combining classifiers, IEEE Trans. PAMI 20 (3) (1998) 226–239.
[21] P. Verlinde, G. Cholet, Comparing decision fusion paradigms using k-nn based classifiers, decision trees and logistic regression in a multimodal identity verification application, in: Proceedings of the Second International Conference on AVBPA, Washington, DC, 1999, pp. 188–193.
[22] U. Dieckmann, P. Plankensteiner, T. Wagner, Sesam: a biometric person identification system using sensor fusion, Pattern Recognition Lett. 18 (9) (1997) 827–833.
[23] S. Mallat, A theory for multiresolution signal decomposition: the wavelet representation, IEEE Trans. Pattern Anal. Mach. Intell. 11 (1989) 674–693.
[24] E.J. Candes, Ridgelets: theory and application: Ph.D. Thesis, Master's Thesis, Department of Statistics, Stanford University, 1998.
[25] E.J. Stollnitz, T.D. Derose, D.H. Salesin, Wavelets for computer graphics: a primer, part 1, IEEE Comput. Graphics Appl. 15 (3) (1995) 76–84.
[26] F. Samaria, A. Harter, Parameterisation of a stochastic model for human face identification, in: IEEE Workshop on Applications of Computer Vision, Sarasota (Florida), 1994.
[27] M.N. Do, Matlab frit toolbox: the finite ridgelet transform for image representation, in: ⟨http://www.ifp.uiuc.edu/minhdo/software⟩.
[28] E.J. Candes, D.L. Donoho, A Surprisingly Effective Nonadaptive Representation for Objects With Edges: Curves and Surfaces, Vanderbilt University Press Edition, 2000.
[29] E.D. Bolker, The finite radon transform, Contemp. Math. 63 (1987) 27–50.
[30] F. Matus, H. Flusser, Image representation via a finite radon transform, IEEE TPAMI 15 (1993) 993–1006.
[31] L. Rabiner, B.-H. Juang, Fundamentals of Speech Recognition, Prentice-Hall, Inc., Upper Saddle River, NJ, USA, 1993.
[32] D. Hond, L. Spacek, Distinctive descriptions for face processing, in: BMVC, 1997, pp. 320–329.
[33] P.C. Yuen, J.H. Lai, Face representation using independent component analysis, Pattern Recognition 35 (6) (2002) 1247–1257.
[34] P. Nicholl, A. Amira, R. Perrot, An automated grid-enabled face recognition system using hybrid approaches, in: Proceedings of the Fifth IEE/IEEE Postgraduate Research Conference on Electronics, Photonics, Communications and Networks (PREP05), 2005, pp. 144–146.
[35] O. Ayinde, Y.H. Yang, Face recognition approach based on rank correlation of gabor-filtered images, Pattern Recognition 35 (6) (2002) 1275–1289.
[36] Histogram equalization: the csharp toolbox, in: ⟨http://codersource.net/csharp_histogram⟩.
[37] S.Z. Li, J. Lu, Face recognition using the nearest feature line method, IEEE Trans. Neural Networks 10 (2) (1999) 439–443.
[38] A. Ross, A. Jain, Information fusion in biometrics, Pattern Recognition Lett. 24 (2003) 2115–2125.
[39] M.V. Erp, L. Schomaker, Variants of the borda count method for combining ranked classifier hypotheses, in: Proceedings of the Seventh International Workshop on Frontiers in Handwriting Recognition, Amsterdam, 2000, pp. 443–452.
[40] A. Jain, K. Nandakumar, A. Ross, Score normalization in multimodal biometric systems, J. Pattern Recognition 38 (1) (2005) 2270–2285.
[41] D. Bouchaffra, V. Govindaraju, S. Srihari, A methodology for mapping scores to probabilities, IEEE TPAMI 21 (9) (1999).

About the Author—Dr. DJAMEL BOUCHAFFRA is an associate professor of Computer Science and Mathematics at Grambling State University. He previously held a position of assistant professor at Oakland University where he was teaching Pattern Recognition, Machine Learning, Artificial Intelligence, Soft Computing, and Discrete Mathematics. Dr. Bouchaffra has been selected as the recipient of the Oakland University Teaching Excellence Award for 2004 as well as the recipient of the School of Engineering Teaching Excellence Award. This recognition acknowledges the genuine admiration of his students and his colleagues for his superior level of professionalism in teaching. His personal views on the art of teaching are published in this Oakland University Newsletter link. On February 2005, he was invited by OU-TV "Focus on Faculty Program" broadcasted on Comcast cable channel 74 (interview link).

His fields of research are Pattern Recognition, Machine Learning, and Adaptive Artificial Intelligence. He is currently working on the development of mathematical models that are capable to: (i) merge statistics and structure (including shape) together within a single Bayesian framework, and (ii) use this fusion to perform adaptive classification of discrete-valued time series. Professor Bouchaffra introduced both the structural and the topological hidden Markov models as two paradigms that perform this fusion. His areas of application include (but are not limited to) biometrics, bioinformatics, speech, handwriting recognition, language modeling, and data mining. He is the director of the Pattern Recognition and Machine Intelligence Laboratory (PARMIL).

He has written several papers in peer-reviewed conferences and premier journals. He chaired several sessions in conferences; he is one of the general chairs of the conference ICSIT'05 (www.icsit.org). He is a regular reviewer for some funding agencies and several journals such as IEEE TPAMI, TNN, TKDE, and Image Processing.

Professor Bouchaffra is a member of the editorial board of the Journal of Pattern Recognition published by Elsevier and an associate editor for the Journal titled "Advances in Artificial Intelligence" published by Hindawi Corporation and the Journal of Engineering Letters published by the International Association of Engineers. Dr. Bouchaffra is a senior member of the IEEE, a member of the IEEE Computer Society and the vice chair of the IEEE/SEM Chapter V.

About the Author — AMIRA is a senior lecturer in Electronic and Computer Engineering in the School of Engineering and Design, Brunel University (from May 2006). Previously he held a lectureship in Computer Science at Queen's University, Belfast (from November 2001). He received his Ph.D. in Computer Science from Queen's University Belfast in 2001. He has been awarded a number of grants from government and industry and has published over 100 publications during his career to date. He was leading the medical imaging and embedded systems using Field Programmable Gate Arrays (FPGAs) research at the Institute of Electronics, Communications and Information Technologies (ECIT), QUB, Belfast. He has been a Ph.D. external/internal examiner for HKUST (Hong Kong), Trinity College Dublin, Queen's University Belfast, Edinburgh and Essex Universities. He has been invited to give a talk at the IEEE CMOS Emerging Technologies Workshop COMSET06, Banff, about his future views of reconfigurable computing with the different emerging technologies. He has been invited to give talks at universities in UK, Europe, USA, and North Africa, at international conferences, workshops and exhibitions (CeBIT02), and contributed to many internationally recognized conferences (DELTA08, Program Chair), (IMVIP2005, Conference Chair), (MWSCAS03, Session Chair), program committee for ParaFPGA07, AHS07, SIGMAP07, IMVIP05, etc. He is a referee for IEEE, IEE, ACM conferences, and journals (ISCAS, ICASSP, FPL, TIM, TSP). He is a senior member of IEEE, member of ACM, and a Fellow of the Higher Education Academy. His research interests include digital image and signal processing: design and implementation, face recognition technologies, medical image analysis, multiresolution analysis, biometrics, custom computing and low power design using FPGAs, System on Chip, and information retrieval.