1

A Review of Hidden Markov Models in Face Recognition

Claudia Iancu and Peter M. Corcoran
College of Engineering & Informatics
National University of Ireland Galway
Ireland

1. Introduction

Hidden Markov Models (HMMs) are a set of statistical models used to characterize the statistical properties of a signal. An HMM is a doubly stochastic process with an underlying stochastic process that is not observable, but can be observed through another set of stochastic processes that produce a sequence of observed symbols. An HMM has a finite set of states, each of which is associated with a multidimensional probability distribution; transitions between these states are governed by a set of probabilities. Hidden Markov Models are especially known for their application in 1D pattern recognition such as speech recognition, musical score analysis, and sequencing problems in bioinformatics. More recently they have been applied to more complex 2D problems and this review focuses on their use in the field of *automatic face recognition*, tracking the evolution of the use of HMMs from the early-1990's to the present day.
Our goal is to enable the interested reader to quickly review and understand the state-of-art for HMM models applied to face recognition problems and to adopt and apply these techniques in their own work.

2. Historical overview and Introduction to HMM

The underlying mathematical theory of Hidden Markov Models (HMMs) was originally described in a series of papers during the 1960's and early 1970's [Baum & Petrie, 1966; Baum et al., 1970; Baum, 1972]. This technique was subsequently applied in practical pattern recognition applications, more specifically in speech recognition problems [Jelinek et al., 1975]. However, widespread understanding and practical application of HMMs only began a decade later, in the mid-1980s. At this time several tutorials were written [Levinson et al., 1983; Juang, 1984; Rabiner & Juang, 1986; Rabiner, 1989]. The most comprehensive of these was the last, [Rabiner, 1989], and provided sufficient detail for researchers to apply HMMs to solve a broad range of practical problems in speech processing and recognition. The broad adoption of HMMs in automatic speech recognition represented a significant milestone in continuous speech recognition problems [Juang & Rabiner, 2005].
The mathematical sophistication of HMMs combined with their successful application to a wide range of speech processing problems has prompted researchers in pattern recognition to consider their use in other areas, such as character recognition, keyword spotting, lip-

reading, gesture and action recognition, bioinformatics and genomics. In this chapter we present a review of the most important variants of HMMs found in the *automatic face recognition literature*. We begin by presenting the initial 1D HMM structures adapted for use in face recognition problems in section 3. Then a number of papers on hybrid approaches used to improve the performance of HMMs for face recognition are discussed in section 4. In section 5 the various 2D variants of HMM are described and evaluated in terms of the recognition rates achieved from each. Finally section 6 includes some recent refinements in the application of HMM techniques to face recognition problems.

3. HMM in face recognition - initial 1D HMM structures

As mentioned in the previous section, HMMs have been used extensively in speech processing, where signal data is naturally one-dimensional. Nevertheless, HMM techniques remain mathematically complex even in the one-dimensional form. The extension of HMM to two-dimensional model structures is exponentially more complex [Park & Lee, 1998]. This consideration has led to a much later adoption of HMM in applications involving two-dimensional pattern processing in general and face recognition in particular.

3.1 Initial research on ergodic and top-to-bottom 1D HMM

In 1993, a new approach to the problem of automatic face recognition based on 1D HMMs was proposed by [Samaria & Fallside, 1993]. In this paper faces are treated as two-dimensional objects and the HMM model automatically extracts statistical facial features. For the automatic extraction of features, a 1D observation sequence is obtained from each face image by sampling it using a sliding window. Each element of the observation sequence is a vector of pixel intensities (or greyscale levels).

Two simple 1D HMMs were trained by these authors in order to test the applicability of HMMs in face recognition problems. A test database was used comprising images of 20 individuals with a minimum of 10 images per person. Images were acquired under homogeneous lighting against a constant background, and with very small changes in head pose and facial expressions. For a first set of tests an ergodic HMM was used. The images were sampled using a rectangular window, size 64 × 64, moving left-to-right horizontally with a 25% overlap (16 pixels), then vertically with 16 pixels overlap and starting again horizontally right-to-left. Using the observation sequence thus extracted, an 8-state ergodic HMM was built to approximately match the 8 distinct regions that seem to appear in the face image (eyes, mouth, forehead, hair, background, shoulders and two extra states for boundary regions). Figure 1 taken from [Samaria & Fallside, 1993] shows the training data used for one subject and the mean vectors for the 8 states found by HMM for that particular subject.

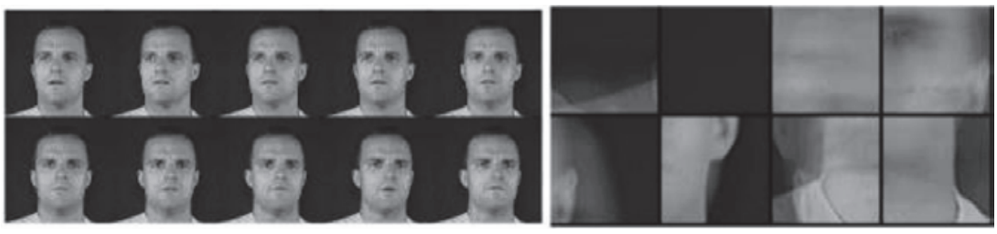


Fig. 1. Training data and states for ergodic HMM [Samaria & Fallside, 1993]

In the second set of tests, a left-to-right (top-to-bottom) HMM was used. Each image was sampled using a horizontal stripe 16 pixels high and as wide as the image, moving top-to-bottom with 12 lines overlap. The resulting observation sequence was used to train a 5-state left-to-right HMM where only transitions between adjacent states are allowed. The training images and the mean vectors for the 5 states found by HMM are presented in Figure 2.

Fig. 2. Examples of training data and states for top-to-bottom HMM from [Samaria & Fallside, 1993]

In both of these models the statistical determination of model features, yields some states of the HMM which can be directly identified with physical facial features. Training and testing were performed using the HTK toolkit[1]. According to these authors, successful recognition results were obtained when test images were extracted from the same video sequence as the training images, proving that the proposed approach can cope with variations in facial features due to small orientation changes, provided the lighting and background are constant. Unfortunately these authors did not provide any explicit recognition rates so it is not possible to compare their methods with later research. It is reasonable, however, to surmise that their experimental results were marginal and are improved upon by the later refinements of [Samaria & Harter, 1994].

3.2 Refinement of the top-to-bottom 1D HMM

In a later paper [Samaria & Harter, 1994] refined the work begun in [Samaria & Fallside, 1993] on a top-to-bottom HMM. These new experiments demonstrate how face recognition rates using a top-to-bottom HMM vary with different model parameters. They also indicate the most sensible choice of parameters for this class of HMM. Up until this point, the parameterization of the model had been based on subjective intuition.

For such a 1D top-to-bottom HMM there are three main parameters that affect the performance of the model: the height of the horizontal strip used to extract the observation sequence, L (in pixels), the overlap used, M (in pixels) and the number of states N of the HMM. The height of the strip, L, determines the size of the features and the length of the observation sequence, thus influencing the number of states. The overlap, M, determines how likely feature alignment is and also the length of the observation sequence. A model with no overlap would imply rigid partitioning of the faces with the risk of cutting across potentially discriminating features. The number of states, N, determines the number of features used to characterize the face, and also the computational complexity of the system.

These experiments were performed using the Olivetti Research Lab (ORL) database, containing frontal facial images with limited side movements and head tilt. The database was comprised of 40 subjects with 10 pictures per subject. The experiments used 5 images

[1] http://htk.eng.cam.ac.uk/

per person for training and the remaining 5 images for testing. The results were reported as error rates, calculated as the proportion of incorrectly classified images. Three sets of tests were done, varying the values of each of the three parameters as follows: $2 \leq N \leq 10$, $1 \leq L \leq 10$ and $0 \leq M \leq L-1$. For M varied, the number of states was fixed at N = 5 and window height L was varied between 2 and 10. According to the tests, the error rates drop as the overlap increases, approximately from 28% to 15%. However a greater overlap implies a bigger computational effort. When L was varied, N was fixed to 5 and the overlaps considered were 0, 1 and L-1. In this case if there is little or no overlap, the smaller the strip height the lower the error rate is, with values between 13% for L = 1 up to 28% for L = 10. However, for sufficiently large overlap the strip height has marginal effect on the recognition performance, the error rate remaining almost constant around 14%. In the third set of tests N was varied, with L = 1 and 0 overlap and L = 8 and maximum overlap (M=L-1). The performance is fairly uniform for values of N between 4 and 10, with an increase in error for values smaller than three.

The conclusions of this paper are: (i) a large overlap in the sampling phase (the extraction of observation sequences) yields better recognition rates; the error rate varies from up to 30% for minimum overlap down to 15% for maximum overlap; (ii) for large overlaps the height of the sampling strip has limited effect. The error rate remains almost constant at 15% for maximum overlap, regardless of the value of L, and (iii) best results are obtained with a HMM with 4 or more states. Error rate drops from around 25% for 1-2 states to 15% from 4 states onward. We remark that these early models were relatively unsophisticated and were limited to fully frontal faces with images taken under controlled background and illuminations conditions.

3.3 1D HMM with 2D-DCT features for face recognition

In [Nefian & Hayes, May 1998], Samaria's version of 1D HMM, is upgraded using 2D-DCT feature vectors instead of pixel intensities. The face image is divided into 5 significant regions, *viz*: hair, forehead, eyes, nose, and mouth. These regions appear in a natural order, each region being assigned to a state in a top-to-bottom 1D continuous HMM. The state structure of the face model and the non-zero transition probabilities are shown in Figure 3.

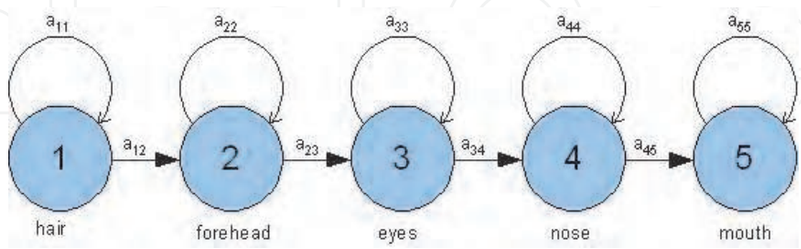


Fig. 3. Sequential HMM for face recognition

The feature vectors were extracted using the same technique as in [Samaria & Harter, 1994]. Each face image of height H and width W is divided into overlapping strips of height L and width W, the amount of overlap between consecutive strips being P, see Figure 4. The number of strips extracted from each face image determines the number of observation vectors.

The 2D-DCT transform is applied on each face strip and the observation vectors are determined, comprising the first 39 2D-DCT coefficients. The system is tested on ORL[2] database containing 400 images of 40 individuals, 10 images per individual, image size 92 × 112, with small variations in facial expressions, pose, hair style and eye wear. Half of the database is used for training and the other half is used for testing. The recognition rate achieved for L=10 and P=9 is 84%. Results are compared with recognition rates obtained using other face recognition methods on the same database: recognition rate for the eigenfaces method is 73%, and for the 1D HMM used by Samaria is also 84%, but the processing time for DCT based HMM is an order of magnitude faster - 2.5 seconds in contrast to 25 seconds required by the pixel intensity method of [Samaria & Harter, 1994].

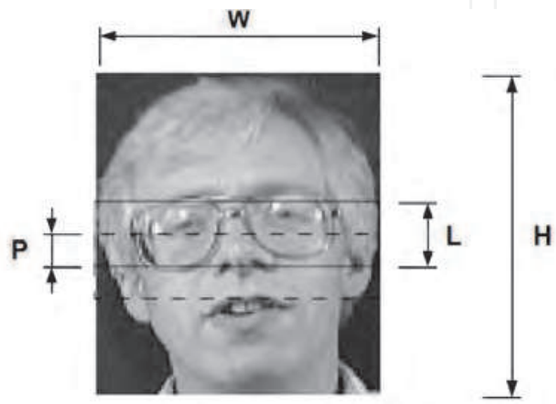


Fig. 4. Face image parameterization and blocks extraction [Nefian & Hayes, May 1998].

3.4 1D HMM with KLT features for face detection and recognition

In a second paper [Nefian & Hayes, October 1998] introduce an alternative 1D HMM approach, which performs the face detection function in addition to that of face recognition. This employs the same topology and structure as in the previous work of these authors, described above, but uses different image features. In contrast with the previous paper, the observation vectors used here are the coefficients of Karhunen-Loeve Transform. The KLT compression properties as well as its decorrelation properties make it an attractive technique for the extraction of the observation vectors. Block extraction from the image is achieved in the same way as in the previous paper. The eigenvectors corresponding to the largest eigenvalues of the covariance matrix of the extracted vectors form the KLT basis set. If μ is the mean of the vectors used to compute the covariance matrix, a set of vectors is obtained by subtracting this mean from each of the vectors corresponding to a block in the image. The resulting set of vectors is then projected onto the eigenvectors of the covariance matrix and the resulting coefficients form the observation vectors.

The system is used both for face detection and recognition by the authors. For face detection, the system is first trained with a set of frontal faces of different people taken under different illumination conditions, in order to build a face model. Then, given a test image, face detection begins by scanning the image with horizontally and vertically overlapping rectangular windows, extracting the observation vectors and computing the probability of

[2] http://www.cl.cam.ac.uk/research/dtg/attarchive/facedatabase.html

data inside each window given the face model, using Viterbi algorithm. The windows that have face model likelihood higher than a threshold are selected as possible face locations. The face detection system was tested on MIT database with 48 images of 16 people with background and with different illuminations and head orientations. Manually segmented faces from 9 images were used for training and the remaining images for testing, with a face detection rate of 90%.

For face recognition this system was applied to the ORL database containing 400 images of 40 individuals, 10 images per individual, at a resolution of 92 × 112 pixels, with small variations in facial expressions, pose, hairstyle and eye wear. The system was trained with half of the database and tested with the other half. The accuracy of the system presented in this paper is increased slightly over earlier work to 86% while the recognition time decreases due to use of the KLT features.

3.5 Refinements to 1D HMM with 2D-DCT features

Following on the work of [Samaria, 1994] and [Nefian, 1999], Kohir & Desai wrote a series of three papers using the 1D HMM for face recognition problems. In a first paper, [Kohir & Desai, 1998], these authors present a face recognition system based on 1D HMM coupled with 2D-DCT coefficients using a different approach for feature extraction than that employed by [Nefian & Hayes, May 1998 & October 1998]. The extracted features are obtained by sliding square windows in a raster scan fashion over the face image, from left to right and with a predefined overlap. At every position of the window over the image (called sub-image) 2D DCT are computed, and only the first few DCT coefficients are retained by scanning the sub-image in a zigzag fashion. The zigzag scanned DCT coefficients form an observation vector. The sliding procedure and the zigzag scanning are illustrated in Figure 5 [Kohir & Desai, 1998].

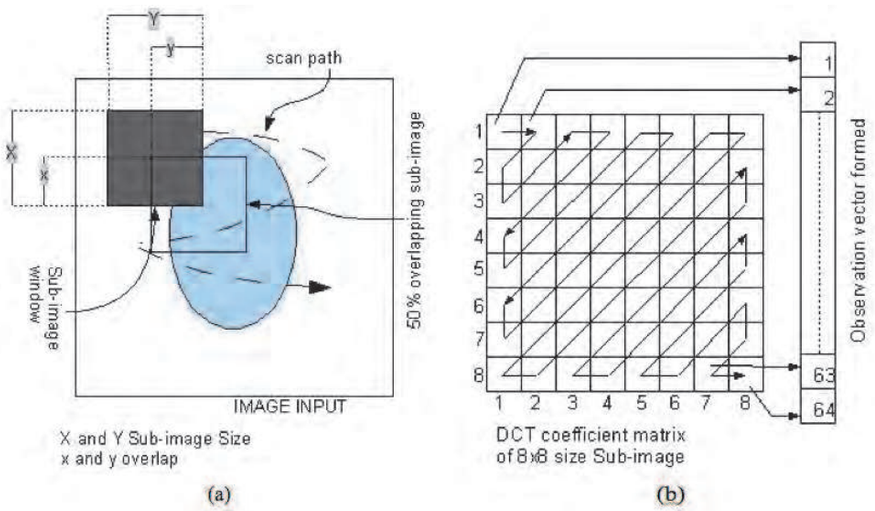


Fig. 5. (a) Raster scan of face image with sliding window. (b) Construction of 1D observation vector from zigzag scanning of the sliding window [Kohir & Desai, 1998].

The performance of this system is tested using the ORL database. Half of the images were used in the training phase and the other half for testing (5 faces for training and the remaining 5 for testing), sampling windows of 8 × 8 and 16 × 16, were used with 50% and

75% overlaps, and 10, 15 and 21 DCT coefficients were extracted. The number of states in the HMM was fixed at 5 as per the earlier work of [Nefian & Hayes, May 1998]. The recognition rates vary from 74.5% for a 16 × 16 window, with a 50% overlap and 21 DCT coefficients to 99.5% for 16 × 16 window, 75% overlap and 10 DCT coefficients.

In a second paper [Kohir & Desai, 1999] these authors further refined their research contribution. To evaluate the recognition performances of the system, 2 new experiments are performed:

- In a first experiment the proposed method is tested with different numbers of training and testing faces per subject. The tests were performed on the ORL database, and the number of training faces was increased from 1 to 6, while the remaining faces were used in the testing phase. A sampling window of 16 × 16 with 75% overlap was used with 10 DCT coefficients as these had provided optimal recognition rates in their earlier work. The recognition rates achieved are from 78.33% for a single training image and 9 testing images up to 99.5% which is the rate obtained when 5 or 6 training images and 5 or 4 testing images are used. It is worth noting that the ORL database comprises frontal face images in uniform lighting conditions and that recognition rates close to 100% are often achieved when using such datasets.
- In a second experiment the system was tested while increasing the number of states in the HMM. Again the ORL database is used, with 5 images for training and 5 for testing. The recognition rates vary as follows: 92% for a 2-states HMM, increasing to 99.5% for a 5-states HMM and stabilizing around 97%-98% when using up to 17 states. The system was also tested with the SPANN database[3] containing 249 persons, each with 7 pictures, with variations in pose, 3 pictures were used for training and the remaining 4 for testing, and the recognition rate achieved was 98.75%.
- A third paper, [Kohir & Desai, 2000] describes the same 1D HMM with DCT features, with a variation in the training phase. In this paper, first a *mean image* is constructed from all the training images, and then each training image is subtracted from the *mean image* to obtain a *mean subtracted image*. The observation vectors are extracted from these *mean subtracted images* using the same window sliding method. The observation vector sequences are then clustered using the K-means technique, and thus an initial state segmentation is obtained. Subsequently, the conventional training steps are followed. In the recognition phase, each test image is first subtracted from the *mean image* obtained during the training phase and recognition is performed on the resulting *mean subtracted image*.

The experiments for face recognition were performed on the same two databases, ORL and SPANN. For ORL database 5 pictures were used for training and the remaining 5 for testing, and the recognition rate obtained is 100%, compared to 88% when the eigenfaces method is used. For SPANN database 3 pictures were used for training and the remaining 4 for testing, the obtained recognition rate was 90%, compared again with the eigenfaces method where a 77% recognition rate was achieved. For the ORL database different resolutions were also tested, the highest recognition rate, 100% being obtained for 96 × 112.

Also, 'new subject rejection' for authentication applications was tested on the ORL database. The database was segmented into 2 sets: 20 subjects corresponding to an 'authorized' subject class - 5 pictures used in training phase and the rest in the testing phase. The

[3] http://www.khayal.ee.iitb.ernet.in/usr/SPANN_DATA_BASE/2D_Signals/Face/faces

remaining 20 subjects are assigned to an 'unauthorized' class - all 10 pictures are used in the testing phase. For each 'authorized' subject a HMM model is built. Also a separate 'common HMM' model is built using all *mean subtracted training images* of all the 'authorized' subjects. For each test face, if the probability of the 'common HMM' is the highest, the input face image is rejected as 'unauthorized', otherwise the input face image is treated as 'authorized'. The results are: 100% rejection of any new subjects and 17% rejection of known subjects (false negatives).

3.6 Refinement of 1D HMM with sequential prunning

As proved by [Samaria & Harter, 1994], the number of states used in a 1D HMM can have a strong influence on recognition rates. The problem of the optimal selection of the structure for an HMM is considered in [Bicego at al., 2003a]. The first part of this paper presents a method of improving the determination of the optimal number of states for an HMM. These authors then proceed to prove the equivalence between (i) a 1D HMM whose observation vectors are modelled with *multiple Gaussians per state* and (ii) a 1D HMM with *one Gaussian per state* but employing a larger number of states. According to the authors, there are several possible methods for solving the first problem, e.g. cross-validation, Bayesian inference criterion (BIC), minimum description length (MDL). These are based on training models with different structures and then choosing the one that optimizes a certain selection criterion. However, these methods involve a considerable computational burden plus they are sensitive to the local-greedy behaviour of the HMM training algorithm, i.e. the successful training of the model is influenced by the initial estimates selected.

The approach proposed by [Bicego et al., 2003a] addresses both the computational burden of model selection, and the initialization phase. The key idea is the use of a decreasing learning strategy, starting each training session from a 'nearly good' situation derived from the previous training session by pruning the 'least probable' state. More specifically, the authors proposed starting the model training with a large number of states. They next run the estimation algorithm and, on convergence, evaluate the model selection criterion. The 'least probable' state is then pruned, and the resulting configuration of the model with one less state is used as a starting point for the next sequence of iterations. In this way, each training session is started from a 'nearly good' estimate. The key observation supporting this approach is that, when the number of states is extremely large, the dependency of the model behaviour on the initial estimates is much weaker. An additional benefit is that using 'nearly good' initializations drastically reduces the number of iterations required by the learning algorithm at each step in this process. Thus the number of model states can be rapidly reduced at low computational cost.

In order to assess the performance of their proposed method, these authors tested the pruning approach and the standard approach (consisting in training one HMM for varying number of states) with BIC criterion and MMDL (mixture minimum description length) [Figueiredo et al., 1999] criterion. These two strategies are compared in terms of: (i) accuracy of the model size estimation, (ii) total computational cost involved in the training phase, and (iii) classification accuracy. In all the HMMs considered in this paper the emission probability density for each state is a single Gaussian. For the accuracy of the model size estimation, synthetically generated test sets of 3 known HMMs were used. The authors set the number of states allowed from 2 to 10. The selection accuracy ranged from 54% to 100% for standard BIC and MMDL, and from 98% to 100% for pruning BIC and MMDL, with up to 50% less iteration required for the latter.

Classification accuracy was tested on both synthetic and real data. For the synthetic data, the test sets used previously to estimate the accuracy of the model size estimation were used, obtaining 92% to 100% accuracy for standard BIC and MMDL compared to 98% to 100% accuracy for pruning BIC and MMDL, with 35% less iterations for pruning. For classification accuracy on real data, two experiments were conducted. The first involves a 2D shape recognition problem, and uses a data set with four classes each with 12 different shapes. The results obtained are 92.5% for standard BIC, 94.37% for standard MMDL, and 95.21% for pruning BIC and MMDL. The second experiment was conducted on the ORL database, using the method proposed by [Kohir & Desai 1998]. The results are 97.5% for standard BIC and MMDL and 97.63% for pruning BIC and MMDL. The classification accuracies are similar, but the pruning method reduces substantially the number of iterations required.

3.7 A 1D HMM with 2D-DCT features and Haar wavelets

In a following paper [Bicego et al., 2003b], a comparison between DCT coding and wavelet coding is undertaken. The aim is to evaluate the effectiveness of HMMs in modelling faces using these two different forms of image features. Each compresses the relevant image data, but employing different underlying techniques. Also, the suitability of HMM to deal with the JPEG 2000 image compression standard is considered by these authors. They adopt the 1D HMM approach introduced by [Kohir & Desai, 1998]. However, the optimum number of states for the model is selected using the sequential pruning strategy presented in [Bicego et al., 2003a] and described in the preceding section. The same feature extraction used by [Kohir & Desai, 1998] is employed, and both 2D DCT and Haar wavelet coefficients are computed.

These experiments have been conducted on the ORL database, consisting of 40 subjects with 10 sample images of each. The first 5 images are used for training the HMM while the remaining 5 are used in the testing phase. The number of states for each HMM is estimated using the pruning strategy. For feature extraction, a 16 × 16 pixel sliding window is used, with 50% and 75% overlaps being tested, and in each case the first 4, 8 and 12 DCT or Haar coefficients are retained. The recognition rate scores for 50% overlap are between 97.4% for 4 coefficients to 100% for 12 coefficients, and for 75% overlap between 95.4% for 4 coefficients to 99.6% for 12 coefficients. Slightly better results were obtained for DCT coefficients throughout the experiments. It is worth noting that unlike [Samaria & Harter, 1994] and [Nefian & Hayes, 1998] in the case of [Kohir & Desai, 1998] the method of extracting observation vectors results in better performance for a 50% overlap than for 75% overlap.

A second experiment was performed to prove the effectiveness of HMM in solving the face recognition problem regardless of the coefficients used, by replacing in the proposed system the wavelet coding with a trivial coding represented by the mean of the square window. The results obtained are 84.9% for 50% overlap and 77.8% for 75% overlap.

4. Hybrid approaches based on 1D HMM

From the discussions of the preceding section it can be seen that 1D HMM can perform successfully in face recognition applications. However, the vast majority of early experiments were performed on the ORL database. The images in this dataset only exhibit very small variations in head pose, facial expressions, facial occlusions such as facial hair and glasses, and almost no variations in illumination. For practical applications a face recognition system must be able to handle significant variations in facial appearance in a

robust manner. Thus in this next section more challenging face recognition applications are described and further HMM approaches are considered from the literature. Specifically, in this section we consider hybrid approaches based on HMMs used successfully in more challenging applications of face recognition.

There are several core problems that a face recognition system has to solve, specifically those of variations in illumination, variations in facial expressions or partial occlusions of the face, and variations in head pose. Firstly an attempt at solving recognition problems caused by facial occlusions is considered [Martinez, 1999]. The solution adopted by this author was to explore the use of *principle component analysis* (PCA) features to characterize 6 different regions of the face and use 1D HMM to model the relationships between these regions. A second group of researchers [Wallhoff et al., 2001] have tackled the challenging task of recognizing side-profile faces in datasets where only frontal faces were used in the training stage. These authors have used a combination of *artificial neural network* (ANN) techniques combined with 1D HMM to solve this challenging problem.

4.1 Using 1D HMM with PCA derived features

A face recognition system is introduced [Martinez, 1999] for indexing images and videos from a database of faces. The system has to tackle three key problems, identifying frontal faces acquired, (i) under differing illumination conditions, (ii) with varying facial expressions and (iii) with different parts of the face occluded by sunglasses/scarves. Martinez's idea was to divide the face into N different regions analyzing each using PCA techniques and model the relationships between these regions using 1D HMMs.

The problem of different lighting conditions is solved in this paper by training the system with a broad range of illumination variations. To handle facial expressions and occlusions, the face is divided into 6 distinct local areas and local features are matched. This dependence on local rather than global features should minimize the effect of facial expressions and occlusions, which affect only a portion of the overall facial region. Each of these local areas obtained from all the images in the database is projected into a primary eigenspace. Each area is represented in vector form. Figure 6 [Martinez, 1999] shows the local feature extraction process.

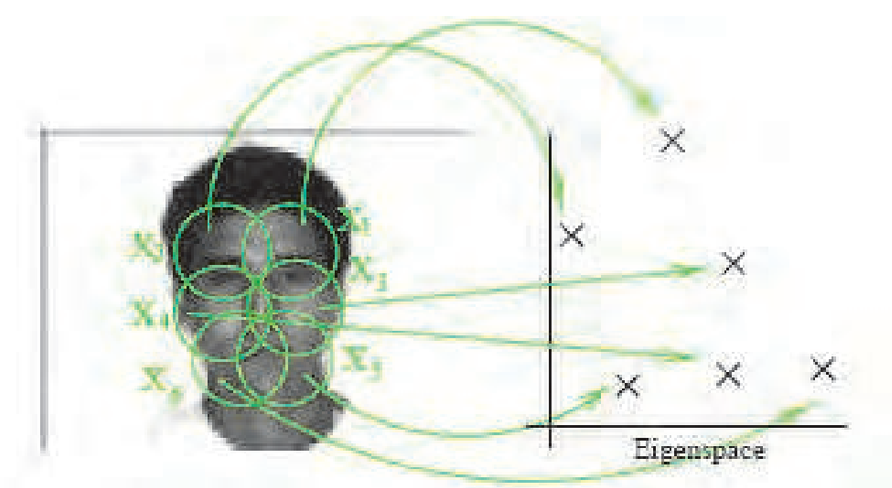


Fig. 6. Projection of the 6 different local areas into a global eigenspace Martinez, 1999].

Note that face localization is performed manually in this research and thus cannot be precise enough to guarantee that the extracted local information will always be projected accurately into the eigenspace. Thus information from pixels within and around the selected local area is also extracted, using a rectangular window. By considering these six local areas as hidden states, a 1D HMM was built for each image in the database. However, a more desirable case is to have a single HMM for each person in the database, as opposed to a HMM for each image. To achieve this, all HMMs of the same person were merged together into a single 1D HMM, where the transition probability from one state to another is *1/number of HMMs per person*. In the recognition phase, instead of using the forward-backward algorithm, the authors used the Viterbi algorithm [Rabiner, 1989] to compute the probability of an observation sequence given a model.
Two sets of tests were performed, using pictures and video sequences. The image database[4] was created by Aleix Martinez and Robert Benavente. It contains over 4,000 colour facial images corresponding to 126 people - 70 men and 56 women. There are 12 images per person, the first 6 frontal view faces with different facial expressions and illumination conditions and the second 6 faces with occlusions (sun-glasses and scarf) and different illumination conditions. These pictures were taken under strictly controlled conditions. No restrictions on appearance including clothing, accessories such as glasses, make-up or hairstyle were imposed on participants. Each person participated in two sessions, separated by 14 days. The same pictures were taken in both sessions. In addition, 30 video sequences were processed consisting of 25 images almost all of them containing a frontal face. Five different tests were run, using 50 people (25 males and 25 females) randomly selected from the database, converted to greyscale images and sampled at half their size, and also using 30 corresponding video sequences. In a first test, all 12 images per person were used in training, and the system was tested with every image by replacing each one of the local features with random noise with mean 0. The recognition rate obtained was 96.83%. For a second test training was with the first six images and testing with the last six images, featuring occlusions. A recognition rate of 98.5% was achieved. In a third test the last six images were used for training and the first six for testing and the resulting recognition rate was 97.1%. A fourth test consisted of training with only two non-occluded images and testing with all the remaining images. A lower recognition rate of 72% was obtained. Finally, the system was trained with all 12 images for each person, and tested with the video sequences, achieving a 93.5% recognition rate.

4.2 Artificial Neural Networks (ANN) in conjunction with 1D HMM

[Wallhoff et al., 2001] approached the challenging task of recognizing profile views with previous knowledge from only frontal views, which may prove a challenging task even for humans. The authors use two approaches based on a combination of Artificial Neural Networks (ANN) and a modelling technique based on 1D HMMs: a first approach uses a synthesized profile view, while a second employs a joint parameter-estimation technique. This paper is of particular interest because of its focus on non-frontal faces. In fact these authors are one of the first to address the concept of training the recognition system with conventional frontal faces, but extending the recognition to include faces with only a side-profile view.

[4] http://www2.ece.ohio-state.edu/~aleix/ARdatabase.html

The experiments are performed on the MUGSHOT[5] database containing the images of 1573 cases, where most individuals are typically represented by only two photographs: one showing the frontal view of the person's face and the other showing the person's right hand profile. The database contains pairs of mostly male subjects at several ages and representatives of several ethnic groups, subjects with and without glasses or beards and a wide range of hairstyles. The lighting conditions and the background of the photographs also change. The pictures in the database are stored as 8-bit greyscale images. Prior to applying the main techniques of [Wallhoff et al., 2001] a pre-processing of each image is conducted. Photographs with unusually high distortions, perturbations or underexposure are discarded; all images are manually labelled so that all faces appear in the centre of the image and with a moderate amount of background, and resized to 64x × 64 pixels. Then two sets are defined: a first set consisting of 600 facial image pairs, *frontal* and right-hand *profile*, are used for training the neural network. A second set with 100 facial image pairs is used for testing. The features used for experiments are pixel intensities. In order to obtain the observation vectors, each image which was resized to 64 × 64 pixels is divided into 64 columns. So from each image 64 observation vectors are extracted. The dimension of the vectors is the number of rows in the image, which is also 64, and these vectors consist of pixel intensities. In the training phase an appropriate neural network is used, estimated by applying the following intuitions: (i) a point in the frontal view will be found in approximately the same row as in the profile view, (ii) considering the right half of the face to be almost bilaterally symmetrical with the left half, only the first 40 columns of the image are used in the input layer to the ANN. Figure 7 taken from [Wallhoff et al., 2001] shows how a frontal view of the face is used to generate the profile view. In the testing phase, a 1D left to right first order HMM is used, allowing self transitions and transitions to the next state only. The models consist of 24 states, plus two non-emitting start and end states.

In the first hybrid approach for face profile recognition there are two training stages. Firstly, a neural network is trained using the first set of 600 images, the frontal image of each individual representing the input and the profile view the output. In this way the neural network is trained to synthesize profiles from the frontal image. In figure 8 [Wallhoff et al., 2001] an example of synthesized profile is shown. In the second training stage, the 100 frontal images are introduced in the neural network and their corresponding profiles are synthesized. Using these profiles, an average profile HMM model is obtained. Then for each testing profile, an HMM model is built using for initialization the average profile model. The Baum-Welch estimation procedure is used for training the HMM.

In a second approach only one training stage is performed, the computation speed being vastly improved as a result. This proceeds as follows: the NN is trained using the frontal images as input; the target outputs are in this case the mean values of each Gaussian mixture used for describing the observations of the corresponding profile image. First, an average profile HMM model is obtained using the 600 training profile images. Using this average model, the mean values for each individual in the training set are computed and used as the target values for the NN to be trained. In the recognition phase, for each frontal face the mean value for profile is returned by the NN. Using this mean and the average profile model, the corresponding HMM is built, then the probability of the test profile image given the HMM model is computed. The recognition rates achieved for the

[5] http://www.nist.gov/srd/nistsd18.cfm

systems proposed in this paper are around 60% for the first approach and up to 49% for the second approach, compared to 70%-80% when humans perform the same recognition task. The approach presented by the authors is very interesting in the context of a mugshot database, where only the two instances, one *frontal* and one *profile* of a face are present. Also the results are quite impressive compared to the human recognition rates reported. However, both ANN and HMM are computationally complex, and using pixel intensities as features also contributes to making this approach very greedy in terms of computing resources.

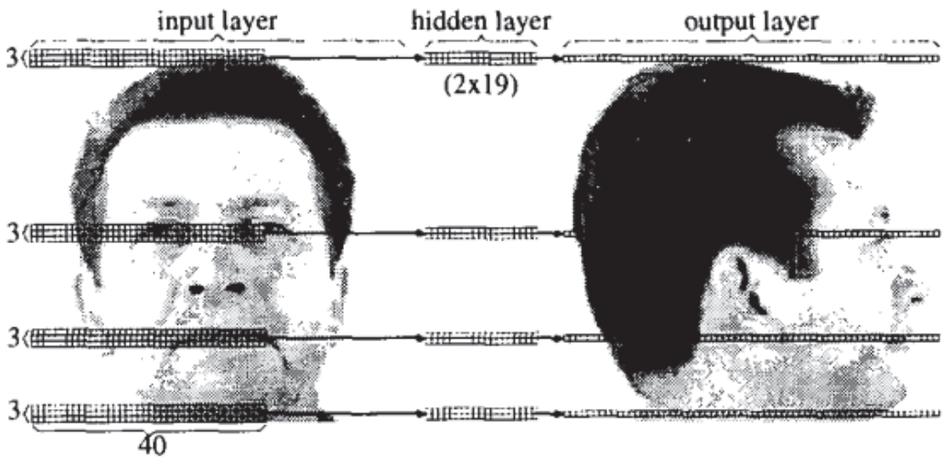


Fig. 7. Generation of a profile view from a frontal view [Wallhoff et al., 2001].

Fig. 8. Example of frontal view, generated and real profile [Wallhoff et al., 2001].

5. 2D HMM approaches

In section 3 and section 4 we showed how 1D HMMs might be adapted for use in face recognition applications. But face images are fundamentally 2D signals and it seems intuitive that they would be more effectively processed with a 2D recognition algorithm. Note however that a fully connected 2D extension of HMM exhibits a significant increase in computational complexity making it inefficient and unsuitable for practical face recognition applications [Levin & Pieraccini, 1992]. As a consequence of this complexity of the full 2D HMM approach a number of simpler structures were developed and are discussed in detail in the following sections.

5.1 A first application of pseudo 2D HMM to Facial Recognition

In his PhD thesis, [Samaria, 1994] was the first researcher to use *pseudo-2D* HMMs in face recognition, with pixel intensities as features. In order to obtain a P2D HMM, a one-dimensional HMM is generalized, to give the appearance of a two-dimensional structure, by allowing each state in a one-dimensional global HMM to be a HMM in its own right. In this way, the HMM consists of a top-level set of super states, each of which contains a set of embedded states. The super states may then be used to model the two-dimensional data in one direction, with the embedded HMMs modelling the data along the other direction. This model is appropriate for face images as it exploits the 2D physical structures of a face, namely that a face preserves the same structure of states from top to bottom – forehead, eyes, nose, mouth, chin, and also the same left-to-right structure of states inside each of these super states. An example of state structure for the face model and the non-zero transition probabilities of the P2D HMM are shown in figure 9. Each state in the overall top-to-bottom HMM is assigned to a left-to-right HMM.

Fig. 9. Structure of a P2D HMM.

In order to simplify the implementation of P2D-HMM, the author used an equivalent 1D HMM to replace the P2D-HMM as shown in figure 10. In this case, the shaded states in the 1D HMM represent end-of-line states with two possible transitions: one to the same row of states - *superstate self-transition* - and one to the next row of states - *superstate to superstate transition*. For feature extraction a square window is used sliding from left-to-right and top-to-bottom. Each observation vector contains the intensity level values of the pixels contained by the window, arranged in a column-vector. In order to accommodate the extra end-of-line state, a white frame is added at the end of each line of sampling. Each state is modelled by one Gaussian with mean and standard deviation set, initialized at the beginning of training, to mid-intensity values for normal states and to white with near zero standard deviation for the end-of-line states. The parameters of the model are then iteratively re-estimated using the Baum-Welch algorithm.

Fig. 10. P2D HMM and its equivalent 1D HMM.

Samaria's experiments were carried out on the ORL database. Different topologies and sampling parameters were used for the P2D-HMM: from 4 to 5 superstates and from 2 to 8 embedded states within each superstate. In addition these experiments considered different sizes of sampling windows with different overlaps ranging from 2 × 2 pixels with 1 × 1 overlap up to 24 × 22 (horizontal × vertical) pixels with 20 × 13 pixels overlap. The highest error rate of 18% was obtained for a 3-5-5-3 P2D-HMM, using a 10 × 8 scanning window with an 8 × 6 overlap, while the smallest error rate of 5.5% was obtained for 3-6-6-6-3 P2D-HMM, with 10 × 8 (and 12 × 8) window and 8 × 6 (and 9 × 6 respectively) overlap. In the same thesis Samaria also tested the standard *unconstrained* P2D HMM, which does not have an end-of-line state. In this case no attempt is made to enforce the fact that the last frame of a line of observations should be generated by the last state of the superstate. The recognition results for the *unconstrained* P2D HMM are similar to those obtained with constrained P2D-HMM, the error rates ranging from 18% to 6%. We remark that Samaria also obtained a 2% error rate for a 3-7-7-5-3 P2D HMM with 12×8 sampling window and 4 × 6 overlap, but considering that for only slightly different overlaps (8 × 6 and 4 × 4) the error rates were 6% and 8.5% respectively, this particular result appears to be a statistical anomaly. It does serve to remind that these models are based on underlying statistical probabilities and that occasional aberrations can occur.

5.2 Refining pseudo 2D HMM with DCT features

In [Nefian & Hayes, 1999] the authors adapted the P2D-HMM developed by [Kuo & Agazzi, 1994] for optical character recognition analysis, showing how it represented a valid approach for facial recognition and detection. These authors renamed this technique as *embedded* HMM. In order to obtain the observation vectors, a set of overlapping blocks are extracted from the image from left to right and top to bottom as shown in figure 11, the observation vector finally consisting of the 6 lower-frequency 2D-DCT coefficients extracted form each image block. Each state in the embedded HMMs is modelled using a single Gaussian.

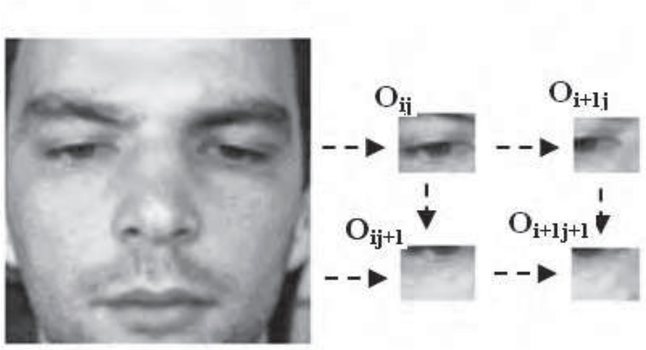


Fig. 11. Face image parameterization and blocks extraction.

For face recognition the ORL database was used. The system was trained with half of the database and tested with the other half. The recognition performance of the method presented in this paper is 98%, improving by more than 10% compared with the best results obtained in using 1D HMM in earlier work [Nefian & Hayes, May 1998, October 1998].

This research also considered the problem of face detection. In the testing phase for detection, 288 images of the MIT database were used, representing 16 subjects with different illuminations and head orientations. A set of 40 images representing frontal views of 40 different individuals from the ORL database is used to train one face model. The testing is performed using a doubly embedded Viterbi algorithm described by [Kuo & Agazzi, 1994]. The detection rate of the system described in this paper is 86%. While this version of HMM appears to be relatively efficient in face detection, it is however computationally very complex and slow, particularly when compared with state of art algorithms [Viola & Jones, 2001].

5.3 Improved initialization of pseudo 2D HMM

Also employing a P2D HMM, [Eickler et al., 2000] describe an advanced face recognition system based on the use of standard P2D HMM employing 2D DCT features is presented. The performance of the system is enhanced using improved initialization techniques and mirror images. It is very important to use a good initial model therefore the authors used all faces in the database to build a 'common initial model'. Then for each person in the database a P2D HMM model is refined using this 'common model'. Feature extraction is based on DCT. The image is scanned with a sliding window of size 8 × 8 from left-to-right and top-to-bottom with an overlap of 6 pixels (75%). The first 15 DCT coefficients are extracted. The use of DCT coefficients allows the system to work directly on images compressed with JPEG standard without a need to decompress these images. The size of the sampling window was chosen as 8 × 8 because the DCT portion of JPEG image compression is based on this window size.

Tests are performed on the ORL database, described previously, with the first 5 images per person used for training and the remaining 5 for testing. Three sets of experiments are performed in this paper. *First Experiment Set:* the system is tested on different quadratic P2D HMM model topologies (4×4 states to 8×8 states) with 1 to 3 Gaussian mixtures to model the probability density functions. The recognition rates achieved range from 81.5% for 4×4 states with 1 Gaussian to 100% for 8 × 8 states with 2 and 3 Gaussians. *Second Experiment Set*: the effect of overlap on recognition rates is tested. An overlap of 75% is used for all training while for testing overlaps between 75% and 0% were used, with a 7×7 HMM and from 1 to 3 Gaussian mixtures. The overall result of this experiment is that recognition rates decrease slightly when the overlap is reduced, however, very good recognition rates of 94.5%-99.5% were still obtained even for 0% overlap, compared with 98.5%-100% for 75% overlap. Thus wide variations in overlap have relatively minor effects on overlap for a sophisticated 7x7 HMM model.

Third Experiment Set: comprises an evaluation of the effect of compression artefacts on the recognition rate. Recognition was performed on JPEG compressed images across a range of quality settings ranging from 100 for the best quality to 1 for the highest compression ratio as shown in figure 12 [Eickler et al., 2000]. The results are as follows: for compression ratios of up to 7.5 to 1, the recognition rates remain constant around 99.5%±0.5%. For compression ratios over 12.5 to 1, the recognition rates drop below 90%, down to approximately 5% for 19.5 to 1 compression ratio.

There are some additional conclusions we can draw from the work of [Eickler et al., 2000]. Firstly, building an initial HMM model using all faces in the database is an improvement over the intuitive initialization used by [Samaria, 1994] or [Nefian & Hayes, 1999], however

this may lead to the dependency of the initial model on the composition of the database. Secondly, these authors obtain excellent results when using JPEG compressed images in the testing phase (overlap 0%), speeding up the recognition process significantly. Note however, in the training stage they use uncompressed images scanned with a 75% overlap and as they have used a very complex HMM model with 49 states the training stage of their approach is resource and time intensive offsetting the benefits of faster recognition speeds.

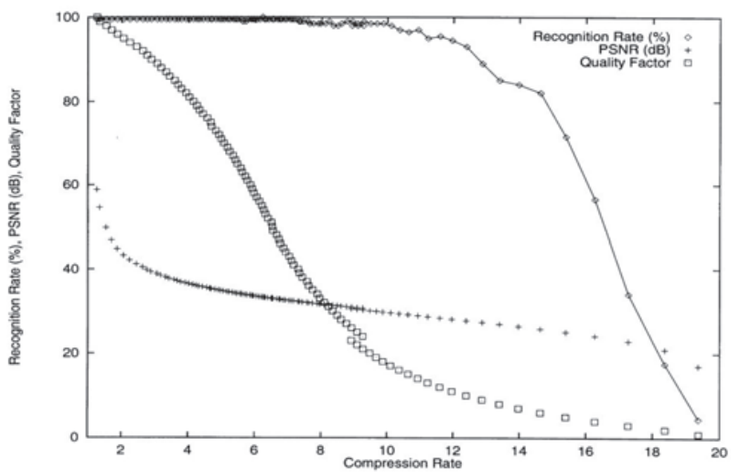


Fig. 12. Recognition rates versus compression rates [Eickler et al., 2000].

5.4 Discrete vs continuous modelling of observation vectors for P2D HMM

In another paper on the subject of face recognition using HMM, [Wallhoff et al., 2001] consider if there is a major difference in recognition performance between HMMs where the observation vectors are modelled as continuous or discrete processes. In the continuous case, the observation probability is expressed as a density probability function approximated by a weighted sum of Gaussian mixtures. In the case of a discrete output probability, a discrete set of observation probabilities is available for each state, and input vector. This discrete set is stored as a set of codebook entries. The codebook is typically obtained by k-means clustering of all available training data feature vectors.

The authors used for their experiments 321 subjects selected from the FERET database[6]. For testing the system, two galleries of images were used: f_a gallery, containing a regular frontal image for each subject, and f_b gallery, containing an alternative frontal image, taken seconds after the corresponding f_a image. First the images are pre-processed, using a semi-automated feature extraction that starts with the manual labelling of the eye and mouth centre-coordinates. The next step is the automatic rotation of the original images so that a line through the eyes is horizontal. After this the face is divided vertically and processing continues on a half-face image. The images are re-sized to the smallest image among the resulting images being 64 × 96 pixels.

For feature extraction, the image is scanned using a rectangular window, with an overlap of 75%. After the DCT coefficients for each block are calculated, a triangular shaped mask is applied and the first 10 coefficients are retained, representing the observation vector. Two

[6] http://www.frvt.org/feret/default.htm

sets of experiments were performed, for continuous and discrete outputs. For the case of *continuous output*, the experiments used 8 × 8 and 16 × 16 scanning windows, and 4 × 4 to 7 × 7 state structures for the P2D HMM. Initially only one Gaussian per state was used. The best recognition rate in this case was 95.95%, for 8 × 8 block size and 7 × 7 states for HMM. When the number of Gaussians was increased form 1 to 3, the recognition rate dropped, maybe due to the fact that only one image per person was used in the training phase. In the case of *discrete output values*, identical scanning windows and HMM were used, and two codebook sizes of 300 and 1000 values were used to generate the observation vectors. The highest recognition rate obtained was 98.13%, for 8 × 8 pixels block size, 7 × 7 states HMM, and a codebook size of 1000. In both cases, continuous and discrete, better results were obtained for the smaller size of scanning window.

5.5 Face retrieval on large databases

After using the combination of 2D DCT and P2D HMM for face recognition on small databases, a new HMM-based measure to rank images within a larger database is next presented, [Eickeler, 2002]. The relation of the method presented to confidence measures is pointed out and five different approximations of the confidence measure for the task of database retrieval are evaluated. These experiments were carried out on the C-VIS database, containing the extracted faces of three days of television broadcast resulting in 25000 unlabeled face images. Normal HMM-based face recognition for database retrieval entails building a model for each person in the database. However, in the case of a very large and unlabeled database, that would imply building a model λ_j for each image O_j in the database, which is not only computationally expensive, but results in poor modelling, considering that a robust model for one person requires multiple training images of that person. In this case, calculating the probability of a query image for each built model $P(O_{query} \mid \lambda_j)$ is simply not practical.

A more feasible method for database retrieval is to train a query HMM λ_{query} using the query images O_{query} of the person searched for ω_{query}, but noting that the probability derived by the Forward-Backward algorithm, $P(O_j \mid \lambda_{query})$ cannot be used as ranking measure for the images in the database because inaccuracies in the modelling of the face images have a big influence on the probability. In order to fix this problem, the ranking of the images uses the query model λ_{query} as a representation of the person being searched for and a set of cohort models Λ_{cohort} representing people not being searched for. An easy way to form the cohort is by using former queries or by taking some images form the database. So instead of calculating $P(O_{query} \mid \lambda_j)$, the probability of an image O_j given the person being searched is used:

$$P(O_j \mid \omega_{query}) \propto P(\lambda_{query} \mid O_j) = \frac{P(O_j \mid \lambda_{Query})}{P(O_j \mid \Lambda_{cohort})} \tag{1}$$

In this research five different confidence measures were used for database retrieval based on this formula. For the confidence measure using normalization, the denominator is replaced:

$$P(O_j \mid \Lambda_{cohort}) = \sum_{\lambda_k \in \Lambda_{cohort}} P(O_j \mid \lambda_k) \tag{2}$$

Another confidence measure uses one filler (common) model instead of a cohort of HMMs for a group of people. The filler model can be trained on all people of the cohort group. If the denominator is set to a fixed probability, it can be dropped from the formula, in which case the confidence measure will be $P(O_j \mid \lambda_{query})$. The fourth confidence measure is based on the sum of ranking differences between the ranking of the cohort models on the query image and the ranking of the cohort models on each of the database images. Finally, the Levenshtein Distance (the Levenshtein distance between two strings is given by the minimum number of operations needed to transform one string into the other) is considered as an alternative measure for the comparison of the rankings of the cohort models for the query image and the database images.

For the experimental part 14 people with 8 to 16 face images each were used as query images, and also as cohort set. A NN-based face detector was used to detect the inner facial rectangles in the video broadcast and the rectangle of each image is scaled to 66 × 86 pixels. In order to remove the background an ellipsoid mask is applied. A P2D HMM with 5 × 5 states is used. The results of the query are evaluated using precision and recall: precision is the proportion of relevant images among the retrieved images while recall is the proportion of relevant images in the database that are part of the retrieval result. In a first experiment a database retrieval for each person of the query set using the normalization is performed and only the precision is calculated considering the database is unlabeled hence an exact number for each person in unknown. For 12 out of 14 people the precision is constant at 100% for around 40 retrieved images (the number of images per person varies between 20 and 300). In a second experiment all five measures were tested for one person. The results are almost perfect for normalization, a little worse but much faster for the filler model. The 'sum of ranking differences' and Levenshtein Distance measures return relatively good results but are inferior to normalization, while the use of a fixed probability gives significantly worse results than all other measures.

5.7 A low-complexity simplification of the Full-2D-HMM

An alternative approach to 2D HMM was proposed by [Othman & Aboulnasr, 2000]. These authors propose a *low-complexity* 2D HMM (LC2D HMM) system for face recognition. The aim of this research is to build a full 2D HMM but with reduced complexity. The challenge is to take advantage of a full 2D HMM structure, but without the full complexity implied by an unconstrained 2D model. Their model is implemented in the 2D DCT compressed domain with 8×8 pixel non-overlapping blocks to maintain compatibility with standard JPEG images. The authors claim a computational complexity reduction from N^4 for a fully connected 2D HMM to $2N^2$ for the LC2D HMM, where N is the number of states. Although the accuracy of the system is not better than other approaches, these authors claim that the computational complexity involved is somewhat less than that required for a 1D HMM and significantly less than that of P2D HMM.

The LC2D HMM is based on 2 key assumptions: (i) the active state at the observation block $B_{k,l}$ is dependant only on immediate vertical and horizontal neighbours, $B_{k-1,l}$ and $B_{k,l-1}$;[7]; (ii) the active states at the 2 observation blocks in anti-diagonal neighbourhood locations, $B_{k-1,l}$ and $B_{k,l-1}$ are statistically independent given the current state. This assumption allows

[7] From a mathematical perspective this assumption is equivalent to a second-order Markov Model, requiring a 3D transition matrix.

separating the 3D state transition matrix into two distinct 2D transition matrices, for horizontal and vertical transitions. This decreases the complexity of the model quite significantly. This *low-complexity* model topology and image scanning are illustrated in figure 13.

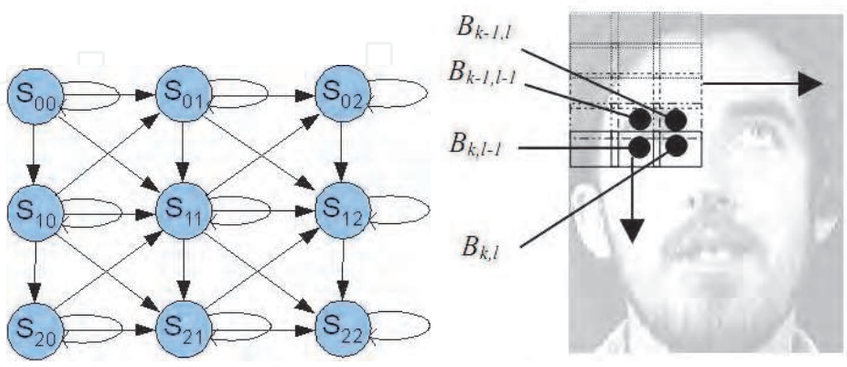


Fig. 13. (a) Image scanning b) Model topology [Othman & Aboulnasr, 2000]

The authors state that the two assumptions are acceptable for non-overlapped feature blocks, but have less validity for very small sized feature blocks or as the allowable overlap increases. The tests were performed on the ORL database. The model for each person was trained with 9 images, and the remaining image was used in the testing phase. Image scanning is performed in a two dimensional manner, with block size set to 8×8. Only the first 9 DCT coefficients per block were used. Different block overlap values were used to investigate the system performance and the validity of the design assumptions. The recognition rates are around 70% for 0 or 1 pixel overlap, decreasing dramatically down to only 10% for a 6-pixel overlap. This is explained because the assumptions of statistical independence, which are the underlying basis of this model, lose their validity as the overlap increases.

5.8 Refinements of the low-complexity approach

In a subsequent publication by the same authors, [Othman & Aboulnasr, 2001], a hybrid HMM for face recognition is introduced. The proposed system comprises of a LC2D HMM, as described in their earlier work used in combination with a 1D HMM. The LC2D HMM carries out a complete search in the compressed JPEG domain, and a 1D HMM is then applied that searches only in the candidate list provided by the first module.

In the experiments presented in this paper, a 6×2 states model was used for the LC2D HMM, and 4 and 5 state top-to-bottom models were used for the 1D HMM. For the 1D HMM, DCT feature extraction is performed on a horizontal 10 × 92 scanning window. For the 2D HMM, a 8×8 block size is used for scanning the image, and the first 9 DCT coefficients are retained from each block. No overlap is allowed for the sliding windows. Tests are performed on the ORL database. In a first series of tests the effects of training data size on the model robustness were studied. The accuracy of the system ranges from 48%-58% when trained with only 2 images per person, to almost 95%-100% if trained with 9 images per person. A second series of experiments provides a detailed analysis of the trade-off between recognition accuracy and computational complexity and determines an optimal operating point for this hybrid approach. This appears to be the first research in this field to

consider such trade-offs in a detailed study and this methodology should provide a useful approach for other researchers in the future.
In a third paper, [Othman & Aboulnasr, 2003], these authors propose a 2D HMM face recognition system that limits the independence assumptions described in their original work to conditional independence among adjacent observation blocks. In this new model, the active states of the two anti-diagonal observation blocks are statistically independent given the current state and knowledge of the past observations. This translates into a more flexible model, allowing state transitions in the transverse direction as shown in figure 14, taken from, [Othman & Aboulnasr, 2003].

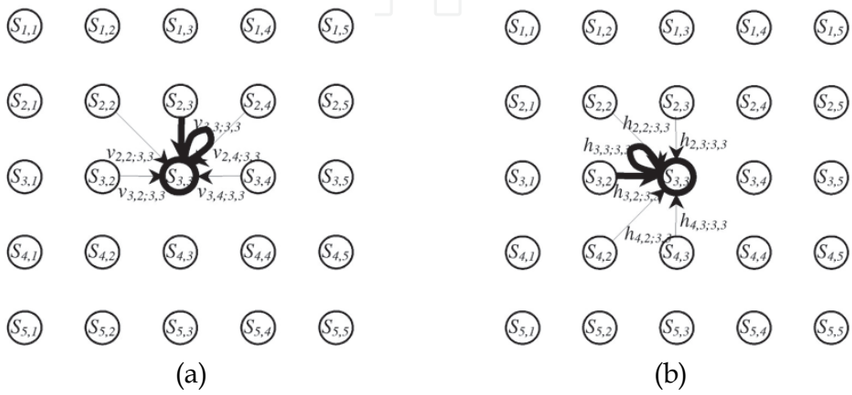


Fig. 14. Modified LC2D HMM [Othman & Aboulnasr, 2003]. (a) Vertical transitions to state S3,3 for 5×5 state model (b) Horizontal transitions to state S3,3 also for 5×5 state model.

This modified LC2D HMM face recognition system is examined for different values of the structural parameters, namely number of states per model and number of Gaussian mixtures per state. These tests are again conducted on the ORL database. The images are scanned using 8×8 blocks and the first 9 2D DCT coefficients comprise the observation vector. The HMMs were trained using 9 images per person, and tested using the 10th image. The test is repeated 5 times with different test images and the results are averaged over a total of 200 test images for 40 persons. Test images are not members of the training data set at any time. The results vary from a very low 4% recognition rate for a 7 × 3 HMM with 64 Gaussian mixtures per state, up to 100% for a 7 × 3 HMM with 4 Gaussian mixtures per state. Best results are obtained for 4 and 8 Gaussians per state. The reason for the poor performance for a higher number of Gaussian mixtures is that the model becomes too discriminating and cannot recognize data with any flexibility, outside the original training set. Finally, the reader's attention is drawn to detailed comments by [Yu & Wu, 2007] on the key assumption of conditional independence in the relationship between adjacent blocks. In this communication, [Yu & Wu, 2007] it is shown that this key assumption is entirely unnecessary.

6. More recent research on HMM in face recognition

While there have been more recent research which applies HMM techniques to face recognition, most of this work has not refined the underlying methods, but has instead combined known HMM techniques with other face analysis techniques. Some work is worth

mentioning, such as that of [Le & Li, 2004] who combined a one-dimensional discrete hidden Markov model (1D-DHMM) with new way of extracting observations and using observation sequences. All subjects in the system share only one HMM that is used as a means to weigh a pair of observations. The Haar wavelet transform is applied to face images to reduce the dimensionality of the observation vectors. Experiments on the AR face database[8] and the CMU PIE face database[9] show that the proposed method outperforms PCA, LDA, LFA based approaches tested on the same databases.

Also worth mentioning is the work of [Yujian, 2006]. In this paper, several new analytic formulae for solving the three basic problems of 2-D HMM are provided. Although the complexity of computing these is exponential in the size of data, it is almost the same as that of a 1D HMM for cases where the numbers of rows or columns are a small constant. While this author did not apply these results specifically to facial recognition problem they appear to offer some promise in simplifying the application of a full 2D HMM to the face recognition problem.

Another notable contribution is the work of [Chien & Liao, 2008] which explores a new discriminative training criterion to assure model compactness combined with ability for accurate to discrimination between subjects. Hypothesis testing is employed to maximize the confidence level during model training leading to a *maximum-confidence* model (MC-HMM) for face recognition. From experiments on the FERET[10] database and GTFD[11], the proposed method obtains robust segmentation in the presence of different facial expressions, orientations, and so forth. In comparison with the maximum likelihood and minimum classification error HMMs, the proposed MC-HMM achieves higher recognition accuracies with lower feature dimensions. Notably this work uses more challenging databases than the ORL database.

Finally we conclude this chapter referring to our own recent work in face recognition using EHMM, presented in [Iancu, 2010; Corcoran & Iancu 2011]. This work can be divided in three parts according to our objectives. The tests were performed on a combined database (BioID, Achermann, UMIST) and on the FERET database. The first objective was to build a recognition system applicable on handheld devices with very low computational power. For this we tested the EHMM-based face recognizer for different sizes of the model, different number of Gaussians, picture size, features, and number of pictures per person used for training. The results obtained for very small picture size (32 × 32), with 1 Gaussian per state and on a simplified EHMM are only 58% recognition for only 1 image per person used for training, when we use 5 pictures per person for training the recognition rates go up to 82% [Corcoran & Iancu 2011]. A second objective was to limit the effect of illumination variations on recognition rates. For this three illumination normalization techniques were used and various combinations of these were tested: histogram equalization (HE), contrast limited adaptive histogram equalization (CLAHE) and DCT in logarithm domain (logDCT). The best recognition rates were obtained for a combination of CLAHE and HE (95.71%) and the worst for logDCT (77.86%) on the combined database [Corcoran & Iancu 2011].

[8] http://www2.ece.ohio-state.edu/~aleix/ARdatabase.html
[9] http://www.ri.cmu.edu/research_project_detail.html?project_id=418&menu_id=261
[10] http://www.frvt.org/feret/default.htm
[11] http://www.anefian.com/research/face_reco.htm

A third objective was to build a system robust to head pose variations. For this we tested the face recognition system using frontal, semi-profile and profile views of the subjects. The first set of tests was performed on the combined database. Here the maximum head pose angle is around 30°. We compared recognition rates obtained when building one EHMM model per person versus one EHMM model per picture. The second set of tests was performed on FERET database which has a much bigger variety of head poses. In this case we used one frontal, 2 semi-profiles and 2 profiles for each subject in the training stage and all pictures of each subject in the testing stage. We compared the recognition rates when building 1 model per person versus 2 models per person versus 3 models per person. We obtained better recognition rates for one model per person for the first set of tests where the database has little head pose variation but better recognition rates for 2 models per person for the second set of tests where the database has a very high head pose variation [Iancu, 2010].

7. Review and concluding remarks

The focus of this chapter is on the use of HMM techniques for face recognition. For this review we have presented a concise yet comprehensive description and review of the most interesting and widely used techniques to apply HMM models in face recognition applications. Although additional papers treating specific aspects of this field can be found in the literature, these are invariably based on one or another of the key techniques presented and reviewed here.

Our goal has been to quickly enable the interested reader to review and understand the state-of-art for HMM models applied to face recognition problems. It is clear that different techniques balance certain trade-offs between computational complexity, speed and accuracy of recognition and overall practicality and ease-of-use. Our hope is that this article will make it easier for new researchers to understand and adopt HMM for face analysis and recognition applications and continue to improve and refine the underlying techniques.

8. References

Baum, L. E. & Petrie, T. (1966). Statistical inference for probabilistic functions of finite state Markov chains. *Annals of Mathematical Statistics*, vol. 37, 1966.

Baum, L. E.; Petrie, T.; Soules, G. & Weiss, N. (1970). A maximization technique occurring in the statistical analysis of probabilistic functions of Markov chains. *Annals of Mathematical Statistics*, vol. 41, 1970.

Baum, L. E. (1972). An inequality and associated maximization technique in statistical estimation for probabilistic functions of Markov processes, Inequalities, vol. 3, pp. 1-8, 1972.

Bicego, M.; Castellani, U. & Murino, V. (2003b). Using hidden markov models and wavelets for face recognition. *Proceedings of Image Analysis and Processing 2003. 12th International Conference on*, pp. 52–56, 2003.

Bicego, M.; Murino, V. & Figueiredo, M. (2003a). A sequential pruning strategy for the selection of the number of states in hidden markov models. *Pattern Recognition Letters*, Vol. 24, pp. 1395–1407, 2003.

Chien, J-T. & Liao, C-P. (2008). Maximum Confidence Hidden Markov Modeling for Face Recognition. *Pattern Analysis and Machine Intelligence, IEEE Transactions on* , Vol. 30, No 4, pp. 606-616, April 2008

Corcoran, P.M. & Iancu, C. (2011). Automatic Face Recognition System for Hidden Markov Model Techniques, *Face Recognition Volume 2, Intech Publishing*, 2011.

Eickeler, S. (2002). Face database retrieval using pseudo 2d hidden markov models. *Fifth IEEE International Conference on Automatic Face and Gesture Recognition, Proceedings*, pp. 58–63, May 2002.

Eickeler, S.; Muller, S. & Rigoll, G. (2000). Recognition of jpeg compressed face images based on statistical methods. *Image and Vision Computing Jour- nal, Special Issue on Facial Image Analysis*, Vol. 18, No 4, pp. 279–287, March 2000.

Figueiredo, M.; Leitao, J. & Jain, A. (1999). On fitting mixture models. *Proceedings of the Second International Workshop on Energy Minimization Methods in Computer Vision and Pattern Recognition, Springer-Verlag*, pp. 54-69, 1999.

Iancu, C. (2010). Face recognition using statistical methods. *PhD thesis*, NUI Galway, 2010.

Jelinek, F.; Bahl, L.R. & Mercer, R.L. (1975). Design of a linguistic statistical decoder for the recognition of continuous speech. *IEEE Transactions on Information Theory*, Vol. 21, No 3, pp. 250 – 256, 1975.

Juang, B.H. (1984). On the hidden markov model and dynamic time warping for speech recognition-a unified view. *AT&T Technical Journal*, Vol. 63, No 7, pp. 1213–1243, September 1984.

Juang, B.H. & Rabiner, L.R. (2005). Automatic speech recognition - a brief history of the technology development. *Elsevier Encyclopedia of Language and Linguistics*, Second Edition, 2005.

Kohir, V.V. & Desai, U.B. (1998). Face recognition using a dct-hmm approach. *Applications of Computer Vision, WACV '98, Proceedings, Fourth IEEE Workshop on*, pp. 226–231, October 1998.

Kohir, V.V. & Desai, U.B. (1999). A transform domain face recognition approach. *TENCON 99, Proceedings of the IEEE Region 10 Conference*, Vol. 1, pp. 104–107, September 1999.

Kohir, V.V. & Desai, U.B. (2000). Face recognition. *IEEE International Symposium on Circuits and Systems*, Geneva, Switzerland, May 2000.

Kuo, S. & Agazzi, O. (1994). Keyword spotting in poorly printed documents using pseudo 2-d hidden markov models. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 16, pp. 842–848, August 1994.

Le, H.S. & Li, H. (2004). Face identification system using single hidden markov model and single sample image per person. *IEEE International Joint Conference on Neural Networks*, Vol. 1, 2004.

Levin, E. & Pieraccini, R. (1992). Dynamic planar warping for optical character recognition. *Proceedings ICASSP 1992,* San Francisco, Vol. 3, pp. 149–152, March 1992.

Levinson, S.E.; Rabiner, L.R. & Sondhi, M.M. (1983). An introduction to the application of the theory of probabilistic functions of a markov process to automatic speech recognition. *Bell System Technical Journal*, Vol. 62, No 4, pp. 1035–1074, April 1983.

Martinez, A. (1999). Face image retrieval using hmms. *IEEE Workshop on Content-Based Access of Image and Video Libraries, (CBAIVL '99) Proceedings*, pp. 35–39, June 1999.

Nefian, A.V. (1999). A hidden markov model based approach for face detection and recognition. *PhD Thesis*, 1999.

Nefian, A.V. & Hayes III, M.H. (Oct. 1998). Face detection and recognition using hidden markov models. *Image Processing, ICIP 98, Proceedings. 1998 International Conference on*, Vol. 1, pp. 141–145, October 1998.

Nefian, A.V. & Hayes III, M.H. (May 1998). Hidden markov models for face recognition. *Acoustics, Speech, and Signal Processing ICASSP '98. Proceedings of the 1998 IEEE International Conference on*, Vol. 5, pp. 2721–2724, May 1998.

Nefian, A.V. & Hayes III, M.H. (1999). An embedded hmm-based approach for face detection and recognition. *Acoustics, Speech, and Signal Processing, ICASSP '99. Proceedings,* IEEE International Conference, 6:3553–3556, March 1999.

Othman, H. & Aboulnasr, T. (2000). Hybrid hidden markov model for face recognition. *4th IEEE Southwest Symposium on Image Analysis and In- terpretation*, pp. 34–40, April 2000.

Othman, H. & Aboulnasr, T. (2001). A simplified second-order hmm with ap- plication to face recognition. *ISCAS 2001 IEEE International Symposium on Circuits and Systems*, Vol. 2, pp. 161–164, May 2001.

Othman, H. & Aboulnasr, T. (2003). A separable low complexity 2d hmm with application to face recognition. *Pattern Analysis and Machine Intelli- gence, IEEE Transactions on*, 2003.

Park, H.S. & Lee, S.W. (1998). A Truly 2D Hidden Markov Model For Off-Line Handwritten Character Recognition. *Pattern Recognition*, Vol. 31, No 12, pp. 1849-1864, December 1998.

Rabiner, L.R. (1989). A tutorial on hidden markov models and selected applications in speech recognition. *Proceedings of IEEE*, Vol. 77, No 2, pp. 257–286, February 1989.

Rabiner, L.R. & Juang, B.H. (1986). An introduction to hidden markov models. *IEEE ASSP Magazine*, Vol. 3, No 1, pp. 4–16, 1986.

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

Samaria, F. & Fallside, F. (1993). Face identification and feature extraction using hidden markov models. *Image Processing: Theory and Applications, Elsevier*, pp. 295–298, 1993.

Samaria, F. & Harter, A.C. (1994). Parameterization of a stochastic model for human face identification. *Applications of Computer Vision, 1994., Pro- ceedings of the Second IEEE Workshop on*, Vol. 77, pp. 138–142, December 1994.

Viola, P. & Jones, M. (2001). Robust real-time object detection, *Technical report 2001/01*, Compaq CRL, 2001.

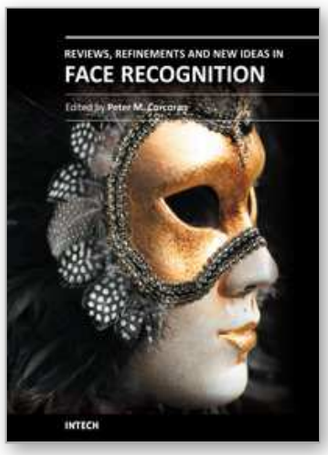
Wallhoff, F.; Eickeler, S. & Rigoll, G. (2001). A comparison of discrete and continuous output modeling techniques for a pseudo-2d hidden markov model face recognition system. *International Conference on Image Processing, Proceedings*, Vol. 2, pp. 685–688, October 2001.

Wallhoff, F., Müller, S. & Rigoll, G. (2001). Hybrid face recognition system for profile views using the mugshot database. *IEEE ICCV Workshop on Recognition, Analysis and*

Tracking of Faces and Gestures in Real-Time Systems, Proceedings, pp. 149–156, July 2001.

Yu, L. & Wu, L. (2007). Comments on 'a separable low complexity 2d hmm with application to face recognition'. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, Vol. 29, No 2, pp. 368–368, February 2007.

Yujian, L. (2007). An analytic solution for estimating two-dimensional hidden Markov models. *Applied Mathematics and Computation*, Vol. 185, No 2, pp. 810-822, February 2007.



Reviews, Refinements and New Ideas in Face Recognition

Edited by Dr. Peter Corcoran

ISBN 978-953-307-368-2
Hard cover, 328 pages
Publisher InTech
Published online 27, July, 2011
Published in print edition July, 2011

As a baby one of our earliest stimuli is that of human faces. We rapidly learn to identify, characterize and eventually distinguish those who are near and dear to us. We accept face recognition later as an everyday ability. We realize the complexity of the underlying problem only when we attempt to duplicate this skill in a computer vision system. This book is arranged around a number of clustered themes covering different aspects of face recognition. The first section on Statistical Face Models and Classifiers presents reviews and refinements of some well-known statistical models. The next section presents two articles exploring the use of Infrared imaging techniques and is followed by few articles devoted to refinements of classical methods. New approaches to improve the robustness of face analysis techniques are followed by two articles dealing with real-time challenges in video sequences. A final article explores human perceptual issues of face recognition.

InTech Europe
University Campus STeP Ri
Slavka Krautzeka 83/A
51000 Rijeka, Croatia
Phone: +385 (51) 770 447
Fax: +385 (51) 686 166
www.intechopen.com

InTech China
Unit 405, Office Block, Hotel Equatorial Shanghai
No.65, Yan An Road (West), Shanghai, 200040, China
中国上海市延安西路65号上海国际贵都大饭店办公楼405单元
Phone: +86-21-62489820
Fax: +86-21-62489821

IntechOpen