Soft Comput (2008) 12:615–621
DOI 10.1007/s00500-007-0253-0

FOCUS

A face recognition system based on Pseudo 2D HMM applied to neural network coefficients

Vitoantonio Bevilacqua · Lucia Cariello · Gaetano Carro · Domenico Daleno · Giuseppe Mastronardi

Published online: 5 October 2007

Abstract Face recognition from an image or video sequences is emerging as an active research area with numerous commercial and law enforcement applications. In this paper different Pseudo 2-dimension Hidden Markov Models (HMMs) are introduced for a face recognition showing performances reasonably fast for binary images. The proposed P2-D HMMs are made up of five levels of states, one for each significant facial region in which the input frontal images are sequenced: forehead, eyes, nose, mouth and chin. Each of P2-D HMMs has been trained by coefficients of an artificial neural network used to compress a bitmap image in order to represent it with a number of coefficients that is smaller than the total number of pixels. All the P2-D HMMs, applied to the input set consisting of the Olivetti Research Laboratory face database combined to others photos, have achieved good rates of recognition and, in particular, the structure 3-6-6-6-3 has achieved a rate of recognition equal to 100%.

Keywords Face recognition · Hidden Markov models · Artificial neural network · Pseudo two-dimension HMM

1 Introduction

Face recognition is a research area that in the recent years has grown in importance due to its numerous potential applicability as a biometric system in commercial and security application. A successful system could be used to prevent unauthorised access to smart cards, credit card and, above all, it could be a powerful tool for criminal identification. In fact, instead of the others biometric methods like fingerprint or iris recognition, it represents a not invasive control mechanism.

These systems require a robust algorithm specially for human face recognition under different lighting conditions, facial expressions and orientations.

The recurring methods existing in literature for face recognition are eigenfaces, neural networks and several Hidden Markov Models (HMMs) based systems.

In particular HMMs are a set of statistical models used to characterise the statistical properties of a signal (Samaria and Young 1994; Samaria 1992, 1994). HMMs are characterised by two interrelated processes:

1. an unobservable Markov chain with a finite number of states, a state transition probability matrix and an initial state probability distribution. This is the principal aspect of a HMM;
2. a set of probability density functions for each state.

The elements that characterised a HMM are:

– $N = |S|$ is the number of states of the model. If S is the set of states, then $S = \{s_1, s_1, \ldots, s_1\}$. $s_i \in S$ is one of the states that can be employed by the model. To observe the system are used T observation sequences, where T is the number of observations. The state of the model at time t is given by $q_t \in S$, $1 < t < T$;
– $M = |V|$ is the number of different observation symbols. If V is the set of all possible observation symbols (also called the codebook of the model), then $V = \{v_1, v_2, \ldots, v_M\}$;

V. Bevilacqua (✉) · L. Cariello · G. Carro · D. Daleno · G. Mastronardi
Department of Electrical and Electronics, Polytechnic of Bari, Via Orabona, 4, 70126 Bari, Italy
e-mail: bevilacqua@poliba.it

– $A = \{a_{ij}\}$ is the state transition probability matrix, where a_{ij} is the probability that the state i became the state j:

$$a_{ij} = p\,(q_t = s_j \,|\, q_{t-1} = s_i), \tag{1}$$

where $1 \le i, j \le N$;

– $B = \{b_j(k)\}$ the observation symbol probability matrix, $b_j(k)$ is the probability to have the observation k when the state is j:

$$b_j(k) = p\,(o_t = v_k \,|\, q_t = s_j), \tag{2}$$

where $1 \le j \le N,\ 1 \le k \le M$;

– $\Pi = \{\pi_1, \pi_2, \ldots, \pi_N\}$ is the initial state distribution:

$$\pi_i = p\,(q_1 = s_i), \tag{3}$$

where $1 \le i \le N$.

Using a shorthand notation, a HMM is defined by the following expression:

$$\lambda = (A, B, \Pi). \tag{4}$$

2 HMM for face recognition

Hidden Markov Models have been successfully used for speech recognition where data are essentially one dimensional because the HMM provides a way of modelling the statistical properties of a one-dimensional signal. To apply the HMM also to process images, that are two-dimensional data, we consider temporal or space sequences: this question has been considered in Samaria and Young (1994); Samaria (1992, 1994), where Samaria suggests to use a space sequence to model an image for HMM. For frontal face images, the significant facial regions are 5: forehead, eyes, nose, mouth and chin (Nefian and Monson 1998a, b).

Each of these facial regions (facial band) is assigned to a state in a left to right 1D continuous HMM. The Left-to-right HMM used for face recognition is shown in Fig. 1. To

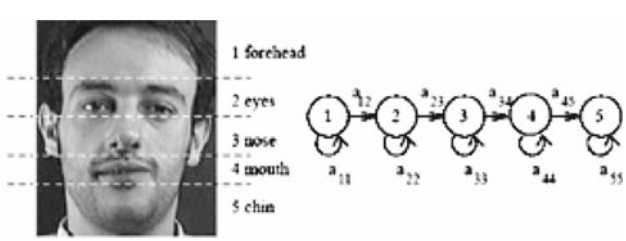


Fig. 1 The significant facial regions

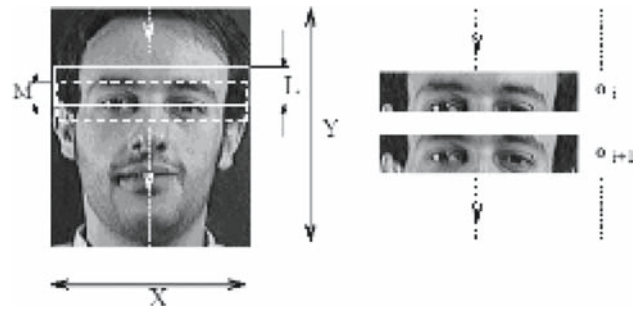


Fig. 2 The facial regions overlapped

recognise the face k we must trained the following HMM:

$$\lambda^{(k)} = (A^{(k)}, B^{(k)}, \Pi^{(k)}). \tag{5}$$

To train a HMM we have used four different frontal face grey scale images for each person. Each face image of width X and height Y is divided into overlapping blocks of height L and width W (see Fig. 2). The amount of overlap between consecutive blocks is M. The number of blocks extracted from each face image equals the number of observation vectors T and is given by:

$$T = \frac{(Y - L)}{L - M} + 1. \tag{6}$$

The choice of parameters M and L can significantly affect the system recognition rate. A high amount of overlap M significantly increases the recognition rate because it allows the features to be captured in a manner that is independent of the vertical position. The choice of parameter L is more delicate. An insufficient amount of information about the observation vector could arise from a small value of the parameter L, while large L increases the probability of cutting across the features. However, the system recognition rate is more sensitive to the M variations than L ones, for this reason is used $M = (L - 1)$.

3 Pseudo-2D HMM

In this paper we will focus on Pseudo two-dimension HMMs (P2D-HMM) which are extensions of the 1D-HMM and are obtained by linking one-dimensional left to right HMMs to form vertical super-states. These super-states model the sequence of rows in the image and the linear 1D-HMMs, which are inside the super-states, model each row (Eickeler et al. 1998; Nefian 1999). The states sequence in the rows is independent from the states sequences of neighbouring rows.

The elements of an P2D-HMM are:

- the number of super-states, N_0, and the set of super-states, $S_0 = \{S_{0,i}\}$, $1 \le i \le N_0$;
- the initial super-state distribution, $\Pi_0 = \{\pi_{0,i}\}$, were $\pi_{0,i}$ are the probabilities of being in super-state i at time zero;
- the super-state transition probability matrix,

$$A_0 = \{a_{0,ij}\}, \tag{7}$$

  where $a_{0,ij}$ is the probability of transitioning from super-state i to super-state j.
- the parameters of the embedded HMMs, which include:
  1. the number of embedded states in the kth super-state, $N_1^{(k)}$, and the set of embedded super-states, $S_1^{(k)} = \{S_{1,i}^{(k)}\}$;
  2. the initial state distribution, $\Pi_1^{(k)} = \{\pi_{1,i}^{(k)}\}$ where $\pi_{1,i}^{(k)}$ are the probabilities of being in state i of super-state k at time zero;
  3. the state transition probability matrix,

$$A_1^{(k)} = \{a_{1,jk}^{(k)}\} \tag{8}$$

  that specifies the probability of transitioning from state k to state j.
- finally, there is the state probability matrix,

$$B^{(k)} = b_i(k)(O_{t_0,t_1}) \tag{9}$$

  for the set of observations where O_{t_0,t_1} represent the observation vector at row t_0 and column t_1. Let

$$\Lambda^{(k)} = \left(A_1^{(k)}, B_1^{(k)}, \Pi_1^{(k)}\right) \tag{10}$$

  be the set of parameters that define the kth super-state. Using a shorthand notation, a P2D-HMM is defined as the triplet:

$$\lambda = (A_0, \Lambda, \Pi_0), \tag{11}$$

  where

$$\Lambda = \Lambda^{(1)}, \Lambda^{(2)}, \ldots, \Lambda^{(N_0)}. \tag{12}$$

Although more complex than a one-dimensional HMM, a P2D-HMM is more appropriate for data that are two-dimensional, and has a complexity proportional to the sum of the squares of the number of states:

$$\sum_{k=1}^{N_0} (N_1^{(k)})^2. \tag{13}$$

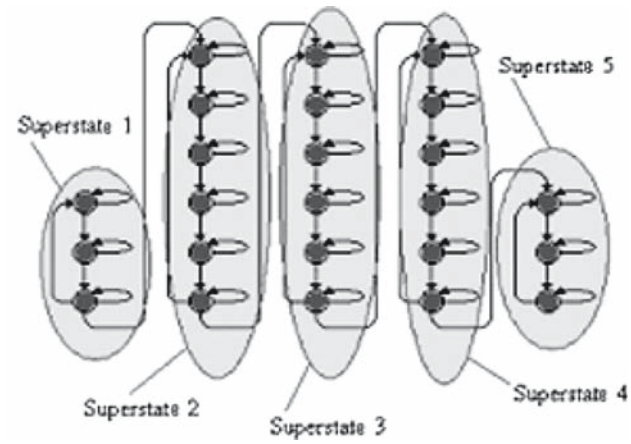


Fig. 3 Pseudo 2D HMM 3-6-6-6-3

In Fig. 3, it is shown an example of a P2D-HMM with a structure 3-6-6-6-3: the super-states 1 and 5 are constituted by a left to right 1D-HMM with 3 states, instead the super-states 2, 3 and 4 are constituted by a left to right 1D-HMM with 6 states.

The first approach to face recognition based on P2D-HMM is described in Samaria and Young (1994) followed from the approaches in Eickeler et al. (2000) and Kohir and Desai (1998) both of them applied to the Olivetti Research Laboratory (ORL) face database. In our paper a structure similar to the Pseudo 2D HMM implemented in the mentioned approaches is used and it has been combined with an artificial neural network.

4 The face recognition system

In this paper we present an hybrid system for face recognition (see Fig. 4), made up of a cascade connection of two different systems: an artificial neural network, existing in Bevilacqua et al. (2006), and different P2D-HMMs.

The system's input is an image of a person that must be recognised and the output is its identification with the corresponding rate of recognition. In order to train the system, we used nine photos for each person and the remaining one to validate its performance. The used database, made of 510 images (ten images for each of the 51 persons to identify), consists of a combination of the ORL database (Cottrell et al. 1987), in which there are frontal images, and other profiles photos of persons camouflaged with dark glasses or bandage. These images are ".bmp" files in grey scales of 92×112 pixels.

To realise the hybrid schema, the following steps have been executed:

1. Training and saving of artificial neural network;

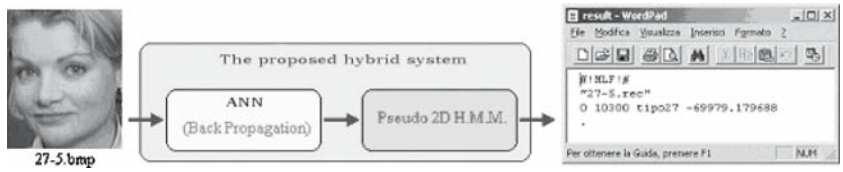


Fig. 4 The proposed hybrid system

Table 1 Characteristic parameters of photos and blocks

X = width photo = 92 pixels	**T** = number of blocks for photos = 103
Y = height photo = 112 pixels	**X**×**Y** = photo dimension =10,304 pixels
L = height block = 10 pixels	**X**×**L** = block dimension = 920 pixels
M = blocks overlapping = 9 pixels	**X**×**M** = overlapping dimension = 828 pixels

2. Transformation of photos in HTK format;
3. Training of different P2D-HMM structures;
4. Identification of the Validation Set subjects.

4.1 Training and saving of the artificial neural network

The considered faces are sequenced in observation windows, following the Samaria model (see Sect. 2). The values of the parameters described in Sect. 3, adjusted to the images manipulated from our system, are represented in Table 1.

The ANN, using the error back propagation (EBP) algorithm, extracts the main features from the image to store them in a sequence of 50 bits, reducing the complexity of the problem and compressing the images in order to represent them with a number of coefficients smaller than pixels.

The image is a facial feature of a frontal face image; from this area we consider 103 segments of 920 pixels that represent the observable states of the model (Bevilacqua et al. 2006). Now all of these sections are divided into features of 230 pixels, that are the input of the network. The first layer is formed by 230 neurons, each neuron per pixel, the hidden layer is composed by 50 units and the last layer by 230 neurons. The network is trained using the values of the input layer equal to the values of the output layer, following the model proposed by Cottrell et al. (1987), so that, when the training is terminated, the 50 bits that are the outputs of the unique hidden layer are the compressed image. For any window of 230 pixels we have an array of 50 elements, so a section of 920 pixels is compressed in a four sub windows of 50 binary values array each.

The weights matrix, referred to the connections between the inputs and the hidden layer, codifies the image bitmap, while the weights matrix associated to the connections between the hidden layer and the outputs, decodes the sequence of bits. So for each of the 103 blocks of 920 pixels (4 × 230) we obtain 103 observation vectors with 200 coefficients (4 × 50) and a rate of compression equal to $\frac{103\times920}{103\times200} = 4.6$.

By observing the schema of Fig. 5 it is possible to note that the "Training Set" used for ANN is composed by 300 photos: 10 face images for each of the first 30 samples of the entire set made of 51 persons. The training function is iterated 200 times for each photo and, at the end of the training phase, the neuron weights are saved in a ".bin" file. Finally the ANN is tested with other images that are similar, but not the same to the training set features.

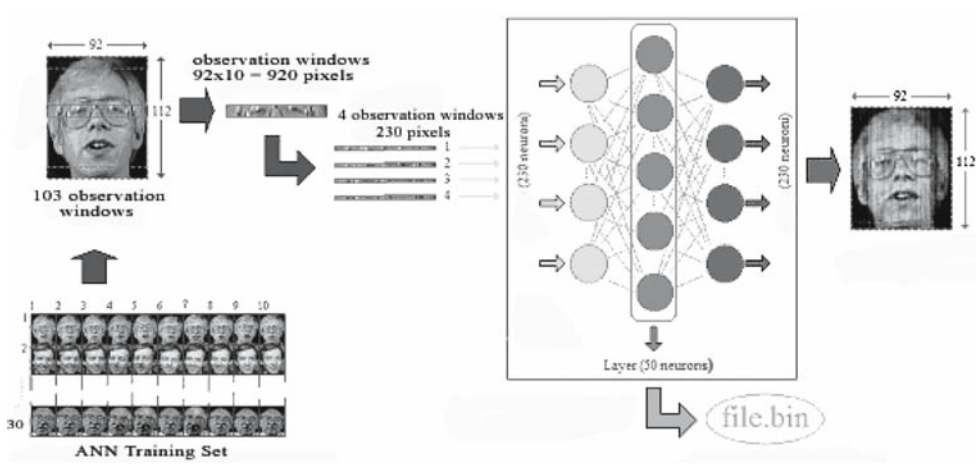


Fig. 5 Schema of the ANN training phase

4.2 Transformation of photos in HTK format

Now we have a face image compressed into an observation vector of 103 elements of 200 binary (1/0) values that will be computed by the P2D Hidden Markov Models.

To train P2D-HMMs the Hidden Markov Model Toolkit v3.3(HTK) (Young et al. 2002) designed by Steve Young from the Cambridge University Engineering Department has been used. So we transformed the ANN output .bin file into another .bin file in HTK format: this step consisted in the creation of a heading, according to the HTK syntax, and in the writing of the 20,600 coefficients (103×200), according the "Little Endian" data storage.

4.3 Training of different P2D-HMM structures

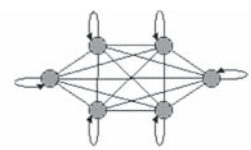
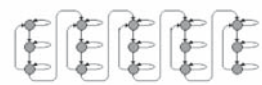
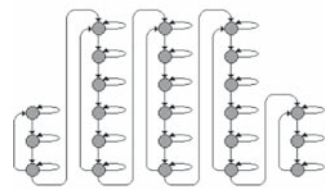
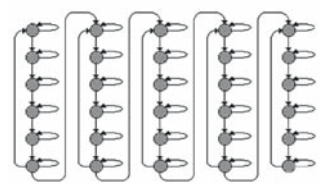
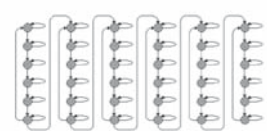
The different Pseudo 2D Hidden Markov Model structures associated with each of the 51 subjects, are represented in Table 2.

The P2D-HMMs were trained by the .bin file created for each photo. The toolkit HTK instruction, that implements this operation, is "HInit".

At the end of the training we obtain the P2D-HMMs for each subject:

- hmm1→P2D-HMM for the subject 1 trained with the first 9 photos in HTK format;
- hmm2→P2D-HMM for the subject 2 trained with the first 9 photos in HTK format;
  …
- hmm51→P2D-HMM for the subject 51 trained with the first 9 photos in HTK format.

Table 2 Different Pseudo 2D Hidden Markov Model structures

HMM 5Ergodic	
Pseudo2D HMM 3-3-3-3-3	
Pseudo2D HMM 3-6-6-6-3	
Pseudo2D HMM 6-6-6-6-6	
Pseudo2D HMM 6-6-6-6-6-6	

4.4 Identification of the validation set subjects

After the P2D-HMMs training process, it is possible to proceed with the recognition phase, according the schema in Fig. 6. The toolkit HTK instruction that implements this operation is "HVite".

In this way, the Viterbi algorithm is applied to each of the 51 P2D-HMMs using the same HTK file (26-5.bmp in Fig. 7) and each of P2D-HMMs returns a logarithmic probability value. The highest probability value identifies the P2D-HMM and so the corresponding recognised sample (see Fig. 7).

At the end of the process, the final result of the identification, that is the recognised person and the consequent logarithmic probability value, are contained in the "result" file.

5 Conclusions and results

We conducted five experiments for each of the different introduced Pseudo 2D HMMs structures. The results that we have obtained are represented in Table 3.

It can be noticed that all the utilised approaches provide satisfactory results and in particular the Pseudo 2D 3-6-6-6-3 is the structure that achieves an identification with a rate of recognition equal to 100%.

Moreover, applying our recognition system relative to the only ORL databases images, the obtained results outperformed, all other methods proposed in the literature. This could be observed in Table 4, which presents a comparison between published results obtained by the most important face recognition algorithms on the ORL database. If we compare our results with the others present in Table 4, we

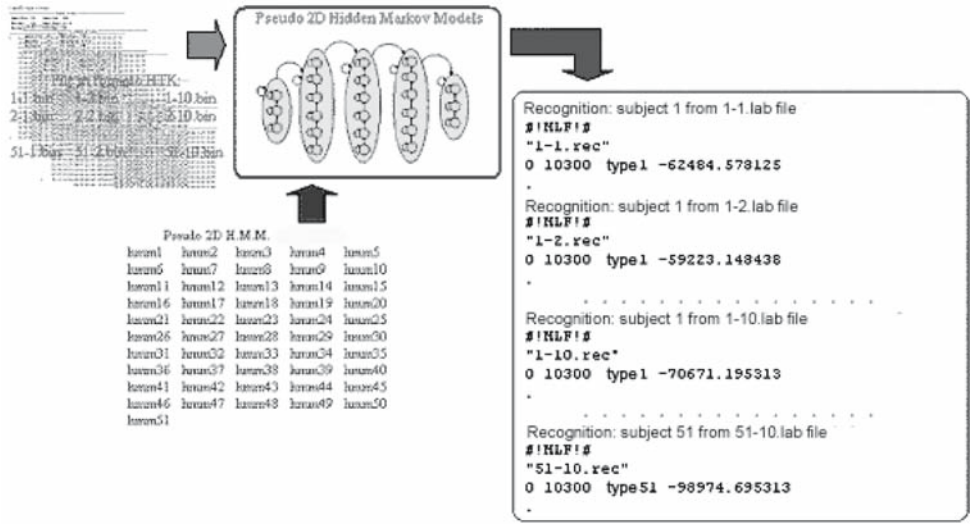


Fig. 6 Schema of recognition phase

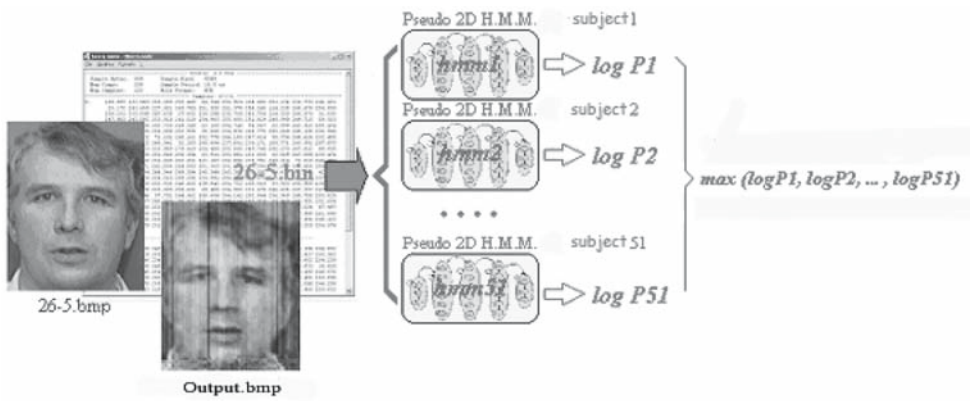


Fig. 7 Example of identification by logarithmic probability

Table 3 Rates of recognition obtained from the different implemented P2D-HMMs

Hidden Markov Models	The exact identification $100 - \frac{n\text{errors}}{5.1}$ (%)
Pseudo 2D **3-3-3-3-3**	99.80 (1 error on 510 photo)
Pseudo 2D **3-6-6-6-3**	**100**
Pseudo 2D **6-6-6-6-6**	99.80 (1 error on 510 photo)
Pseudo 2D **6-6-6-6-6-6**	99.80 (1 error on 510 photo)
5Ergodic	98.82 (6 error on 510 photo)

Table 4 Comparative results on ORL database

Methods	Recognition rate (%)	Reference	Year
Local Gabor wavelet	66.0	Huang et al. (2004)	2004
Eigenface	72.1	Yin et al. (2005)	2005
Fisherface	76.3	Yin et al. (2005)	2005
PCA	82.8	Huang et al. (2004)	2004
ICA	85.0	Huang et al. (2004)	2004
Spectroface	86.4	Huang et al. (2004)	2004
Pseudo 2D HMM feature: DCT coefficients	99.5	Eickeler et al. (2000)	1998
Ergodic HMM + DCT	99.5	Kohir and Desai (1998)	1998
Pseudo 2D HMM + neural network coefficients	100	This paper	2006

can conclude that a combined employment of the ANN and P2D-HMMs improves a more efficient and surer personal identification process.

References

Bevilacqua V, Mastronardi G, Pedone A, Romanazzi G, Daleno D (2006) A Hidden Markov Model for recognition using artificial neural networks, LNCS 4113. Springer, Heidelberg, pp 126–134

Cottrell GW, Munro P, Zipser D (1987) Learning internal representations from grey scale image: an example of extensional programming. In: Ninth annual conference of the cognitive science society, pp 462–473

Eickeler S, Mller S, Rigoll G (1998) High performance face recognition using pseudo 2-D Hidden Markov Models. In: European control conference

Eickeler S, Mller S, Rigoll G (2000) A recognition of jpeg compressed face images based on statistical methods. Image Vision Comput 18:279–287

Huang J, Yuen PC, Lai JH, Li C (2004) Face recognition using local and global features. EURASIP J Appl Signal Process 4:530–541

Kohir V, Desai U (1998) Face recognition using DCTHMM approach, workshop on advances in facial. In: Workshop on advances in facial image analysis and recognition technology (AFIART), Freiburg, Germany

Nefian AV, Monson H (1998a) Hidden Markov Models for face recognition. In: IEEE international conference on acoustics, speech and signal processing (ICASSP98), Seattle, USA, pp 2721–2724

Nefian AV, Monson H (1998b) Detection and recognition using Hidden Markov Models. In: International conference on image processing, Chicago, IL, USA, pp 141–145

Nefian A (1999) A Hidden Markov Model-based approach for face detection and recognition. Ph.D. thesis, Georgia Institute of Technology, Atlanta

Rabiner L, Huang B (1993) Fundamentals of speech recognition. Prentice-Hall, Englewood Cliffs

Samaria F, Young S (1994) Face HMM based architecture for face identification. Image Comput Vision 12:537–583

Samaria F (1992) Face identification using Hidden Markov Model. In: 1st year report Cambridge University Engineering Department, London, UK

Samaria F (1994) Face recognition using Hidden Markov Models. Master's thesis, University of Cambridge, UK

Yin B, Bai X, Shi Q, Sun Y (2005) Enhanced fisherface for face recognition. J Inform Comput Sci 2:591–595

Young S, Evermann G, Hain T, Kershaw D, Moore G, Odell J, Ollason D, Povey D, Valtchev V, Woodland P (2002) The HTK Book, Cambridge University Engineering Department, Cambridge, UK