IDIAP RESEARCH REPORT

LOCAL FEATURES AND 1D-HMMS FOR FAST AND ROBUST FACE AUTHENTICATION

Fabien Cardinaux [(a)]

IDIAP–RR 05-17

APRIL 2005

SUBMITTED FOR PUBLICATION

(a) cardinau @ idiap.ch

IDIAP Research Institute **www.idiap.ch**
Rue du Simplon 4 P.O. Box 592 1920 Martigny – Switzerland
Tel: +41 27 721 77 11 Fax: +41 27 721 77 12 Email: info@idiap.ch

IDIAP Research Report 05-17

LOCAL FEATURES AND 1D-HMMS FOR FAST AND ROBUST FACE AUTHENTICATION

Fabien Cardinaux

APRIL 2005

SUBMITTED FOR PUBLICATION

Abstract. It has been previously demonstrated that systems based on Hidden Markov Models (HMMs) are suitable for face recognition. The proposed approaches in the literature are either HMMs with one-dimensional (1D-HMMs) or two-dimensional (2D-HMMs) topology. Both have shown some serious drawbacks. The 1D-HMM approaches typically use a whole row (or column) of an image as observation vector and by consequence do not allow horizontal (or vertical) alignment. 2D-HMM approaches present some implementation issues because of the computational cost. In this paper, we propose a 1D-HMM approach which allow the use of local features and we will demonstate the accuracy of this approach on the so-called BANCA database.

1 Introduction

An identity authentication system has to discriminate between two kinds of events: either the person claiming a given identity is the true claimant or the person is an impostor. This is in contrast to an identification system, which attempts to find the identity of a given person out of a pool of people. Both verification and identification systems can be thought of as falling in the general research area of face recognition (FR).

Many techniques have been proposed for FR; we can dissociate holistic approaches where one feature vector describes the entire face and local feature approaches where a face is represented by a set of feature vectors and each only describes a region of the face. Popular methods to handle local features for FR include Hidden Markov Model (HMM) based approaches. Many variations of the HMM have been introduced, including 1D-HMM [13], Pseudo 2D-HMM (P2D-HMM) [5, 9], Low Complexity 2D-HMM (LC 2D-HMM) [11] and Gaussian Mixture Model (GMM) [3] (which can be considered as a simplified version of HMMs).

Approaches based on HMMs have shown promising performances. However, none of them combines robustness and low computational cost. The P2D-HMM obtains the best performances in [11, 2]; however, it is computationally intensive, making it inappropriate for most of applications on current hardware. GMMs are less complex and faster than HMMs with the cost of lower accuracy. For 1D-HMM based approaches, typical observation sequence represents consecutive horizontal strips and conserves rigid horizontal spatial constraints. As a consequence, typical 1D-HMM approaches are sensitive to a imperfectly localized faces (due to a non-perfect face detection) and face deformation (due to different face expressions).

In this paper we investigate an alternative 1D-HMM structure which deals with observation vectors representing a block of the image instead of a whole strip in traditional 1D-HMM approaches. The blocks of a same line are then treated independantly with no spacial constraints, making the model robust to misalignment.

2 HMMs for Face Recognition

2.1 Gaussian Mixture Models

GMMs can be considered as one state HMM. In the GMM approach, all feature vectors are assumed to be independent. Given the GMM parameter set λ, the likelihood of a set of T feature vectors $X = \{\mathbf{x}_t\}_{t=1}^{T}$ is found with

$$P(X|\lambda) = \prod_{t=1}^{T} P(\mathbf{x}_t|\lambda) \tag{1}$$

where

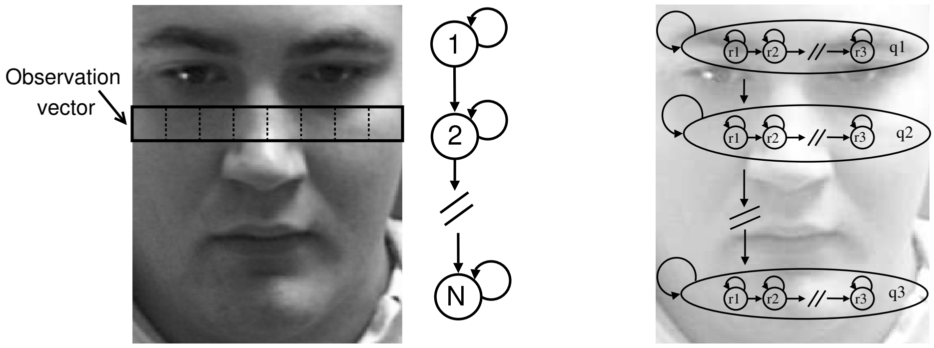
$$P(\mathbf{x}|\lambda) = \sum_{k=1}^{N_G} m_k \, \mathcal{N}(\mathbf{x}|\mu_k, \Sigma_k) \tag{2}$$

$$\lambda = \{m_k, \mu_k, \Sigma_k\}_{k=1}^{N_G} \tag{3}$$

Here, $\mathcal{N}(\mathbf{x}|\mu, \Sigma)$ is a D-dimensional gaussian density function [4] with mean μ and diagonal covariance matrix Σ. N_G is the number of gaussians and m_k is the weight for gaussian k (with constraints $\sum_{k=1}^{N_G} m_k = 1$ and $\forall\, k : m_k \geq 0$).

2.2 1D-HMM

The one-dimensional HMM (1D HMM) is a particular HMM topology where only self transitions or transitions to the next state are allowed (see Fig. 1(a) for an example). This type of HMM is also known as a top-bottom



(a) 1D HMM topology (b) P2D HMM topology

Figure 1: Figure 1(a) is an example of 1D-HMM sampling window and topology and Figure 1(b) represents a P2D HMM: the emission distributions of the vertical HMM are estimated by horizontal HMMs. q_i represent the states of the main HMM and r_j represent the embedded HMMs states

HMM [13] or left-right HMM in the context of speech recognition [12]. The model is characterized by the following:

1. N, the number of states in the model; each state corresponds to a region of the face; $S = \{S_1, S_2, \ldots, S_N\}$ is the set of states. The state of the model at row t is given by $q_t \in S$, $1 \leq t \leq T$, where T is the length of the observation sequence.

2. The state transition matrix $A = \{a_{ij}\}$. The topology of the 1D HMM allows only self transitions or transitions to the next state:

$$a_{ij} = \begin{cases} P(q_t = S_j | q_{t-1} = S_i) & \text{for } j = i, \ j = i+1 \\ 0 & \text{otherwise} \end{cases}$$

3. The emission probability distribution $B = \{b_j(\mathbf{x}_t)\}$, where

$$b_j(\mathbf{x}_t) = p(\mathbf{x}_t | q_t = S_j) \tag{4}$$

In compact notation, the parameter set of the 1D HMM is:

$$\lambda = (A, B) \tag{5}$$

If we let Q to be a state sequence $q_1, q_2, \cdots, q_T$, then the likelihood of an observation sequence X is:

$$\begin{aligned} P(X|\lambda) &= \sum_{\forall Q} P(X, Q|\lambda) && (6) \\ &= \sum_{\forall Q} \prod_{t=1}^{T} b_{q_t}(\mathbf{x}_t) \prod_{t=2}^{T} a_{q_{t-1}, q_t} && (7) \end{aligned}$$

The calculation of this likelihood according to the direct definition in Eqn. (7) involves an exponential number of computations. In practice the Forward-Backward procedure is used [12]; it is mathematically equivalent, but considerably more efficient.

2.3 P2D-HMM

Emission probabilities of 1D-HMMs are typically represented using mixtures of gaussians. In the case of P2D-HMM, the emission probabilities of the HMM (now referred to as the "main HMM") are estimated through a secondary HMM (referred to as an "embedded HMM"). Figure 1(b) is an example of P2D-HMM topology. The states of the embedded HMMs are in turn modeled by a mixture of gaussians. This approach was used for face identification in [5, 13] and the training process is described in detail in [10].

3 Proposed approach

3.1 Proposed 1D-HMM

The observation sequence for traditional 1D-HMM is typically composed of vectors that represent consecutive horizontal strips (see Figure 1 for an example) . As a consequence, 1D-HMM system has rigid horizontal constraints and does not allow the system to deal with face deformations, when the expression change and imperfectly aligned faces, when localization has not been perfectly done.

For the proposed 1D-HMM structure, the observation vectors are extracted from blocks similarly to the GMM and P2D-HMM approaches. Let us denote the sequence of N_S observation vectors representing the consecutive horizontal strips of an image as $X = \{\mathbf{x}_s\}_{s=1}^{N_S}$. Each strip can itself be represented as a sequence of N_B observation vectors $\mathbf{x}_s = \left\{\mathbf{x}_s^b\right\}_{b=1}^{N_B}$ representing the consecutive blocks composing the strip. To model the sequence of horizontal strips X, we use a 1D-HMM with the emission probabilities represented using mixtures of gaussians. If we make the assumption that the feature vectors representing the blocks of a same strip are independent and are generated by the same distribution, the likelihood of the strip s for the state S_j can be estimated with:

$$P(\mathbf{x}_s|q_s = S_j) = \prod_{b=1}^{N_B} P(\mathbf{x}_s^b|\lambda_j) \tag{8}$$

Then, the likelihood of an observation sequence X is:

$$P(X|\lambda) = \sum_{\forall Q} \prod_{s=1}^{N_S} \prod_{b=1}^{N_B} P(\mathbf{x}_s^b|\lambda_j) \prod_{s=2}^{N_S} a_{q_{s-1},q_s} \tag{9}$$

where:

$$P(\mathbf{x}_b|\lambda_j) = \sum_{k=1}^{N_G} m_k^s \, \mathcal{N}(\mathbf{x}_s^b|\mu_k^s, \Sigma_k^s) \tag{10}$$

As shown in figure 2, this 1D-HMM structure can be seen as a particular case of 2D-HMM where the embedded HMM has only one state and equal transition probabilities. Note that the traditional 1D-HMM structure can also be represented as a 2D-HMM with only one embedded HMM state which emits the entire strip observation vector at once and with null self transition probability [16].

3.2 Application to Face Authentication

Let us denote the HMM parameter set for client C as λ_C, and the parameter set describing a generic face (non-client specific) as $\lambda_{\overline{C}}$. Given a claim for client C's identity and a set of feature vectors X supporting the claim (extracted from the given face), we find an opinion on the claim using:

$$\Lambda(X) = \log P(X|\lambda_C) - \log P(X|\lambda_{\overline{C}}) \tag{11}$$

where $P(X|\lambda_C)$ is the likelihood of the claim coming from the true claimant and $P(X|\lambda_{\overline{C}})$ is an approximation of the likelihood of the claim coming from an impostor. The generic face model is trained using data from many people. The authentication decision is then reached as follows: given a threshold τ, the claim is accepted when $\Lambda(X) \geq \tau$ and rejected when $\Lambda(X) < \tau$.

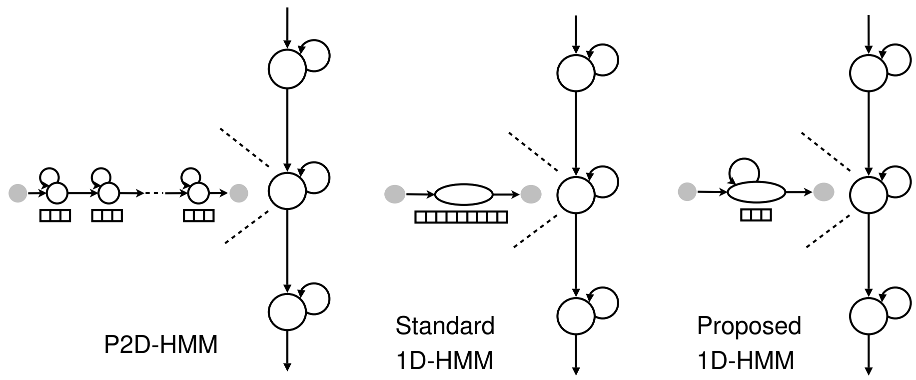


Figure 2: 1D-HMMs interpreted as P2D-HMM. The left figure shows the general P2DHMM where the emission probabilities of the main HMM are estimated through an embedded HMM. The figure in the middle shows a particular case of P2D-HMM which corresponds to a standard 1D-HMM where the embedded HMM consists just of one state and only one observation vector representing a whole strip is used. In the left figure the represented P2D-HMM corresponds to the proposed 1D-HMM where the observation vectors represents a block and the transition probabilities are equal.

3.3 HMM Training

It has been previously shown that the traditionally used Maximum Likelihood (ML) training approach has problems estimating robust model parameters when there are only a few training images available [2]. Considerably more precise models can be obtained through the use of Maximum *a Posteriori* (MAP) training [7]. A generic (non-client specific) model is *adapted* using client data. Given a set of training vectors, X, the probability density function (pdf) $P(X|\lambda)$ and the prior pdf of λ, $P(\lambda)$, the MAP estimate of model parameters, $\lambda_{\texttt{MAP}}$, is defined as:

$$\lambda_{\texttt{MAP}} = \arg\max_{\lambda} P(\lambda|X) \quad (12)$$

$$= \arg\max_{\lambda} P(X|\lambda)P(\lambda) \quad (13)$$

Assuming λ to be uniform is equivalent to having a non-informative $P(\lambda)$, reducing the solution of $\lambda_{\texttt{MAP}}$ to the standard ML solution. Thus, the difference between ML and MAP training is in the definition of the prior distribution for the model parameters to be estimated. Further discussion on MAP training is given in [2].

4 Experiments Setup

4.1 Preprocessing and Feature extraction

Based on given eye positions, a gray-scale 80×64 (rows $\times$ columns) face window is cropped out of each image. When using manually found eye positions, each face window contains the face area from the eyebrows to the mouth; moreover, the location of the eyes is the same for each face window (via geometric normalization). Fig. 1(a) shows an example face window.

Histogram equalization is used to normalize the face images photometrically. We then extract 2D Discrete Cosine Transform (DCT) [8] features from each image face. The feature extraction process is summarized as follows. The face window is analyzed on a block by block basis; each block is $N_P \times N_P$ (here we use N_P=8) and overlaps neighbouring blocks by a configurable amount of pixels. The degree of overlap has three main effects:

1. With a large overlap, the 2D DCT coefficients from a set of (horizontally or vertically) consecutive blocks will not vary abruptly.
2. A large overlap will increase dependance between consecutive blocks,

Test Sessions	Train Sessions			
	1	5	9	1,5,9
C: 2-4 I: 1-4	Mc			
C: 6-8 I: 5-8	Ud	Md		
C: 10-12 I: 9-12	Ua		Ma	
C: 2-4,6-8,10-12 I: 1-12	P			G

Table 1: Usage of the seven BANCA protocols (C: client, I: impostor). The numbers refer to the id of each session.

3. When using a large overlap, the parts of each face are in effect "sampled" at various degrees of translation, resulting in models which should be robust to minor translations of the faces.

A feature vector is then constructed with the 15 first DCT coefficients.

Note that in [2] DCTmod2 feature extraction [14] has been used; compared to traditional 2D DCT, the first three coefficients are replaced by their respective *delta features* in order to reduce the effects of illumination direction changes. DCTmod2 features make the system more robust; however the *delta features* are dependant on the number of overlapped pixels. Since in our system we use different horizontal and vertical overlaps (see section 5), the delta features would not be consistant. This motivated the choice of using standard 2D DCT features.

4.2 BANCA Database and Experiment Protocols

The multi-lingual BANCA database [1] was designed to evaluate multi-modal identity authentication with various acquisition devices under several scenarios. The database is comprised of four separate corpora, each containing 52 subjects; the corpora are named after their country of origin. Each subject participated in 12 recording sessions in different conditions and with different cameras. Each of these sessions contains two video recordings: one true claimant access and one impostor attack. Five "frontal" (not necessarily directly frontal) face images have been extracted from each video recording. Sessions 1-4 contain images for the *controlled* condition, while sessions 5-8 and 9-12 respectively contain *degraded* and *adverse* conditions.

According to the original experiment protocols, there are seven distinct configurations that specify which images can be used for training and testing. Table 4.2 describes the usage of different sessions in each configuration.

We believe that the most realistic cases are when we train the system in controlled conditions and test it in different conditions; hence in this paper we only performed experiments with configurations Mc, Ud, Ua and P. This limitation to four different scenarios should also make the results easier to interpret.

According to the BANCA experiment protocols, experiments should be performed on each corpus independently. The protocols further dictate that the subjects in each corpus are equally split into the validation and test sets. Subjects in the validation set are used to optimize the authentication system (e.g. to find the optimal number of gaussians and the decision threshold), while subjects from the test set are used for final performance evaluation. Note that this amounts to using only 26 subjects in the final stage. To increase the number of subjects, we merged the English and French corpora, resulting in a total of 104 subjects. In a similar manner to the original protocols, the resulting population was then divided into two groups of 52 subjects.

Authentication systems make two types of errors: a False Acceptance (FA), which occurs when the system accepts an impostor, or a False Rejection (FR), which occurs when the system refuses a true claimant. The performance is generally measured in terms of False Acceptance Rate (FAR) and False Rejection Rate (FRR).

To aid the interpretation of performance, the two error measures are often combined using the Half Total Error Rate (HTER), defined as:

$$\mathrm{HTER} = (\mathrm{FAR} + \mathrm{FRR})\,/2$$

Feature Vectors	Models	Protocol				Time for authentication of one claim (5 images)
		Mc	Ud	Ua	P	
Blocks DCTmod2 [2]	P2D-HMM	4.6	15.3	13.1	13.5	19.9s
Blocks DCT	P2D-HMM	5.6	18.8	13.2	15.1	19.9s
Blocks DCT	GMM	8.7	21.6	23.7	18.8	1.1s
Strips DCT	Standard 1D-HMM	6.6	20.1	20.0	16.6	1.3s
Blocks DCT	**Proposed 1D-HMM**	**5.4**	**16.1**	**17.2**	**15.1**	**2.5s**

Table 2: HTER performance for **manual face localization**
.

Since in real life the decision threshold has to be chosen *a priori*, it is selected to obtain Equal Error Rate (EER) performance (where FAR=FRR) on the validation set; it is then used on the test set to obtain a HTER figure.

5 Results and Discussions

For each client model, the training set was composed of five images extracted from the same video sequence. We artificially increased this to ten images by mirroring each original image. The generic model was trained with 571 face images (extended to 1142 by mirroring) from the Spanish corpus of BANCA (containing faces different from the English and French corpora), thus making the generic model independent of the subjects present in the client database. We compare the proposed 1D-HMM approach to GMM, standard 1D-HMM and P2D-HMM systems. Similarly to the systems presented in [2], an overlap of four pixels is used for the GMM approach while the strips are overlaped by seven pixels for the standard 1D-HMM and the blocks are overlaped by seven pixels for the P2D-HMM. For the proposed 1D-HMM approach, we choose to use an overlap of seven pixels between consecutive horizontal strips and we don't use horizontal overlap between the blocks. This choice is motivated by the assumption of independance between the blocks made in equation (8) since an overlap between concecutive blocks would increase the dependance.

5.1 Manual Face Localization

Table 2 shows the results in terms of HTER for manual face localization (we use the manually annotated eye center positions) and the time needed to perform an authentication. The authentication time is given in seconds for a claim which corresponds to five images in the BANCA protocol; the time includes the pre-processing and the experiments were performed on a Pentium IV 3 Ghz. It can be seen that the proposed 1D-HMM performs better than standard 1D-HMM or GMM approaches. The performances are similar to the P2D-HMM when the same features are used, however the computational cost for the proposed 1D-HMM is much less important. We note that better performances are obtained if we use the P2D-HMM with DCTmod2 features.

5.2 Automatic Localization

For automatic face localization experiments, we use the face detector proposed by Fröba and Ermst in [6]. The detector employs local features based on the *Modified Census Transform*, which represent each location of the image by a binary pattern computed from a 3×3 pixel neighbourhood. Face detection is carried out by analyzing all possible windows in the given image at different scales; each window is classified as either containing a face or not. The classification is performed by a cascade classifier similar to the approach proposed by Viola and Jones [15]; training of the classifier is accomplished using a version of the boosting algorithm. In our experiments the eye positions are inferred from the position and scale of the window with the best score at the last stage of the classifier. Note that this assumes that at most only one face is present in each image.

If all the windows were classified as containing the background, we consider that the given image does not contain a face and we perform the authentication using, if available, other images supporting the claim. If all given images are deemed not to contain a face, the claim is considered to have come from an impostor.

Results presented in Table 5.2 show that the proposed 1D-HMM is less affected by imperfect localization than the standard 1D-HMM which conserves a rigid spatial constraint in a same strip. The P2D-HMM approach

Feature Vectors	Models	Protocol			
		Mc	Ud	Ua	P
Blocks DCTmod2 [2]	P2D-HMM	6.5	15.9	14.7	14.7
Blocks DCT	P2D-HMM	5.7	18.8	16.5	16.5
Blocks DCT	GMM	11.0	25.2	27.9	22.5
Strips DCT	Standard 1D-HMM	20.8	29.3	31.0	27.3
Blocks DCT	**Proposed 1D-HMM**	**9.0**	**20.4**	**22.0**	**18.2**

Table 3: HTER performance for **automatic face localization**

.

System	Protocol			
	Mc	Ud	Ua	P
V 1D-HMM	5.4	16.1	17.2	15.1
H 1D-HMM	6.7	22.0	23.6	19.4
Combination	5.4	15.7	18.1	14.9

(a) HTER performance for **manual face localization**

System	Protocol			
	Mc	Ud	Ua	P
V 1D-HMM	9.0	20.4	22.0	18.2
H 1D-HMM	12.6	27.4	32.5	25.4
Combination	9.1	21.5	25.1	19.4

(b) HTER performance for **automatic face localization**

Table 4: HTER performance for Vertical (V) and Horizontal (H) 1D-HMM.

which performs an horizontal alignment between the blocks is more robust to imperfect localization with the cost of an authentication time much more important.

5.3 Vertical and Horizontal HMM

In previous sections of this paper, we made the choice to perform the main segmentation of the face image vertically. We made this choice since the main decomposition of the face is instinctively from top to the bottom (forehead, eyes, nose, mouth). However, the opposite choice has been made in [5]. It is interresting to see what are the performances if we use the proposed 1D-HMM to perform horizontal segmentation. In this case the image is decomposed in vertical strips, themself decomposed in blocks.

Since the computation time for this model is relatively low, we can combine Vertical and Horizontal 1D-HMM in order to improve the performances. The combined score for a claim is computed through the weighting sum of the likelihoods of horizontal and vertical models:

$$S_{comb} = wP(X_v|\lambda_v) + (1-w)P(X_h|\lambda_h) \tag{14}$$

where $P(X_v|\lambda_v)$ and $P(X_h|\lambda_h)$ are respectively the likelihoods estimated through the vertical and horizontal HMMs, w is the weight factor determined empirically on the validation set and S_{comb} is the combined score.

Results are presented in Table 4, they demonstrate that the natural decomposition of the face from top to bottom is also the most efficient for automatic authentication using 1D-HMM. Furthermore, we notice that the combination of horizontal and vertical models does not significantly improves the performance.

6 Conclusion

A alternative 1D-HMM variation is proposed for face recognition. This model allows the use of local features (blocks) as observation vectors instead of using a whole strip of the image for standard 1D-HMM implementation.

The experiments performed for a face authentication application demonstrate that this model is significantly more robust than the standard 1D-HMM. Due to its low complexity, it is also eight times faster than a P2D-HMM with the cost of a lower accuracy when an automatic localization system is used.

Two implementations of the proposed 1D-HMM are investigated, while the first one performs a vertical segmentation of the face, the second performs an horizontal segmentation. The results clearly demonstrate the superiority of the vertical segmentation.

Since feature extraction approaches including *delta features*, such as DCTmod2 [14], have shown to perform better than standard DCT decomposition, as a future work we plan to investigate a similar feature extraction that could be independant of the degree of overlap, and thus be consistant even if the horizontal and vertical amounts of overlap are not equal.

Acknowledgments

The author thanks Y. Rodriguez and S. Marcél for providing the face detector and for fruitful discussions and the Swiss National Science Foundation for supporting this work through the National Center of Competence in Research (NCCR) on Interactive Multimodal Information Management (IM2).

References

[1] E. Bailly-Baillière, S. Bengio, F. Bimbot, M. Hamouz, J. Kittler, J. Mariéthoz, J. Matas, K. Messer, V. Popovici, F. Porée, B. Ruiz, and J.-P. Thiran. The BANCA database and evaluation protocol. In *4th International Conference on Audio- and Video-Based Biometric Person Authentication, AVBPA*, Guilford, UK, 2003.

[2] F. Cardinaux, C. Sanderson, and S. Bengio. User Authentication via Adapted Statistical Models of Face Images. *To appear in IEEE Transaction on Signal Processing*, 2005.

[3] F. Cardinaux, C. Sanderson, and S. Marcel. Comparison of MLP and GMM Classifiers for Face Verification on XM2VTS. In *4th International Conference on Audio- and Video-Based Biometric Person Authentication, AVBPA*, pages 911–920, Guilford, UK, 2003.

[4] R. O. Duda, P. E. Hart, and G. S David. *Pattern Classification*. Wiley, 2001.

[5] S. Eickeler, S. Müller, and G. Rigoll. Recognition of jpeg compressed face images based on statistical methods. *Image and Vision Computing*, 18(4):279–287, 2000.

[6] B. Fröba and A. Ernst. Face detection with the modified census transform. In *Proc. IEEE Int. Conf. on Automatic Face and Gesture Recognition (AFGR)*, pages 91–96, 2004.

[7] J. L. Gauvain and C.-H. Lee. Maximum a posteriori estimation for multivariate gaussian mixture observation of markov chains, 1994.

[8] R. C. Gonzales and R. E. Woods. *Digital Image Processing*. Addison-Wesley, 1992.

[9] A. Nefian and M. Hayes. Face recognition using an embedded hmm. In *Proceedings of the IEEE Conference on Audio and Video-based Biometric Person Authentication, AVBPA*, pages 19–24, 1999.

[10] A Nefian and H. Monson. Maximum likelihood training of the embedded HMM for face detection and recognition. In *IEEE International Conference on Image Processing*, volume 1, pages 33–36, Vancouver, BC, Canada, 2000.

[11] H. Othman and T. Aboulnasr. A separable low complexity 2d hmm with application to face recognition. *IEEE Trans. on Pattern Recognition and Machine Intelligence*, 2003.

[12] L. R . Rabiner. A tutorial on hidden markov models and selected applications in speech recognition. In A. Waibel and K.-F. Lee, editors, *Readings in Speech Recognition*, pages 267–296. Kaufmann, San Mateo, 1990.

[13] F. Samaria. *Face Recognition Using Hidden Markov Models*. PhD thesis, University of Cambridge, 1994.

[14] C. Sanderson and K. K. Paliwal. Fast features for face authentication under illumination direction change. *Pattern Recognition Letters*, 24(14):2409–2419, 2003.

[15] P. Viola and M. Jones. Rapid object detection using a boosted cascade of simple features'. In *Proc. Int. Conf. Computer Vision and Pattern Recognition (CVPR)*, volume 1, pages 511–518, Seoul, 2001.

[16] K. Weber. *HMM Mixtures (HMM2) for Robust Speech Recognition*. PhD thesis, Ecole Polytechnique Federale de Lausanne, Lausanne, Switzerland, 2003.