Pseudo two-dimensional Hidden Markov Models for face detection in colour images

Stéphane Marchand-Maillet Bernard Mérialdo

Department of Multimedia Communications
EURECOM Institute
06904 Sophia-Antipolis, France
Stephane.Marchand@eurecom.fr
http://www.eurecom.fr/~marchand

To be presented in the
2nd Int. Conf. on Audio- and Video-based Biometric Person Authentication

Abstract

This paper introduces the use of Hidden Markov Models (HMM) as an alternative to techniques classically used for face detection. Our aim is to locate faces in colour images of a video sequence in view to indexing. The use of HMM in pattern recognition is first briefly reviewed and the mapping of these models onto our problem is presented. Pseudo two-dimensional HMM are presented and shown to be efficient and well-suited tools for performing face detection in a context where no constraints on face orientation are given. Issues about efficient face modelling are discussed and illustrated with practical examples.

1 Introduction

Automated face detection is the first necessary step for identifying persons present in an image. The aim is to reliably extract areas surrounding faces in the original image. Identification techniques such as Eigenfaces [10] typically assume that input images are normalised both in terms of size and illumination. The quality of face detection will therefore condition accurate subsequent identification.

A generic approach for localising faces in an image is as follows. Sub-images are selected at different locations and tested to represent a face or otherwise via a filter which associates with the centre of the sub-image the probability for a face to be centred at that point. The most commonly used filter for such a task is a Neural Network trained with face and non-face images (see *e.g.*, [8]). This approach represents an exhaustive search and can be refined with a top-down approach where pre-localisation takes place using subimages located on a coarse grid and subsequent interesting locations are re-investigated using a finer grid. Such systems are generally sensitive to face occlusion and in [4], an alternative is suggested that re-composes faces from face parts (*e.g.*, eyes and mouth) extracted in a preprocessing stage.

Another filter is given by the Eigenfaces approach. The likelihood of the subimage to contain a face is given by its distance from the "face state" previously determined. In [10], the combination of Eigenfaces and Neural Networks in the context of face detection and identification is also presented.

Most of applications consider grey-scale images as input so that they base the detection of faces on their shape only. This leads to the problem of false detections due to face-like parts of the background. The characterisation of a face therefore needs to be defined in a way that avoids confusion. In turn, techniques become more sensitive to face orientation. In this paper, we present an alternative approach based on Hidden Markov Models (HMM) where the aim is to embed extra information given by *e.g.*, colour and gradients at the face location. In Section 2, we first review the use of HMM in pattern recognition. Following this line, we present developments of such models adapted to the context of face detection in colour images (Section 3). The aim is to locate faces within images of a video sequence for indexing. In this context, no constraints are given on the background or the scale and orientation of the face in the image. In this paper, the aim is to test the ability of HMM to perform such a task rather than presenting a formal

face detection system. The technique is first introduced using one-dimensional models and extended to pseudo-dimensional models to take advantage of the specific structure of the images under investigation. Based on our experiments, Section 4 discusses extensions both in terms of the structure of the HMM used and in terms of characterisation of a face within an image through the design of feature vectors.

2 Previous work using HMM

Hidden Markov Models [5, 6] are stochastic models which provide a high level of flexibility for modelling the structure of an observation sequence. They allow for recovering the (hidden) structure of a sequence of observations by pairing each observation with a (hidden) state. State duration is left free so that HMM represent a powerful technique for realising elastic matching when imposing constraints on the topology of state transitions.

It is now acknowledged that the use of HMM is fundamental in automated speech analysis and recognition [3]. In this type of applications, the signal is mono-dimensional whereas pattern recognition in images requires two-dimensional operators. Nevertheless, mono-dimensional Hidden Markov Models (1DHMM) have been successfully applied to keyword spotting in binary document images [1]. The sequential aspect of written words is exploited (*i.e.*, from left to right). Each column of a word image is mapped onto a feature vector considered as a multi-dimensional observation. The sequence of such observations is then matched against different left-right models, each representing a keyword to be recognised. Recognition is based on the selection of the model of keyword that fits best the image in question. An extension of this work is found in [2]. The two-dimensional structure of the image is accounted for using a *pseudo* two-dimensional model (Pseudo 2D HMM or P2DHMM). This extension is shown to add robustness of the detection system against variations of size and slant of the fonts present in the document image.

Similarly, the potential of HMM for performing face recognition is demonstrated in [9]. The idea is again to exploit the sequential (vertical) structure of a human face. The image is divided in a sequence of overlapping horizontal stripes and the sequence of these stripes (*e.g.*, eyes-nose-mouth) is labelled using a 1DHMM. Results reported indicate that the use of HMM provides a suitable alternative to techniques classically used for this type of applications.

3 HMM for face detection

Based on the applications described above, we now show how HMM can be used in the context of face detection in colour images.

3.1 1DHMM

In a first approach, modelling is done at the line level. In an image containing a face, two types of lines are distinguished. Namely, lines composed of background pixels only and lines composed of a sequence of background and face pixels (see Figure 1(A)). These lines are labelled Background Line and Face Line, respectively. Two 1DHMM, λ_1 and λ_2 are therefore

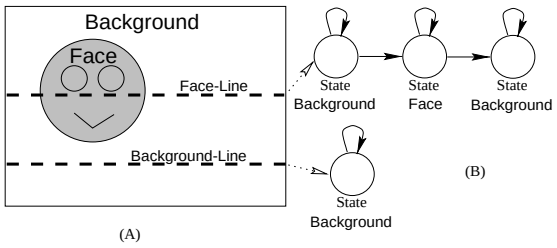


Figure 1: (A) A typical face image and (B) its model using 1DHMM.

used for modelling these lines (Figure 1(B)). To each state s_j^i of model λ_i is associated an output probability distribution $b_j^i(o)$ which represents the probability of producing observation o_{xy} (*i.e.*, values at pixel (x, y)) while being in state s_j^i. In our context, observations consist in feature vectors including chosen characteristics at a given pixel (*e.g.*, chrominance Cr-Cb values). Since observations are multi-dimensional, output probability distributions are assumed to be continuous and are approximated by Gaussian mixtures of the form:

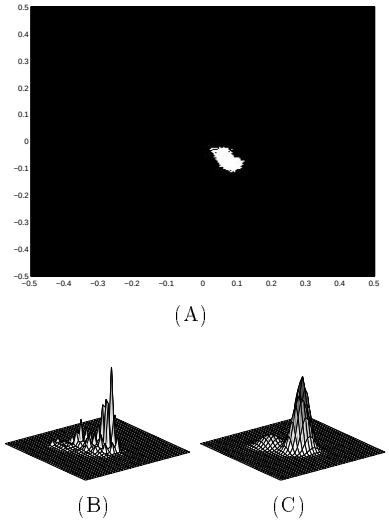
$$b_j^i(o_{xy}) = \sum_{m=1}^{M} c_{jm}^i \mathcal{N}(o_{xy}, \mu_{jm}^i, \Sigma_{jm}^i), \qquad (1)$$

where M is the number of mixtures, c_{jm}^i is the mixture coefficient for the mth mixture at state s_j^i and $\mathcal{N}(o_{xy}, \mu_{jm}^i, \Sigma_{jm}^i)$ is a Gaussian density with mean vector μ_{jm}^i and covariance matrix Σ_{jm}^i.

Both models are trained using training lines extracted from images segmented by hand. For each model λ_i, training consists in iteratively adjusting HMM parameters (state transition and output probabilities) using Baum-Welsh re-estimation procedure, in order to maximise $P[L_y|\lambda_i]$, the goodness-of-fit of model λ_i to the given line $L_y = \{o_{xy}\}_{x=1\cdots X}$.

Figure 2(A) shows the distribution of face colours in the (Cr,Cb) plane extracted from our training set.

The white areas show values that have a non-zero probability. This example shows that chrominance components are good cues for face pixel location since the white region is well-localised. Figure 2(B) details this part of the histogram and Figure 2(C) shows its approximation by Gaussian mixtures obtained with $M = 3$.



(A) (B) (C)

Figure 2: Face chrominance distribution and its modelling ($M = 3$).

Recognition is achieved by selecting for each line L_y of the test image $I = \{L_y\}_{y=1 \cdots Y}$, which of models λ_1 or λ_2 fits best L_y. More formally, we use Viterbi procedure for calculating $P[L_y|\lambda_i]$, $i = 1, 2$ and select the model which corresponds to the maximum of these values. Once the best model is selected, the line is segmented using the structure recovered by the HMM (*i.e.*, the sequence of hidden states s_j^i leading to the highest value of $P[L_y|\lambda_i]$). In our example, the result is therefore a binary image containing pixels labelled with Face state and pixels labelled with Background state. Figure 3 shows the segmentation of a test image obtained using the model presented in Figure 1 and the luminance-chrominance components (YCrCb) as feature vectors. Background pixels of the segmented image have been whitened for illustrating the segmentation.

This example illustrates the efficiency of the segmentation while using a very simple model. It also shows that the segmentation lacks coherence in the vertical orientation. Extra regions (red coloured part of the robe) are segmented that do not fit with the model presented in Figure 1. Morphological processing of the binary image may overcome this problem. However, a more formal approach is used in this work

(A) (B)

Figure 3: (A) Original image. (B) Segmentation using the 1DHMM presented in Figure 1.

and is based on the following principle. The fact that no relationship is given between a line and subsequent lines implies that the 2D structure of the image is ignored in the model. The next section presents a pseudo two-dimensional technique that introduces such a dependency between lines.

3.2 P2DHMM

Hidden Markov Models retrieve optimal transitions between discrete states using a given topology for connectivity relationships between states. It is known that a fully connected two-dimensional structure for the HMM would lead to exponential complexity when retrieving the best (2D-) state sequence for producing a given two-dimensional sequence of observations. In [7], it is also assessed that not only such structure would lead to an unmanageable complexity but the retrieval of the best 2D state sequences is not ensured. We therefore use a *pseudo* two-dimensional model (P2DHMM) Λ where dependency in the second dimension is made at the line level. More formally, 1DHMM λ_i are again used for modelling each line and considered as super-states S_i of a vertical (upper-level) 1DHMM. The aim of this (vertical) HMM is to impose constraints on the optimal sequence of (horizontal) line models which maximise $P[I|\Lambda]$. In other words, the image is now considered as a whole when retrieving the sequence of hidden states from which its (2D-)structure will be recovered. The P2DHMM corresponding to the face image model presented in Figure 1 is show in Figure 4.

This model is again trained using images from a training set segmented by hand. Re-estimation of parameters is done using two nested Baum-Welsh procedures. For each line L_y of an image I, $P[L_y|S_i]$, the probability of generating the L_y using the 1DHMM λ_i (*i.e.*, while being in super-state S_i) is calculated. By this mean, the output probability distribution corresponding to (super-)state S_i is formed so that training of the (vertical) line-level 1DHMM can be performed

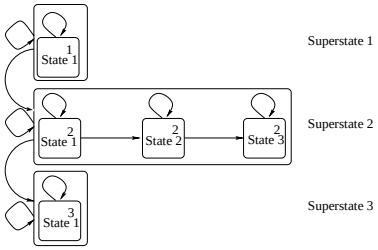


Figure 4: Pseudo two-dimensional HMM Λ corresponding to the face image model shown in Figure 1.

using another Baum-Welsh procedure.

Similarly, recognition is done using two nested Viterbi procedures. Likelihood of model λ_i to generate each line L_y (*i.e.*, the probability of being in super-state S_i at line L_y) is calculated and the most likely super-state sequence determined. Each line is then segmented similarly to the 1D case using the 1DHMM corresponding to the super-state in this sequence. The result is shown in Figure 5 where the test image is segmented using the model depicted in Figure 4.

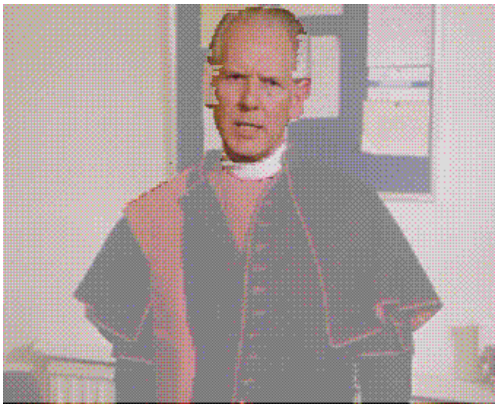


Figure 5: Segmentation of the test image (Figure 3(A)) using the P2DHMM presented in Figure 4.

By imposing a vertical structure to the line model used, coherence in the two-dimensional structure is better assumed by the model.

4 Models and discussion

In this section, we present and discuss important features that are to be included in HMM in order to perform accurate face detection.

In our context, the aim of Hidden Markov Models is to associate a label with each pixel for performing image segmentation. In the specific case of face detection and in the simplest case, one wishes to obtain a binary image indicating the location of face pixels in the original image. It is therefore necessary that output probabilities are coupled between different line models. Following the model proposed in Figure 4, states s_1^1, s_2^1, s_2^3 and s_3^1 all represent Background pixels. A single output probability will therefore be used by all these states in order to have optimal characterisation during training. Similarly, at the line level, 1DHMM used for modelling super-states Background Line are to be equivalent. During training, both models will therefore be updated with the same parameters. In practice, this is done using the concept of object, so that one item is defined and its occurrences are symbolled by pointers.

At each state of a HMM, a feature vector is emitted with a certain probability given by the output probability distribution. This distribution is modelled by Gaussian mixtures as given by Equation (1). Different strategies can be adopted for handling the type of components of the feature vectors. In the case where the feature vector contain independent groups of components (*e.g.*, chrominance Cr-Cb and gradient norm), the covariance matrix Σ_{jm}^i can be forced to be block-diagonal during parameter re-estimation. Another way of handling this case is to consider each group of components as an independent stream of observation and to combine these streams as a weighted sum. The output probability distribution then becomes

$$b_j^i(o_{xy}) = \sum_{s=1}^{S} w_{js}^i \sum_{m=1}^{M_s} c_{jsm}^i \mathcal{N}(o_{xy}, \mu_{jsm}^i, \Sigma_{jsm}^i). \quad (2)$$

Weights w_{js}^i therefore represent the amount of information contained in stream s at state s_j^i. Estimation of these parameters can be embedded in the Baum-Welsh procedure so that these weights automatically adapt to the context. By this mean, one may include in each state any characteristic he feels it is relevant and the HMM will automatically select those which are actually relevant. For example, gradient information at a pixel may be useful for detecting state change. Although it is related to luminance, it can be considered as an independent component of the feature vector. It is however difficult to assess what is the importance of this component within each state. Training with multi-stream output probability distribution as given by Equation (2) which helps in resolving this problem.

This can be used in a model like that shown in Figure 6, where the Border state will impose conditions on the gradient at state change. Note that transitions from the state Border onto itself are not permitted in super-states S_3. Similarly, transitions from super-

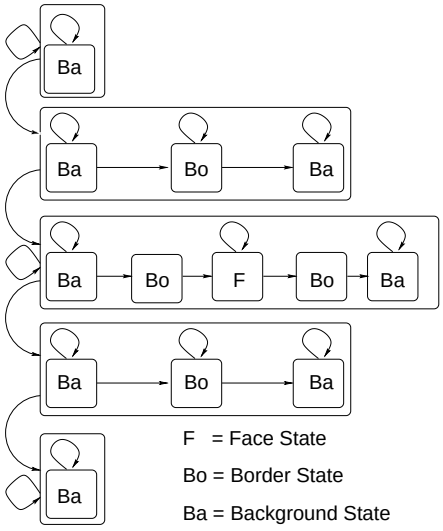


Figure 6: P2DHMM including border characterisation.

states S_2 and S_4 onto themselves are not allowed. This ensures that the width of the border will be of one pixel exactly. Using this simple principle, one may explicitly specify state duration (*i.e.*, minimal, exact or maximal width of a segmentation region) by duplicating states or by defining transition that allow for skipping some states in the sequence. This flexibility in terms of designing the structure allows for the use of P2DHMM in specific applications without complicating the theoretical aspect. For example, Figure 7 shows an instance of false segmentation. This is due to the face that more than one skin-coloured area are present and it is possible to connect them vertically using a false Face state path. More formally, at lines where the second area is present, the probability for the line to be generated by the model Face Line (*i.e.*, super-state 2 in Figure 4) is high. Since no return to this super-state is permitted once it has been left (left-right model), super-state Face Line is kept for lines between the two areas and one pixel only is used for representing the Face state. Forcing the Face state duration to be longer by duplicating this state resolves this problem as shown in Figure 7(B) where the model Face Line includes 20 times the Face state.

Another important aspect of the use of HMM in face detection is that they result in a precise segmentation of the original image. If a face area is to be extracted for subsequent face identification, one will be able to remove (or lower) the effect of the background from the subimage. Thus making the identification more precise. By contrast, techniques based on Neural Networks do not provide such a capability and

(A) (B)

Figure 7: Forcing state duration to 20 pixels.

typically remove the effect of the background using a fixed mask where corner are chopped off (see *e.g.*, [8]). Figure 8 shows different examples of faces segmented using P2DHMM. Only Face pixels have been left intact and the minimal surrounding box have been extracted. Pictures have been normalised to the same height for display and their quality is related to their size within the original image.

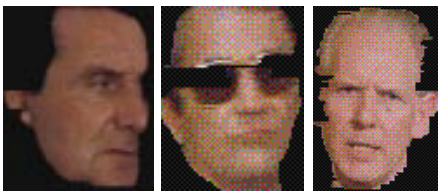


Figure 8: Segmented faces extracted using P2DHMM.

Note that even when eye-glasses are present, the face is still well-segmented. Since our models are not based on geometrical considerations, face orientation does not influence face detection. However, a limitation of this principle is illustrated by Figure 9. In this example, skin-colour pixels are present at almost every line of the image. Since the dependency between lines is introduced at the line level rather than the pixel level, this allows for horizontally disconnected regions to be segmented. This is the case in this example where super-state Face line is used for almost all the lines.

The next step will therefore be to introduce dependency at the pixel level for refining face characterisation using geometrical criteria. We are currently working at defining a way to add this dependency within the feature vector so that computation load remain reasonable.

5 Conclusion

In this paper, we investigated the problem of face detection in colour images. The aim in to include such application in an automated video-indexing process.

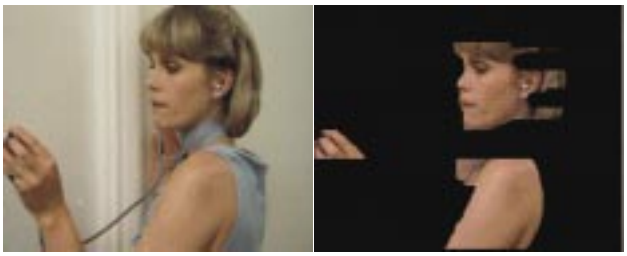


Figure 9: A case of unsucessful segmentation.

In this context and unlike in most face detection applications, no prior knowledge about the face orientation and scale is available. After reviewing briefly the use of HMM for pattern recognition in 1D and 2D signals, we detailed the application of P2DHMM in our context. It was shown that such model offer a high level of flexibility in term of face orientation and background quality. We also presented a technique which allow for the introduction of several characteristics and automatically selects the best ones for the context in question. It was shown that HMM have the capability of segmenting this input image at the pixel level. Subsequent face identification may then take advantage of this capability by lowering the effect of what is known to be background pixels.

We concluded the paper by presenting some limitations that we are currently working at overcoming.

Acknowledgements

Eurecom's research is partially supported by its industrial partners: Ascom, Cegetel, France Telecom, Hitachi, IBM France, Motorola, Swisscom, Texas Instruments, and Thomson CSF.

References

[1] S.-S. Kuo and O. E. Agazzi. Automatic keyword recognition using Hidden Markov Models. *Journal of Visual Communication and Image Representation*, 5(3):265–272, 1994.

[2] S.-S. Kuo and O. E. Agazzi. Keyword spotting in poorly printed documents using Pseudo 2-D Hidden Markov models. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, PAMI-16(8):842–848, 1994.

[3] K.-F. Lee. *Automatic speech recognition*. Kluwer Academic Publishers, Dordrecht, The Netherlands, 1989.

[4] S. Oka, M. Kitabata, Y. Ajioka, and Y. Takefuji. Grouping complex face parts by nonlinear oscillations. In *Proceedings of the European Symposium on Artificial Neural Networks*, pages 395–400, Bruges, Belgium, April 22-24 1998.

[5] L. R. Rabiner. A tutorial on Hidden Markov Models and selected applications in speech recognition. *Proceedings of the IEEE*, 77(2):257–285, 1989.

[6] L. R. Rabiner and B.-H. Juang. *Fundamentals of Speech Recognition*. Prentice Hall, Englewood Cliffs, NJ, 1993.

[7] G. Rigoll, S. Müller and C. Neukirchen. Spotting of handwritten symbols in complex environments using Pseudo-2D Hidden Markov Models. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, Santa Barbara, CA, USA, 1998.

[8] H. A. Rowley, S. Baluja and T. Kanade. Neural-Network-based face detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, PAMI-20(1):23–38, 1998.

[9] F. S. Samaria and A. C. Harter. Parameterisation of a stochastic model for human face identification. In *Proceeding of the Second IEEE Workshop on Applications of Computer Vision*, Sarasota, Florida, December 1994.

[10] M. Turk and A. Pentland. Eigenfaces for recognition. *Journal of Cognitive Neuroscience*, 3(1):71–86, 1991.