In: Proceedings of the 2nd IEEE Workshop on Applications of Computer Vision. December 1994, Sarasota (Florida).

Parameterisation of a Stochastic Model for Human Face Identification

F. S. Samaria*†

A. C. Harter†

*Engineering Department
University of Cambridge
Trumpington Street
Cambridge, UK CB2 1PZ

†Olivetti Research Limited
Old Addenbrooke's Site
24a Trumpington Street
Cambridge, UK CB2 1QA

Abstract

Recent work on face identification using continuous density Hidden Markov Models (HMMs) has shown that stochastic modelling can be used successfully to encode feature information. When frontal images of faces are sampled using top-bottom scanning, there is a natural order in which the features appear and this can be conveniently modelled using a top-bottom HMM. However, a top-bottom HMM is characterised by different parameters, the choice of which has so far been based on subjective intuition. This paper presents a set of experimental results in which various HMM parameterisations are analysed.

1 Introduction

Stochastic modelling of non-stationary vector time-series based on HMMs has been very successful for speech applications [5]. Recently it has been applied to a range of image recognition problems [7, 9]. Previously reported work [6] has investigated the use of HMMs to model human faces for identification purposes. Faces can be intuitively divided into regions such as the mouth, eyes, nose, etc., and these regions can be associated with the states of an HMM. The identification performance of a top-bottom HMM compares favourably with some of the well-known algorithms, for example eigenfaces as detailed in [8]. However, the HMM parameterisation in the work presented so far was arrived at by subjective intuition. This paper presents experimental results which show how identification rates vary with HMM parameters, and which indicate the most sensible choice of parameters. The paper is organised as follows: section 2 gives an overview of the HMM-based approach; section 3 details the training and recognition processes; section 4 describes the experimental setup; section 5 presents the identification results; section 6 concludes the paper.

2 The HMM-Based Approach

2.1 The Sampling Technique

An HMM provides a statistical model for a set of observation sequences [4]. To use HMMs in the context of face identification, an observation sequence is extracted from a face using the sampling technique illustrated in figure 1. A one-dimensional (1D) vector

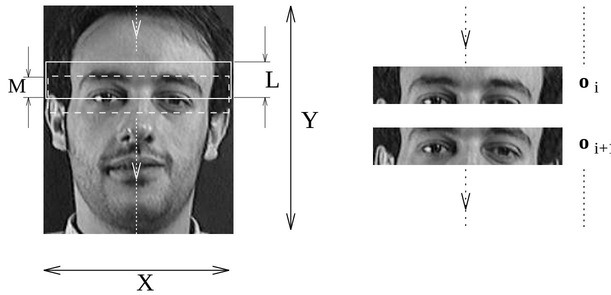

Figure 1: Sampling Technique

series of pixel observations $\mathbf{O} = \mathbf{o}_1 \ldots \mathbf{o}_T$ is generated, where each observation $\mathbf{o}_i$ contains the values of the pixels in the block of lines arranged in a column-vector. Therefore each observation vector is a block of L lines and there is an M-line overlap between successive observations. The length of the observation sequence T is given by:

$$T = \phi\left(\frac{Y-L}{L-M}\right) + 1 \tag{1}$$

where $\phi(x)$ is the largest integer r such that $r \leq x$ (i.e. ϕ is the round down function). This means that if it is not possible to fit an integral number of observations in the image, then some of the bottom lines are not used. Assuming that each face is in an upright, frontal position, features will occur in a predictable order, i.e. forehead, then eyes, then nose, and so on. This ordering suggests the use of a top-bottom (non-ergodic) model, where only transitions between adjacent states in a top-bottom manner are allowed as discussed in [8].

2.2 A Parameterised HMM Model

For face images of fixed size there are three HMM parameters which affect the performance of the model: the number of HMM states N, the height of the sampling window L and the amount of overlap M. Using shorthand notation, a model with such parameters will be defined as:

$$\mathcal{H} = (N, L, M)$$

In the work reported so far, little experimental evidence was put forward to support the choice of specific values for $\mathcal{H}$ and only subjective attempts were made to justify the choice of the parameters. The parameterisation of the model can determine how successful the model is. In [6] it was argued that the number of states N should be 5, since by inspection approximately that many distinct regions appear in the face. In that paper, preliminary segmentation and recognition results were presented in support of the choice $N = 5$. In more recent work [8] the segmentation of the training data was used subjectively to choose the value of L. Here experimental results are presented to support the choice of specific parameters for $\mathcal{H}$.

3 Applying the HMM Technique

3.1 Training Phase

All the HMM-based experiments reported in this paper were carried out using the *HTK: Hidden Markov Model Toolkit V1.3* developed by Young [10] at the Cambridge University Engineering Department. The training process for each of the S subjects in the database consists of the following steps which are summarised in the diagram of figure 2:

1. J training images are collected for the kth subject in the database and are sampled to generate J distinct observation sequences.

2. A common prototype HMM model λ_0 is constructed with the purpose of specifying the number of states in the HMM, the state transitions allowed and the size of the observation sequence vectors.

3. A set of initial parameter values using the training data and the prototype model are computed iteratively. On the first cycle, the data is uniformly segmented and matched with each model state. On successive cycles, the uniform segmentation is replaced by Viterbi [2] alignment. The outcome of this process is an initial HMM estimate $\lambda_e^{(k)}$ which is used as input to the re-estimation stage.

4. HMM parameters are re-estimated using the Baum-Welch [1] method. The model parameters are adjusted so as to locally maximise the probability of observing the training data, given each corresponding model. The outcome of this process is the HMM $\lambda^{(k)}$ which is used to represent subject k in the database.

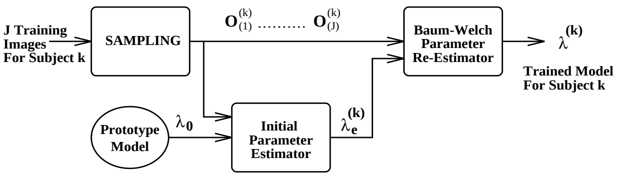

Figure 2: Block diagram of training technique

3.2 The Recognition Phase

Recognition is carried out via a simple Viterbi recogniser. A collection of HMMs each representing a different subject is matched against the test image and the highest match is selected. The recognition process consists of the following steps which are summarised in the diagram of figure 3:

1. The unknown test image is sampled to generate an observation sequence $\mathbf{O}_{test}$.

2. The observation sequence is matched against each face model by calculating the model likelihoods:

$$P(\mathbf{O}_{test} \mid \lambda^{(k)}), \quad 1 \leq k \leq S$$

3. The model with the highest likelihood is selected and this model reveals the identity of the unknown face.

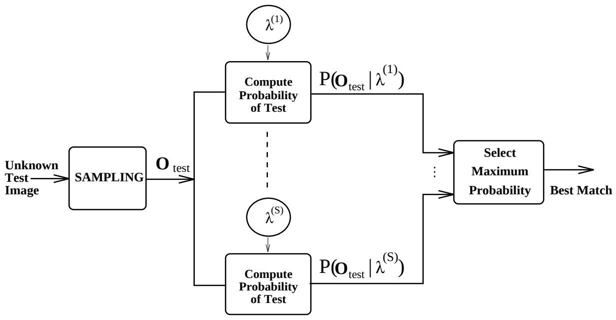

Figure 3: Block diagram of face recogniser

4 Experimental Setup

4.1 The Subject Database

Experiments with different models $\mathcal{H}$ were carried out using the Olivetti Research Ltd (ORL) database of faces. The database consists of 400 images, 10 each of 40 different subjects. The subjects are either Olivetti employees or Cambridge University students. The age of the subjects ranges from 18 to 81, with the majority of the subjects being aged between 20 and 35. There are 4 female and 36 male subjects. Subjects were asked to face the camera and no restrictions were imposed on expression; only limited side movement and limited tilt were tolerated. For most subjects the images were shot at different times and with different lighting conditions, but always against a dark background. Some subjects are captured with and without glasses. The images were manually cropped and rescaled to a resolution of 92x112, 8-bit grey levels. Five images of each subject were used for training and five for testing, giving a total of 200 training and 200 test images.

4.2 Experiment Plan

For each model $\mathcal{H} = (N, L, M)$, the results of the 200 identification tests are reported as an error rate. Each error rate is calculated as the proportion of the images which are misclassified, where a lower error rate obviously indicates a better model. An exhaustive exploration of all possible combinations of N,L and M would require a considerable number of experiments. It is evident that only a subset of the full parameter range needs to be investigated, as inter-parameter dependency constrains the parameter range for meaningful results. For example, the size of the window L directly constrains the possible values of the overlap M ($0 \leq M \leq L-1$). Moreover, both L and M determine the length of the observation sequence T as can be seen from equation 1 and this constrains the choice of number of states N. Given the image size, it was decided to experiment with parameters in the following range:

$$\begin{array}{rcl} 2 \leq & N & \leq 10 \\ 1 \leq & L & \leq 10 \\ 0 \leq & M & \leq L-1 \end{array}$$

When experimenting within this parameter range, it was assumed that, to a certain extent, parameters could be varied independently.

5 Parameterisation Results

5.1 Varying the Overlap M

A model with no overlap implies that training and test faces are partitioned into rigid, arbitrary regions with the risk of cutting across potentially discriminating features. In a top-bottom model with no overlap, features require accurate alignment for successful results. Alignment in images of the same subject is preserved either if the features occupy the exact same position in all the images or if the features are vertically displaced by a number of pixels which is a multiple of L. Unless the images are preprocessed, the features will normally not be in the same position. Therefore in most cases alignment is preserved only if the vertical displacement is a multiple of L. Overlap during the sampling process has the following main functions:

1. The overlap determines how likely feature alignment is and it is expected that a large overlap would increase the likelihood of preserving the alignment.

2. Given a fixed image size and window height, the overlap determines the length of the observation sequence T as can be seen from equation 1.

A larger value of M produces a larger T, because the face regions are oversampled hence increasing the length of the training and test data observations. The accuracy of the model estimates depends on the number of observations T in the training data. If T is small, the accuracy will be limited because there are not enough occurrences of model events. It may therefore be advantageous to use a larger value of M.

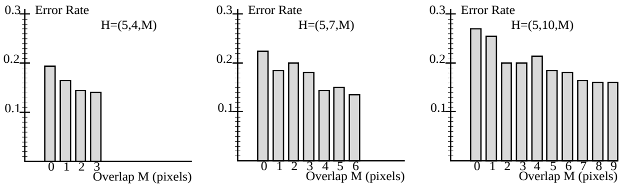

Figure 4: Results obtained for varying M

In order to determine the effect of M on the recognition performance, a comprehensive set of experiments were run with the number of states fixed to $N = 5$ as discussed in section 2.2, window height in the range $2 \leq L \leq 10$ and every possible overlap $0 \leq M \leq L-1$. The results summarised in figure 4 are representative. Recognition performance appears to improve as the overlap increases, which is in accordance with expectations. A greater overlap, however, implies a larger value of T and the order of calculations required in the identification process varies linearly with T.

5.2 Varying the Window Height L

The window height L has the following functions:

1. It determines the size of the features that the model extracts.
2. For a fixed image size and overlap, L determines the length of the vector series as can be seen from equation 1.

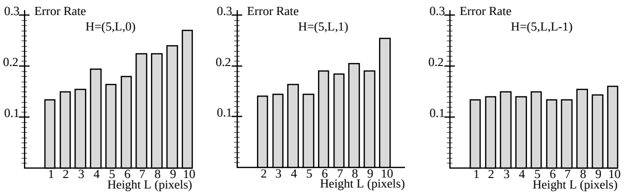

Figure 5: Results obtained for varying L

The experiments reported in this section consider the cases with no overlap, one line overlap and maximum overlap $L-1$. The number of states was still kept to $N = 5$. If the sampling window height is sufficiently smaller than the image height Y, then the length of the observation sequence will be large. In this case the value of L is expected to have a limited effect on the identification performance since the overlap guarantees that features are aligned. The histograms of figure 5 show the results obtained for the experiments mentioned above. From the results it appears that, for sufficiently large overlap, the window height has a marginal effect on the recognition performance. The effect of the window height becomes more noticeable when there is little or no overlap. In both cases, as the window size increases the error rate also increases.

5.3 Varying the Number of States N

The number of states N in a top-bottom HMM determines the number of features used to characterise the face. If the number of observations in a sequence is very large, then we can choose a large N to capture more features. However, the computational complexity of the identification algorithm varies as N^2 and therefore the smaller the value of N the faster the identification.

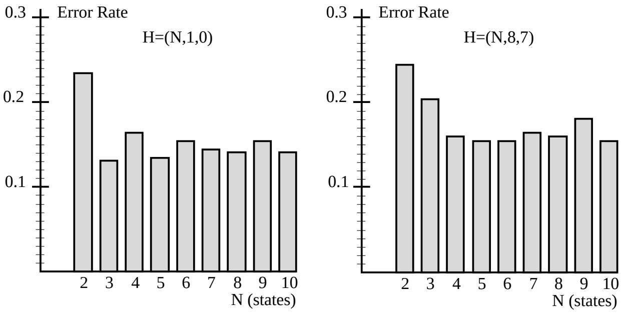

Figure 6: Results obtained for varying N

In the experiments presented so far, it was assumed that $N = 5$ was a reasonable value to use. The experiments presented in this section analyse the variation of recognition performance as the number of states varies. Two cases are investigated: the smallest possible window with $L = 1$ and a medium size window $L = 8$ with maximum overlap $M = 7$. The results are presented in figure 6. The performance is fairly uniform for the values $4 \leq N \leq 10$, while the error rate increases for values of N smaller than 4.

6 Conclusions and Summary

HMMs have been used with some success in the area of face identification. However, some of the aspects that determine the parameterisation of the model had so far been assessed subjectively and intuitively. This paper has presented some experiments through which different parameterisations have been assessed using their success rates in identifying 200 images from a database of 40 people. The preliminary results reported in this paper seem to indicate that:

- Large overlap in the sampling results in better recognition performances.
- As the overlap becomes noticeable, the effect of the window height is limited.
- Best results are obtained with 4 or more states.

Present work is concentrating on extending the insights gained from experimenting with top-bottom HMMs to a technique using pseudo-2-dimensional HMMs [3]. Preliminary experiments using the same test data have given already very encouraging results, with error rates below 0.05.

Acknowledgements

Many thanks to Steve Young, for closely supervising this work.

This work was supported by a Trinity College Internal Graduate Studentship and an Olivetti Research Limited CASE award. Their support is gratefully acknowledged.

References

[1] L.E. Baum. An inequality and associated maximization technique in statistical estimation for probabilistic functions of markov processes. *Inequalities*, III:1–8, 1972.

[2] G.D. Forney. The viterbi algorithm. *Proceedings of the IEEE*, 61,3:268–278, March 1973.

[3] S. Kuo and O.E. Agazzi. Machine vision for keyword spotting using pseudo 2d hidden markov models. *Proceedings of ICASSP'93*, V:81–84, 1993.

[4] L.R. Rabiner. A tutorial on hidden markov models and selected applications in speech recognition. *Proceedings of the IEEE*, 77,2:257–286, January 1989.

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

[6] F. Samaria. Face segmentation for identification using hidden markov models. In *British Machine Vision Conference 1993.* BMVA Press, 1993.

[7] F. Samaria and F. Fallside. Face identification and feature extraction using hidden markov models. In G. Vernazza, editor, *Image Processing: Theory and Applications.* Elsevier, 1993.

[8] F Samaria and S. Young. A hmm-based architecture for face identification. *TO APPEAR IN: Image and Vision Computing*, 12:8, October 1994.

[9] J. Yamato, J. Ohya, and K. Ishii. Recognizing human action in time-sequential images using hidden markov model. *Proceedings of CVPR'92*, pages 379–385, 1992.

[10] S.J. Young. *The HTK Hidden Markov Model Toolkit: Design and Philosophy.* Technical Report TR.152, Cambridge University Engineering Department, 1993.