Face Identification and Feature Extraction Using Hidden Markov Models

F. Samaria[a,b] and F. Fallside[a]

[a]Cambridge University Engineering Department, Trumpington Street,
Cambridge CB2 1PZ, United Kingdom, tel.+44 223 332752

[b]Olivetti Research Lab, Old Addenbrookes Site, 24a Trumpington Street,
Cambridge CB2 1QA, United Kingdom, tel.+44 223 343000

Abstract

This paper details work done on automatic face identification. A new approach to the problem is proposed involving the use of Hidden Markov Models. We illustrate how these models allow the automatic extraction of facial features and the classification of face images. Some experiments are presented to support the plausibility of this approach. Successful results were obtained under the constraints of homogeneous lighting and constant background.

1. INTRODUCTION

Substantial research effort has gone into trying to understand how to build a successful model for face identification, both by psychologists [1] and information scientists. The importance of face identification, other than for the intrinsic challenge of the task, stems from its numerous potential applications: workstation security [2], criminal identification, video-mail retrieval, building security and home video actor-based indexing. We present a new architecture to classify face images that uses Hidden Markov Models (HMMs). Faces are treated as two dimensional objects and the model automatically extracts *statistical* facial features. However, it can at the same time make use of *structural* information, yielding an *hybrid* model.

2. PROPOSED APPROACH

We propose a method that automatically extracts facial features from a set of training data and successively uses them to identify test images. We achieve this using a particular context-dependent classifier [7] based on Hidden Markov Models (HMMs). Rabiner [4] provides a comprehensive tutorial on HMMs. They have been extensively and successfully used for speech recognition applications [3]. The one dimensional (1D) nature of speech signals (along the time axis) is suited to analysis by HMM. A 1D observation sequence can be obtained from an image by sampling it using a sliding window, where each element of the sequence is a vector of pixel intensities. The way in which the sliding window passes over the image determines the ordering of the sequence. The windowed data generates a spatial sequence [5] that can be used to train an HMM, which will then partition the sequence into a number of feature states. The term *feature state* here refers to a block of pixels (of the same size as the sliding window) which has sufficient statistical significance within the image to result in the HMM assigning a state to represent it.

An HMM is primarily characterised by a transition probability matrix **A** and an output probability matrix **B** [4]. After training an HMM on a sequence generated as above, each matrix will capture a different aspect of the face image. Matrix **A** will record the occurrences of transitions from one feature state to another across the face and the thickness of the various features. The latter can be seen by considering the terms representing the probability of going from a state to itself, *i.e.* the probability of staying in the same feature state. Matrix **B** will record what the feature states look like for each subject. Both matrices will provide strong discrimination for the various subjects.

3. INITIAL EXPERIMENTAL RESULTS

Two sets of simple spatial HMMs were built to test some of the ideas presented in the previous section. In both cases, a separate HMM was trained for each subject in the database and the resulting models were used to classify unknown images. The training and testing were carried out using the *HTK: Hidden Markov Model Toolkit V1.3* [8].

3.1. Ergodic Models

A total of 10 training images (of size 256×256, 8-bit grey levels) for each subject were used to train an HMM. Each image was sampled into an observation sequence using a 64×64 window. The window initially moved left to right stepping 48 pixels at a time, until it hit the right margin of the picture. It then moved 48 pixels down and started sampling right to left, and so on. This implies an overlap of 16×64 pixels when sliding horizontally and 64×16 pixels when sliding vertically.

Figure 1: Training data and magnified states for ergodic model

An 8-state ergodic HMM was built. The reason for choosing 8 states was that, by inspection, approximately 8 distinct regions seem to appear in the face image (eyes, mouth, forehead, hair, background, shoulders and two extra states for boundary regions). Figure 1 shows some training data, together with the 8 states found by the HMM. Some states, such as the background, the eye and the neck, are easily recognisable. Others have been generated by superposition of features and are less distinguishable. They nonetheless characterise the subject, providing strong discrimination.

3.2. Left-to-right Models

A total of 5 training images (of size 184×224, 8-bit grey levels) per subject were used to train each HMM. Each image was sampled into an observation sequence using a block of 16 lines, moving vertically 4 lines at a time (*i.e.* with a 12-line overlap). Each still image was effectively converted into a sequence of overlapping line blocks spatially ordered in a top-to-bottom sense. The overlapping was present to allow the model to capture features in a manner independent of their vertical position (a disjoint partitioning of the image could result in features occurring across block boundaries being truncated).

This sequence was then partitioned into 5 states, each representing a *feature band*. We wish to establish a correspondence between statistical feature bands and facial bands as understood by humans. Assuming that each face is in an upright, frontal position, we expect facial bands to occur in a predictable order, *i.e.* forehead, then eyes, then nose, and so on. This form of ordering suggests the use of a left-to-right HMM where only transitions between adjacent states in a left-to-right manner will be allowed. Figure 2 shows the training data for one subject in the database and the means for the 5 states found by the HMM for that subject. It is interesting to notice that we can identify some

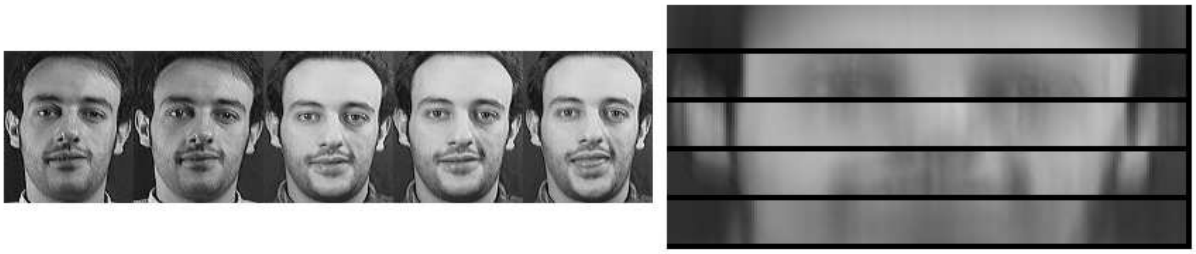

Figure 2: Training data and magnified states for left-to-right model

of the states as physical facial features. We can, for instance, identify the second state as being the eye band. This means that the eyes have sufficient statistical significance to induce the HMM to assign a state to represent them. Experiments [6] have shown evidence that one of the salient feature areas of the face for identification used by humans is the eye region. The remaining states extracted by the HMM can also be identified with other facial regions. Some states "leak" into their next one: it is possible, for example, to see part of the eye state lightly appearing in the next band. This is due to those observation vectors (or line blocks) which have, as far as state allocation is concerned, ambiguous statistical properties and actually give a better statistical match if placed in the next state. They will therefore contribute to the overall mean for that state vector thus causing the leaking effect.

4. CONCLUSIONS

Successful identification results were obtained with a database of 20 subjects when the test image was extracted from the same video sequence as the training images. This demonstrates that, provided lighting and background are kept the same, the model can cope with variations in facial features due to small orientation changes. This has been a considerable problem in all template-based approaches to face identification.

This method also provides a way of extracting statistical facial features. It is interesting to note that such statistical features may correspond to ordinary features (*e.g.* the eyes). Therefore this approach, if appropriately tuned, could be used to extract facial features automatically.

Acknowledgements

This work is supported by a Trinity College Internal Graduate Studentship and an Olivetti Research Lab CASE award. Their support is gratefully acknowledged. We wish to thank many of our colleagues in Cambridge: Andy Hopper, Leslie French, Phil Elwell and Martin Brown of Olivetti Research, for useful discussions and for the image capture software; Steve Young and the Speech Group at CUED for the HMM software; Owen Jones and Gabor Megezy of the Pure Mathematics Department, Barney Pell of the Computer Lab, and Steve Hodges, Gopal Chand and Charlie Allen of CUED for the many ideas discussed together.

5. REFERENCES

[1] V. Bruce. *Recognising Faces.* Laurence Erlbaum Associates, 1988.

[2] R. Gallery and T.I.P. Trew. An architecture for face classification. *IEE Colloquium on 'Machine Storage and Recognition of Faces', Digest No: 1992/017*, 2:1–5, 1992.

[3] T.W. Parsons. *Voice and speech processing.* McGraw-Hill, 1986.

[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] F. Samaria. *Face Identification using Hidden Markov Models.* 1st Year Report, Cambridge University Engineering Department, 1992.

[6] J. Shepherd, G. Davies, and H. Ellis. Studies of cue saliency. In G. Davies, H. Ellis, and J. Shepherd, editors, *Perceiving and remembering faces*, pages 105–131. Academic Press, 1981.

[7] C.W. Therrien. *Decision estimation and classification.* Wiley, 1989.

[8] S.J. Young. *HTK: Hidden Markov Model Toolkit V1.3,Reference Manual.* Cambridge University Engineering Department, 1992.