Buletinul Stiintific al Universitatii "Politehnica" din Timisoara, ROMANIA
Seria AUTOMATICA si CALCULATOARE
PERIODICA POLITECHNICA, Transactions on AUTOMATIC CONTROL and COMPUTER SCIENCE
Vol.47 (61), 2002, ISSN 1224-600X

Identity Confirmation System Based on CTAG and HMM

Robert Stefan Győrödi*, Stefan Holban**, Tom Coffey*** and Cornelia Aurora Győrödi****

* Department of Electronic and Computer Engineering, University of Limerick, National Technological Park, Limerick, Ireland
Phone: +353 (0) 61 212-095, Fax: +353 (0) 61 338-176, E-Mail: Robert.Gyorodi@ul.ie, WWW: http://www.ece.ul.ie/robert

** Computer & Software Engineering Department, Politehnica University of Timisoara, Bd. Vasile Parvan 2, 1900, Timisoara, Romania
Phone: +40 (0) 56 204-333 ext 713, Fax: +40 (0) 56 204-333 ext 713, E-Mail: stefan@cs.utt.ro, WWW: http://www.cs.utt.ro/~stefan

*** Department of Electronic and Computer Engineering, University of Limerick, National Technological Park, Limerick, Ireland
Phone: +353 (0) 61 212-268, Fax: +353 (0) 61 338-176, E-Mail: Tom.Coffey@ul.ie, WWW: http://www.ece.ul.ie/

**** Department of Computer Science, Faculty of Electrotehnics and Informatics, University of Oradea, Str. Armatei Romane 5, 3700, Oradea, Romania
Phone: +40 (0) 59 432-830, E-Mail: cgyorodi@univ.uoradea.ro

Abstract *– **Face recognition from still images and video sequences is emerging as an active research area with numerous commercial and law enforcement applications. Hidden Markov Models (HMM) have been successfully used for speech and action recognition where the data that is to be modeled is one-dimensional. Although attempts to use these one-dimensional HMMs for face recognition have been moderately successful, images are two-dimensional (2-D) and consequently a two dimensional model should perform better. In recent researches there were some relatively successful attempts for using embedded HMMs for this purpose. In this paper we present an approach for face recognition using an embedded HMM and compare this new approach to other HMM-based methods.***

Keywords: ***Face recognition, embedded HMM, CTAG, Identity Confirmation.***

I. INTRODUCTION

In the light of recent events, face recognition from still images and video sequences has become a high priority research area with numerous law enforcement and commercial applications. Face recognition systems can be used for non-intrusive person authentication, for example in accessing a workstation, an ATM or even restricted areas, and also could be used to recognize people in specific areas (like airports). A robust face recognition system must operate under a variety of conditions, such as varying illuminations and backgrounds, and it must be able to handle non-frontal facial images of males and females of different ages and races.

As noted in [2], the most significant facial features occur in a natural order, from top to bottom, even if the image undergo small rotations in the image plane, and/or rotations in the plane perpendicular to the image plane. Therefore, the image of a face may be modeled using a one-dimensional HMM by assigning each of these regions to a state as illustrated in Fig. 1.

In this model, the states themselves are not directly observable. What is remarkable are observation vectors that are statistically dependent upon the state of the HMM. These vectors are obtained from L rows that are extracted sequentially from the top of the image to the bottom. Since the length of each row is fixed, and the height of a face image is proportional to its width, this HMM is restricted to fixed-size face images. Although used to model two-dimensional data, this one-dimensional HMM achieved recognition rates of 85% [2][3].

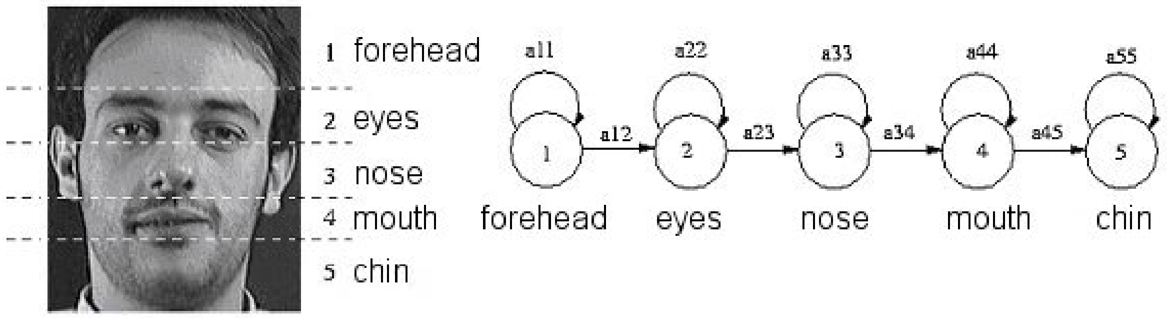

Fig. 1

In this paper, we describe a new approach to face recognition using an embedded HMM as introduced by Kuo and Agazzi [1]. Unlike previous HMM approaches to face recognition, which use pixel intensities [6] or two-dimensional Discrete Cosine Transform (2D-DCT) [7] to form the observation vectors, our embedded HMM uses observation vectors that are composed of Color Texture Adjacent Graf coefficients (CTAG) [9]. Compared with other methods, our proposed system offers a more flexible framework for face recognition, and can be used more efficiently in scale invariant systems.

II. THE EMBEDDED HMM APPROACH

A one-dimensional HMM is a Markov chain with a finite number of unobservable states [11]. Although the Markov states are not directly observable, each state has a probability distribution associated with the set of possible observations. Thus, when the HMM is in state *i*, the output (observation) is determined according to a given conditional probability density function, often Gaussian or a Gaussian mixture. What is necessary to statistically characterize an HMM is:

- a state transition probability matrix
- an initial state probability distribution
- a set of probability density functions associated with the observations for each state

A one-dimensional HMM may be generalized, to give it the appearance of a two-dimensional structure, by allowing each state in a one-dimensional HMM to be a HMM. In this way, the HMM consists of a set of *super states*, along with a set of *embedded* states. The super states may then be used to model two-dimensional data along one direction, with the embedded HMM modeling the data along the other direction. This model differs from a true two-dimensional HMM since transitions between the states in different super states are not allowed. Therefore this is referred to as an embedded HMM [7]. The elements of an embedded 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, $\Pi_0 = \{\pi_{0,i}\}$ where $\pi_{0,i}$ are the probabilities of being in super state *i* at time zero.
- The super state transition probability matrix,
  $$A_0 = \{a_{0,ij}\}$$
  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
  - The number of embedded states in the k^{th} super state, $N_1^{(k)}$, and the set of embedded states, $S_1^{(k)} = \{ S_{1,i}^{(k)}\}$.
  - 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.
  - The state transition probability matrix,
    $$A_1^{(k)} = \{ a_{1,jk}^{(k)} \}$$
    That specifies the probability of transitioning from state *k* to state *j*.
- Finally, there is the state probability matrix,
  $$B^{(k)} = \{ b_i^{(k)}(O_{t0,t1})\}$$
  for the set of observations where $O_{t0,t1}$ represent the observation vector at row t_0 and column t_1. In a *continuous density* HMM, the states are characterized by continuous observation density functions. The probability density function that is typically used is a finite mixture of the form
  $$b_i^{(k)}(O_{t_0,t_1}) = \sum_{m=1}^{M} c_{im}^{(k)} N(O_{t_0,t_1}, \mu_{im}^{(k)}, U_{im}^{(k)})$$
  where $1 \le i \le N_1^{(k)}$, $c_{im}^{(k)}$ is the mixture coefficient for the m^{th} mixture in state *i* of super state *k*. $N(O_{t0,t1}, \mu_{im}^{(k)}, U_{im}^{(k)})$ is a Gaussian pdf with mean vector $\mu_{im}^{(k)}$ and covariance matrix $U_{im}^{(k)}$.

Let $\Lambda^{(k)} = \{ \Pi_1^{(k)}, A_0, \Lambda \}$ be the set of parameters that define the k^{th} super state. Using a shorthand notation, an embedded HMM is defined as the triplet

$$\lambda = (\Pi_0, A_0, \Lambda)$$

where $\Lambda = \{\Lambda^{(1)}, \Lambda^{(2)}, ..., \Lambda^{(N_0)}\}$ [7].

The state structure of the face model and the non-zero transition probabilities of the embedded HMM are shown in Fig. 2.

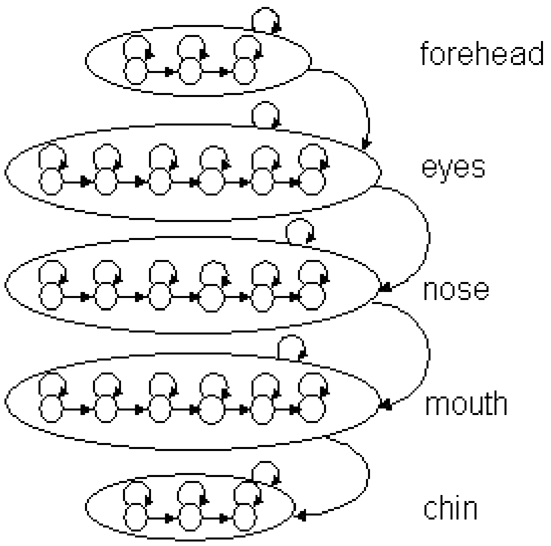

Fig. 2. Embedded HMM state structure for face recognition.

Each state in the overall top-to-bottom HMM is assigned to a left-to-right HMM. This model is appropriate for face images since it exploits an important facial characteristic, namely that frontal faces preserve the same structure of "super states" from top to bottom, and also the same left-to-right structure of "states" inside each of these "super states".

III. EXPERIMENTS AND RESULTS

A. Observation vector extraction

For the usage of this model, the observation sequence is generated using the technique shown in Fig. 3, where a *P*x*L* window scans the image left to right, and top to bottom. The overlap between adjacent windows is *M* lines in the vertical direction and *Q* columns in the horizontal direction.

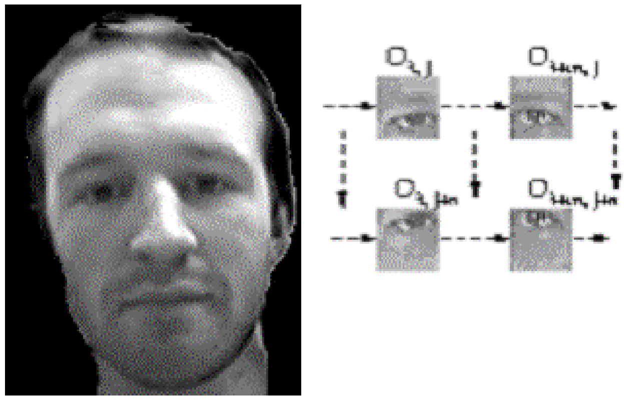

Fig. 3. Face blocks extraction.

B. The proposed feature extraction method

In [3] Samaria constructed the observation vectors from all the pixel values from each block, and therefore the dimension of the observation vectors was *P*x*L*. But the pixel values do not represent robust features (e.g. they are sensitive to image noise, change in illumination etc.)

In [7] Nefian uses the 2D-DCT coefficients of each immage block as observation vectors. The compression and decorrelation properties of the 2D-DCT for natural images makes it suitable for this usage.

In this paper we propose the usage of CTAG [9] coefficients as observation vectors. These are also suitable for observation vectors because they actually describe the texture of the considered region. In Fig. 4 we have a sample of CTAG coefficients, computed for a band from a face image. Based on the experiments conducted we will derive the most significant parameters that influence the usage of CTAG coefficients for observation vectors in the process of face recognition.

C. The Experiment

For each individual in the database, an embedded HMM was computed using a set of images representing different instances of the same face. The entire training process is described in detail in [8].

After extracting the observation vectors corresponding to the test images, the probability of the observation sequence given an embedded HMM face model is computed via a doubly embedded Viterbi recognizer. The model with the highest likelihood is selected and this model reveals the identity of the unknown face.

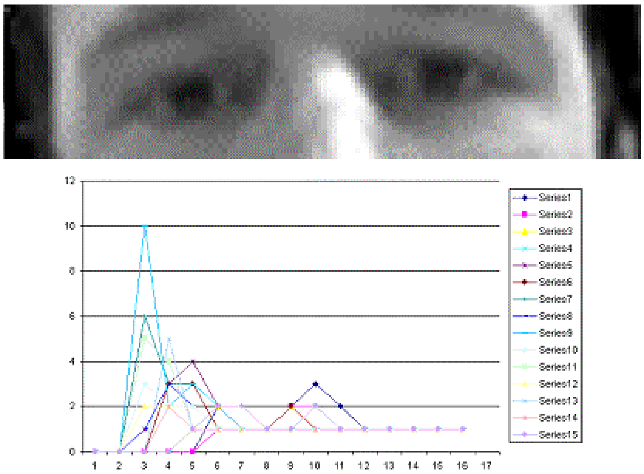

Fig. 4. CTAG coefficients for a face image band.

The face recognition system was tested on the face database from Olivetti Research Ltd. (400 images of 40 people having a resolution of 92x112 pixels). Half of the images were used in training and the other half for testing. The database contains face images of people of different ages, males and females, having different face expressions, hair cuts and some of them wearing spectacles.

D. Results comparison

The recognition rate using eigenfaces, that depends on the number of eigenfaces used, is between 73%, using less than 5 eigenfaces, to almost 90% when using 200 eigenfaces.

The recognition rate of the 2D-HMM method described by Samaria [3], depends on the structure and the actual instances used, and is between 90% and 95%, but it has the drawback of using large observation vectors, that leads to longer recognition times.

The recognition ratio of the system presented by Nefian [7] grows up to 98% using observation vectors based on DCT or KLT, and the recognition time drops considerably in comparison with that proposed by Samaria.
The efficiency of this system is mainly due to a more efficient observation vector and also to the usage of a more efficient HMM structure. A shorter observation vector evidently leads to the reduction of the complexity of the training and also of the recognition. On the other hand, the efficiency of the embedded HMM is proved in [8]. Additionally, the embedded HMM allows a parallel implementation of the decoding algorithm (the doubly embedded Viterbi algorithm) that further reduces the recognition time.

The proposed method in this paper, namely to calculate the observation vectors based on the CTAG coefficients, leads us to somewhat lower recognition rates, compared with those reported by Samaria [3] and Nefian [8], but still a reasonable value, somewhere around 83%, depending on the chosen parameters.
The recognition time, on the other hand presents a slight improvement, having a relative recognition time of 88%, as shown in Fig. 5. This fact can be explained by a reazonable observation vector length and a lower complexity of computing them.

	Observation Vector Length	Recognition Time (relative)	Recognition Rate
Samaria (pixels)	400 (20 x 20)	575%	93%
Nefian (DCT)	9 (3 x 3)	100%	98%
CTAG	28 (14 x 2)	175%	83%
CTAG	14 (14 x 1)	88%	83%

Fig. 5. Comparison of methods based on obs. vect. length / rec. time / rec. rate.

To be able to compare the proposed method with those proposed by Samaria and Nefian, we implemented them.

E. Proposed model parameters

For the proposed method we identified some parameters that influence the recognition ratio:

1. The dimension of the image block – a too small dimension cannot capture sufficient information about the structure of the texture, to be relevant to our feature extraction model, and one that is too big leads to an insufficient number of observation vectors for the embedded states.
2. The color range that is used in extracting the CTAG coefficients – one improvement in this area could be an auto-adjustable range, which is dependent on the region on which the method is applied (e.g. for skin we could have a different range).
3. The granularity of the color range that is used in extracting the CTAG coefficients – we arrived at the conclusion that there are no significant improvements if we increase the granularity, so we tried to lower as much as possible the granularity – this to reduce the processing time.
4. The overlapping on x and y directions – in this case we concluded that slightly lowering the overlapping, this also leading to reduced

processing time, has no significant negative impact, so we increased the window shifting from 2x2 in the case of DCT (Nefian [8]) to 5x5 for CTAG.

The experimental results, using an embedded HMM and CTAG, applied on the ORL database, and the variation of the recognition ratio on the parameters of the model, are shown in Fig. 6. The results shown in this figure support all the considerations presented earlier when we discussed the identified parameters.

IV. CONCLUSION

Based on the experiments that we done, compared with methods based on templates or even standard HMM based face detection and recognition, the embedded HMM approach is more flexible in relation to scale variations and natural deformations, both on vertical and horizontal directions.

Also the performances of an embedded HMM based approach are equal or even superior to other methods having, in the same time, a lower complexity too, and this for a large variety of feature extraction techniques used for obtaining the observation vectors (CTAG, DCT, KLT, pixel values).

Through our experiments we confirm the results obtained by Samaria [3] and mainly those obtained by Nefian [8], showing also that the embedded HMM approach can be extended by using different feature extraction techniques (like the CTAG).

Future work will be directed towards building an identification system that will use non-intrusive techniques, such as those described in this paper for person identification purposes. At the beginning this system will be used only for illustration purposes, and work will be carried on improving the results in non-ideal situations.

Block dimension	Color range *	Levels **	Shifting	Correction factor ***	Recognition ratio
25 × 25	(16,0,16)	(2,2,2)	5 × 5	10	63.3%
25 × 25	(16,0,16)	(2,2,2)	5 × 5	10	73.3%
25 × 25	(32,0,8)	(3,2,2)	5 × 5	10	76.7%
25 × 25	(32,0,8)	(3,2,2)	2 × 2	10	76.7%
25 × 25	(16,0,16)	(2,2,2)	2 × 2	10	66.7%
30 × 30	(16,0,16)	(2,2,2)	2 × 2	10	73.3%
30 × 30	(14,16,16)	(2,2,2)	5 × 5	10	76.7%
30 × 40	(14,16,16)	(2,2,2)	5 × 5	10	63.3%
30 × 30	(14,16,16)	(2,2,2)	5 × 5	10	76.7%
30 × 30	(14,16,16)	(2,2,2)	5 × 5	40	83.3%
30 × 30	(14,16,16)	(2,2,2)	5 × 5	100	80.0%
30 × 30	(10,128,8)	(2,2,2)	5 × 5	40	36.7%
30 × 30	(12,16,16)	(3,2,2)	5 × 5	40	70.0%
30 × 30	(14,16,16)	(1,2,2)	5 × 5	40	83.3%
30 × 30	(14,16,16)	(1,2,2)	2 × 2	40	83.3%
35 × 35	(14,16,16)	(1,2,2)	5 × 5	40	80.0%

Fig. 6. Experimental results using eHMM + CTAG.

* (Number of values, Initial value, Step)

** (Number of levels, Initial value, Step)

*** Corection factor of the values for a better discrimination between vectors (the elements of the observation vector are multiplied with this value)

REFERENCES

[1] O. Agazzi, S. Kuo, E. Levin şi R. Pieraccini – ”Connected and degraded text recognition using planar HMM”, International Conference on Acoustics, Speech and Signal Processing ’93, vol 5, pp. 113-116, 1993.

[2] F. Samaria – ”Face segmentation for identification using hidden Markov models”, British Machine Vision Conference, 1993.

[3] F. Samaria şi F. Fallside – ”Face identification and feature extraction using hidden Markov models”, Image Processing: Theory and Applications, 1993.

[4] F. Samaria şi F. Fallside – ”Automated face identification using hidden Markov models”, Proceedings of the International Conference on Advanced Mechatronics, pp. 1-9, 1993.

[5] F. Samaria şi A. Harter – ”Parametrisation of stohastic model for human face identification”, Proceedings of the Second IEEE Workshop on Application of Computer Vision, pp. 138-142, 1994.

[6] F. Samaria – ”Face Recognition Using Hidden Markov Models”, PhD thesis, University of Cambridge, 1994.

[7] A.V. Nefian, M.H. Hayes – ”Face Recognition Using an Embedded HMM”, Proceedings of the IEEE Conference on Audio and Video-based Biometric Person Authentication, pp. 19-24, Martie 1999.

[8] A.V. Nefian – ”A Hidden Markov Model-Based Approach for Faces Detection and Recognition”, PhD thesis, Georgia Institute of Technology , 1999.

[9] Robert Győrödi, Dr. Mark H. Fisher, Juliana Camapum - “Multi-scale color texture description” - 12th International Conference on Systems Engineering, Vol1, Coventry University, 9-11 September 1997, UK, pp. 284-289.

[10] Robert Győrödi – “Contributions to Feature Selection in the Image Recognition and Interpretation Process”, PhD thesis, Politehnica University of Timisoara, Romania, 2001.

[11] L. Rabiner – “A tutorial on Hidden Markov Models and selected applications in speech recognition”, Proceedings of IEEE, vol. 77, pp. 257-286, February 1989.