Video-Based Face Authentication Using Appearance Models and HMMs

Ke-Zhao Chen[1], Yao-Jen Chang[2], and Chia-Wen Lin[1]
[1]Department of Computer Science & Information Engineering
National Chung Cheng University
Chiayi 621, Taiwan
cwlin@cs.ccu.edu.tw
[2]Advanced Technology Center, Computer & Communications Research Laboratory
Industrial Technology Research Institute
Hsinchu 310, Taiwan
KevinChang@itri.org.tw

***Abstract*—In this paper, we propose a novel face authentication scheme using the Active Appearance Model (AAM) and the Hidden Markov Model (HMM). The proposed face authentication system can be divided into two parts. First, the AAM is used to extract the low-dimensional feature vectors including combined texture and shape information of individual face images. The extracted feature vectors are further classified into several clusters using vector quantization. The clustered feature vectors are then characterized using HMMs to make full use of the temporal information across the face images. After all parameters in the HMMs are calculated, we can dynamically determine the thresholds for face authentication. An iterative algorithm is also proposed to automatically determine a suitable number of HMM states and a suitable number of observation classes to achieve good authentication accuracy. The experimental results show the efficacy of the proposed method.**

I. INTRODUCTION

The rapid advance of computing power and multimedia technology has already enabled many images processing applications. Especially, using biometric information for security purposes is an increasing trend. Currently biometric features, including face, speech, iris, fingerprints, and palm prints, etc., have been widely utilized for person identification or authentication. However, when we use fingerprints and palm prints to identify a person, the direct contact of the body will easily cause a source of infection of the epidemic disease. On the other hand, when the instrument infected by sweat, it will also be difficult to preserve and decrease the recognition rate. Although the iris recognition achieves highest rate in the domain of biometrics-based recognition, it will make users uncomfortable since it launches a line of light to scan the eyeballs. Thus, it is only applied to the most secure usage such as military applications. In addition, the use of speech recognition is easier than other recognition methods, but it is often affected by the external noise. Among all biometrics-based recognition methods, face recognition is a natural and convenient method to use. Not only the users will not have any negative feeling in the recognition phase, but the recognition rate is also promising. Thus, it has long been receiving broad attentions in biometrics-based identification and authentication research fields [1][2].

The face recognition methods can be roughly divided into three categories: the feature-based methods [1], the template-based methods [2], and the model-based methods [3]. The feature-based methods are the earliest and intuitive methods. In the training phase, several salient facial features are detected by using a face detector, and the relation of distances between the feature points is taken as the feature vector and stored in the database. In the verification phase, the minimum distance approach is utilized to measure the similarity between the feature vectors of the test image and the trained image. The template-based methods use the entire face template to recognize faces and usually outperform the feature-based approaches. The model-based methods such as Active Shape Model (ASM) based and AAM-based are also popular in recent years. Not only a compact representation is provided for faces by decoupling shapes and textures, effective tracking mechanisms are also presented for capturing the dynamics of deformable objects.

Traditionally, most of previous face identification methods only focus on the single face image. But it is vulnerable to attacks with fake or photoed face pictures. Thus, one way to resolving this problem is to use the face video instead of a single image. Video-based face authentication cannot only be taken as a live verification to prevent fraud, but potentially it can improve the discriminability by making use of temporal information across the video sequence. Several video-based authentication approaches have been proposed recently. The method presented in [4] proposes to use the SVM classifier for video-based face recognition, but it is just a simple voting mechanism similar to image-based recognition with multiple images. In [5] a method using the trained identity surfaces with the multi-view dynamic faces for recognition is proposed. The method presented in [6] instead uses a trained probabilistic appearance manifold with a set of rotation faces. Although these methods all use the temporal information across a video sequence, the information of face rotation is not a good password for face authentication. This password is not secure since the action of face rotation can be easily seen by others and the degree of freedom is small. Tang and Li proposed to segment video according to speech alignment [7]. Face features are extracted separately from each video segment for face authentication. This approach uses the temporal information across a video sequence, but dynamics is still not fully explored. Liu and Chen used adaptive HMMs to learn the

dynamics of face rotation sequences [8]. This paper shows that the HMM can be successfully applied to model temporal information, but the behavior of long-term unintentional face rotation may not be suitable for most face authentication applications. Based on aforementioned observations, we propose a new video-based face authentication by using appearance models for feature representation and the HMM for exploring the temporal feature models.

The rest of this paper is organized as follows. In Sec. 2, we briefly describe AAM and HMM that form the foundation of this work. The proposed video-based face authentication scheme is elaborated in Sec. 3. Preliminary experimental results are presented in Sec. 4. Finally, Sec. 5 concludes the paper.

II. REVIEW OF AAM AND HMM

The concept of Statistical Appearance Model (SAM) was proposed in [9]. It can model both the shape and texture of an image of an object. The models are generated by combining a model of shape variation with a model of the texture variation in a normalized frame, and it is usually used to synthesize a complete image of an object or structure. The shape and the shape-free texture of a face can be extracted using the statistical appearance models, and the shape-free texture means that the texture of a face is not dependent on the shape of a face. Fig. 1 shows that each training example can be split into a set of landmark points and a shape-free texture patch. Alternatively, some methods like Eigenfaces or Fisherfaces are only used to extract the lower dimensional texture to interpret a face, but discarding the shape information which is also very useful to interpret a face.

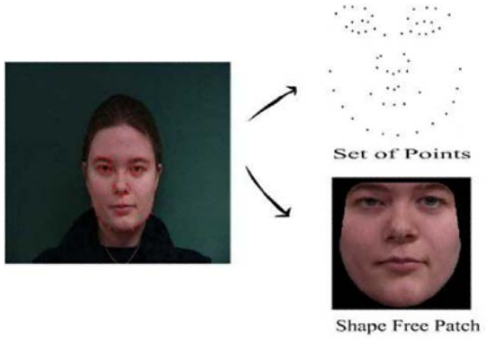


Fig. 1. A set of points and the shape-free patch of a face [9].

AAM [10][11] is an extension to the ASM. Instead of manually selecting feature points in SAM, it is used to extract the representative features of the object in an image automatically. The appearance model is built based on a set of labeled images, where the landmark points are marked on each face. After labeling all the sample images, the Procrustes analysis is utilized to align each face according to a mean shape. Principal Component Analysis (PCA) is then applied to effectively reduce the dimensionality for both shapes and shape-free textures. Thus, the shape $\mathbf{x}$ and texture $\mathbf{g}$ of a face image can be represented by a compact feature vector $\mathbf{c}$ as

$$\begin{aligned} \mathbf{x} &= \bar{\mathbf{x}} + \mathbf{Q}_s \mathbf{c} \\ \mathbf{g} &= \bar{\mathbf{g}} + \mathbf{Q}_g \mathbf{c} \end{aligned} \tag{1}$$

where $\bar{\mathbf{x}}$ is the mean shape, $\bar{\mathbf{g}}$ is the mean of shape-free texture, and $\mathbf{Q}_s, \mathbf{Q}_g$ are matrices describing the modes of variation derived from the training set [11].

The Hidden Markov Model (HMM) is a statistical model used to characterize a signal as a parametric random process. In past ten years, the HMM approach has been successfully and widely used for many applications because the models are very rich in mathematical structure. In the speech recognition domain, Rabiner [13] proposed the approaches for speech recognition by using HMM. The method proposed in [14] uses the embedded HMM for facial expression. In the face recognition domain, Liu proposed an approach for video-based face recognition by using adaptive HMM [8], and each HMM is adapted with the test video sequence.

An HMM can be viewed as an unobservable Markov chain with a finite number of states. An HMM can be described by a transition probability matrix $\mathbf{A}$, an initial state probability distribution $\boldsymbol{\pi}$, and a set of probability density functions for observations $\mathbf{B}$. Fig. 2 shows the diagram of HMM with three states.

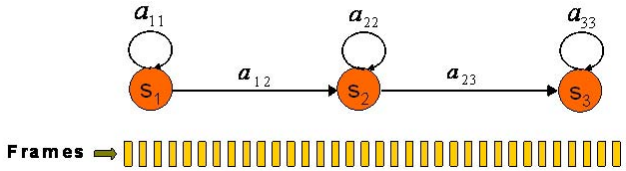


Fig. 2. A 3-state left-right HMM.

III. PRPOPOSED VIDO-BASED FACE AUTHENTIFICATION

A. System Overview

Fig. 3 shows the block diagram of our proposed face authentication system. The proposed scheme can be divided into two steps. First, the features are extracted by using the appearance models. Second, the AAM parameters are classified by HMM for authentication. During the authentication process, the user is asked to show his/her front face to the camera and say a secrete password. A face video with mouth motions corresponding to the secret password is captured for authentication. At the first stage, a skin-color-based face detector is used to locate the face region which is used to determine the initial AAM shape model. Based on the initial model, an iterative AAM mode refinement algorithm is performed to locate accurate feature point locations as well as to extract the low dimensional features of each face image. The model parameters of all face images in a video sequence are then clustered using vector quantization to obtain a reduced number of observation vectors. These observation vectors are fed into a set of HMM classifiers to decide whether the incoming face sequence matches the features of an authorized user.

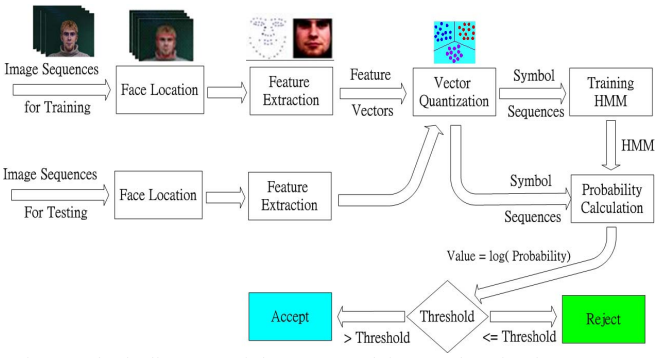


Fig. 3. Block diagram of the proposed face authentication system.

In this work, we propose an adaptive scheme to determine the thresholds in authenticating a face. We also propose an iterative algorithm with a set of training sequences to determine a suitable

hidden state number in HMM and a suitable class number of observations.

B. Face authetication Using HMM

As shown in Fig. 3, the sequences in the database are divided into two sets: one set for training and the other for testing. In the training stage, we first extract the features of each training face image and use these features to construct an appearance model. The appearance parameter **c** in (1) can then be easily obtained by using this model. After all the appearance parameters are computed, we use a vector quantizer to separate these feature vectors into N clusters. Thus, we can obtain all the observations **O**, and each training face image sequence can be represented as a symbol sequence. Finally, we use these symbol sequences for training the HMM.

When we train the HMMs, a transition probability matrix **A**, an initial state probability distribution **B**, and a set of probability density functions $\boldsymbol{\pi}$ need to be initialized. Here we set the initial value of $\boldsymbol{\pi}$ = [1,0,0,…,0], because the left-to-right HMM is suitable for our application. We then use the training symbol sequences and the Expectation Maximization (EM) algorithm to calculate the final parameter vector (**A**, **B**, $\boldsymbol{\pi}$) of each trained HMM. After the face authentication system is built, we need to determine the parameters of HMM. Finally, we use the test sequences with the trained HMM parameters to test the system and check if the system is stable for face authentication.

C. Determining the HMM model parameters

While testing the face authentication system, we need to determine a suitable threshold *TH*, the hidden state number *S* in HMM, and the class number of observations *C*. The threshold *TH* for face authentication is first considered. Taking Fig. 4 as an example in which four persons are involved in the database, "*A_A*" is a sequence which represents person *A* says *A*'s password; "HMM *A_A*" represents a pre-trained HMM obtained from training by a number of "*A_A*" sequences; p_1 is the probability that is computed from using the "*A_A*" sequences to test the trained "HMM *A_A*" model. In order to find a suitable threshold for "HMM *A_A*," the probabilities, p_1, p_2, p_3, and p_4 are calculated first. The "HMM *A_A*" model has to accept "*A_A*" and reject "*A_B*," "*A_C*," and "*A_D*." Thus, the threshold for "HMM *A_A*" must be higher than the value $\log(p_1)$ and lower than the values, $\log(p_2)$, $\log(p_3)$, and $\log(p_4)$. We can roughly separate these log-probabilities into two classes. The log-probability (i.e., p_1) in the first class must be accepted by "HMM *A_A*" and the other class of log-probability values (e.g., p_2, p_3 and p_4) must be rejected for "HMM *A_A*." We can then compute the mean values, μ_1 and μ_2, of the corresponding classes, respectively. The difference between μ_1 and μ_2 is also calculated. In our method, 25 equally spaced thresholds are determined as shown in Fig. 5. Finally, the following iterative algorithm with these 25 thresholds is performed to find a suitable state number of HMM and a suitable class number of observations.

Step 1. Set an initial state number.

Step 2. Given the fixed state number, change the class number and the threshold in order, and then use the training sequences to test each HMM. We can find a best class number according to the ROC curve.

Step 3. Use the result obtained in **Step 2** as the fixed class number. Change the state number and the threshold in order, and then use the training sequences to test each HMM. We can find a best state number according to the ROC curve and use it to update the state number.

Step 4. Repeat **Step 2** and **Step 3** until the state number of HMM and the class number of observations converge.

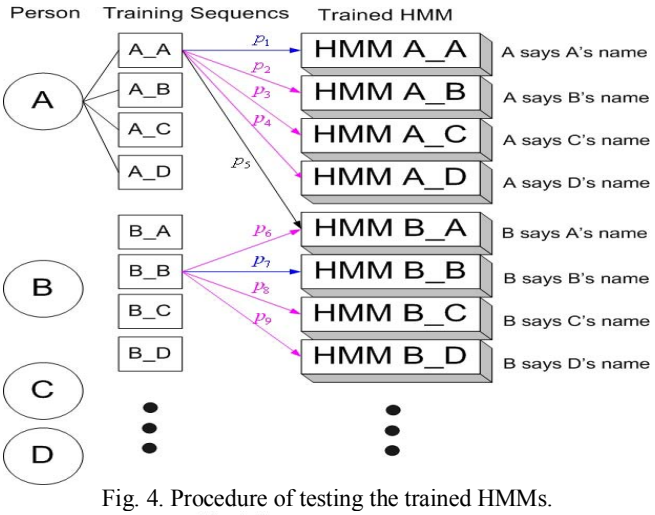


Fig. 4. Procedure of testing the trained HMMs.

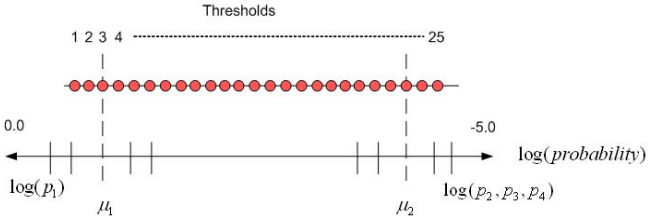


Fig. 5. Determining the 25 thresholds for HMM.

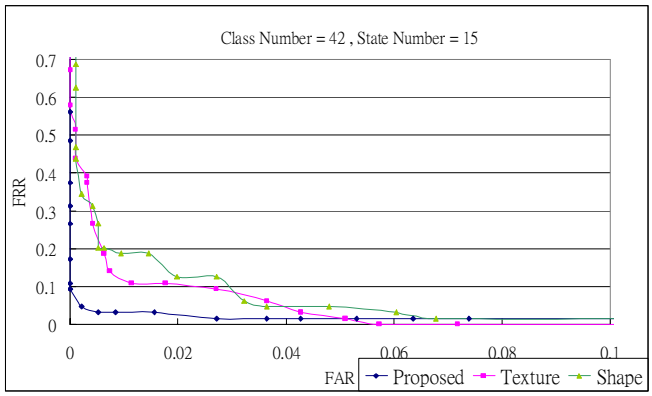
IV. Experimental Results

We constructed a database which contains 128 face sequences (64 training sequences and 64 test sequences) with about 4500 images. Totally 16 HMMs were trained for face authentication. Fig 6 illustrates a few sample face images of our database which involves four different persons. In the database, each person says a secrete sentence as the password. In our experiments, we simply used each user's name as his own password. For example, as shown in Fig. 4, the face model, HMM *A_A*, only accepts the case that person *A* says *A*'s password. On the other word, both person *A* says others' passwords and other persons say *A*'s password will get rejected by the "HMM *A_A*" model.

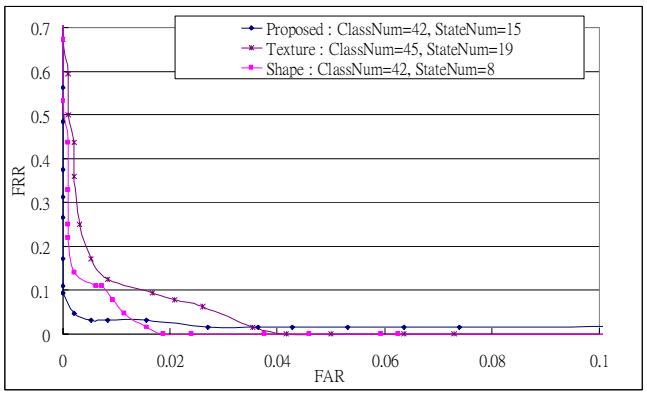
We use the proposed iterative algorithm to find a suitable state number of HMM and a suitable class number of observations. Our experiments show that when the combination of the state and class numbers (*S*, *C*) = (15,42), both FAR (False Acceptance Rate) and FRR (False Rejection Rate) have best performance. We compare the ROC curves (FAR on the horizontal axis and FRR on the vertical axis) of three methods: the proposed method, the method which uses the AAM shape model and HMMs, and the method which uses the AAM texture model and HMMs (similar to PCA method). Fig. 7(a) compares the ROC curves of our proposed method and the other methods with the same state and class numbers which are optimum to the proposed method. Fig. 7(b) shows the performance comparison of the three methods in which

the optimum state and class numbers are selected for each corresponding method. The experiments show that the proposed method achieves significantly better FRR and FAR combinations most of the time.

Fig. 6. Sample face images in our database.



(a)



(b)

Fig. 7. Performance comparison of the proposed method, AAM shape + HMM, and AAM texture + HMM: (a) (state number, class number) = (15,42) for all the three methods; (b) each method uses its optimum class number and state number.

V. Conclusion

In this paper, we proposed a novel video-based face authentication scheme using AAM and HMMs, in which AAM is used to extract the low dimensional features of a face image. After extracting model parameters of each face image in a face video sequence, a reduced set of observations of each sequence are obtained using vector quantization to cluster all these feature vectors. Using HMMs to characterize the temporal dynamics of these observations, we can extract useful features for face authentication. We proposed a scheme to adaptively determine the thresholds used in the system. We have also proposed an iterative algorithm to determine a suitable hidden state number in HMM and a suitable class number of observations using test sequences. The experimental results show that the proposed method achieves very low FAR and FRR.

References

[1] A. J. Goldstein, L. D. Harmon, and A. B. Lesk, "Identification of human faces," *Proc. IEEE*, vol. 59, no. 5, pp. 748-760, May 1971.

[2] R. Chellappa, C. L. Wilson and S. Sirohey, "Human and machine recognition of faces: A survey," *Proc. IEEE*, vol. 83, no. 5, pp. 705-741, May 1995.

[3] A. Nefian, *A Hidden Markov Model-Based Approach for Face Detection and Recognition*, *PhD thesis*, Georgia Institute of Technology, Atlanta, GA, August 1999.

[4] Z. Li, H. Ai, and G. Xu, "Training support vector machines for video based face recognition," In *Proc. IEEE Int. Conf. Image and Graphics*, 2002.

[5] Y. Li, S. Gong, and H. Liddell, "Video-based online face recognition using identity surfaces," in *Proc. IEEE Int. Conf. Recognition, Analysis, and Tracking of Faces and Gestures in Real-Time Systems*, pp. 40-46, July 2001.

[6] K.-C. Lee, J. Ho, M.-H. Yang, and D. Kriegman, "Video-based face recognition using probabilistic appearance manifolds," in *Proc. IEEE Int. Conf. Computer Vision and Pattern Recognition*, vol. 1, pp. 313-320, June 2003.

[7] X. Tang and Z. Li, "Frame synchronization and multi-level subspace analysis for video based face recognition," in *Proc. IEEE Int. Conf. Computer Vision and Pattern Recognition*, vol. 2, pp. 902-907, July 2004.

[8] X. Liu and T. Chen, "Video-based face recognition using adaptive hidden Markov models," in *Proc. IEEE Int. Conf. Computer Vision and Pattern Recognition*, vol. 1, pp. 340-345, June 2003.

[9] T. F. Cootes and C. J. Taylor, *Statistical Models of Appearance for Computer Vision*, *PhD thesis*, Manchester, U.K., March 2004.

[10] T. F. Cootes, G. J. Edwards, and C. J. Taylor, "A comparative evaluation of active appearance model algorithms," in *Proc. British Machine Vision Conference*, 1998.

[11] T. F. Cootes, A. Hill, C. J. Taylor, and J. Haslam, "The use of active shape models for locating structures in medical images," *Image & Vision Computing*, July 1994.

[12] H. Kang, T. F. Cootes, and C. J. Taylor, "A comparison of face verification algorithms using appearance models," in *Proc. BMVC*, vol. 2, pp. 477-486, 2002.

[13] L. R. Rabiner, "A tutorial on hidden Markov models and selected applications in speech recognition," *Proc. IEEE*, vol. 77, no. 2, pp. 257-286, Feb. 1989.

[14] X. Zhou, X. Huang, B. Xu, and Y. Wang, "Real-time facial expression recognition based on boosted embedded hidden Markov model," in *Proc. IEEE Int. Conf. Image and Graphics*, pp. 290-293, Dec. 2004.