World Academy of Science, Engineering and Technology
International Journal of Computer and Information Engineering
Vol:7, No:10, 2013

Video-Based Face Recognition Based On State-Space Model

Cheng-Chieh Chiang, Yi-Chia Chan, Greg C. Lee

***Abstract*—**This paper proposes a video-based framework for face recognition to identify which faces appear in a video sequence. Our basic idea is like a tracking task - to track a selection of person candidates over time according to the observing visual features of face images in video frames. Hence, we employ the state-space model to formulate video-based face recognition by dividing this problem into two parts: the likelihood and the transition measures. The likelihood measure is to recognize whose face is currently being observed in video frames, for which two-dimensional linear discriminant analysis is employed. The transition measure estimates the probability of changing from an incorrect recognition at the previous stage to the correct person at the current stage. Moreover, extra nodes associated with head nodes are incorporated into our proposed state-space model. The experimental results are also provided to demonstrate the robustness and efficiency of our proposed approach.

***Keywords*—**2DLDA, face recognition, state-space model, likelihood measure, transition measure.

I. Introduction

VIDEO data are currently widely obtained in many kinds of application, such as from a camera in a handheld device for capturing everyday lives, from a web camera in a laptop for sending video messages, and from surveillance cameras in city streets for ensuring the security of the general public. When we design a video-based application, it is important to understand who appears in a video. Thus, face recognition is often a key technology in many video-based applications, which aims to recognize which persons are observed in a video sequence.

This paper deals with the problem of face recognition in a video sequence, with the assumption that face areas have been localized. In principle, video-based face recognition can be regarded as a fusion of recognition results in a set of consecutive still images. However, video frames actually contain more information. For example, a single face may keep moving with different poses in a video so that appearance features of these head poses may be helpful for the face recognition. Incorporating all of the relationships for different head poses appearing in a video should have the potential to overcome difficulties encountered when attempting to identify persons appeared in a video.

Assume that there are *K* persons in the system and these persons may appear with different head poses. Our basic idea is analogous to a tracking task - to track a selection of the *K* candidates over time according to the observations of visual features in video frames. This prompted us to employ the state-space model [1], [2], which is well known and widely used for visual tracking, to construct a probabilistic graphical model for video-based face recognition, by dividing this problem into two parts: likelihood and transition measures. The former is like a traditional task of face recognition in a still image that involves making a decision about whose face is currently being observed, while the latter estimates the probability of a change from the recognized state at the previous stage to each of possible states at the current stage. The transition measure makes it possible to change recognition results from an incorrect decision to the correction one. Moreover, our formulation also considers extra nodes associated with head poses in the probabilistic graphical model such that the transition measure also involves extra information of head poses to improve the video-based face recognition in our approach.

The remainder of this paper is organized as the follows. Section II discusses some related works on face recognition. Section III briefly introduces the basic concept of the state-space model. Then, Section IV proposes our formulation for solving the face recognition in video based on a state-space model. The likelihood and the transition measures of our proposed state-space model are presented in Section V and VI, respectively. Section VII presents several experimental results to demonstrate the performance of our proposed approach. Finally, Section VIII draws conclusions and discusses future work.

II. Related Work

The problem of face recognition in a still image has traditionally been approached using subspace learning, such as Eigenface [3], PCA and LDA [4], two-dimensional LDA (2DLDA) [5]. Learning a manifold subspace to efficiently represent face images has also presented a great performance for face recognition. Li et al. employed the classical locally linear embedding to extract the most discriminant features of face images for face recognition [6]. He et al. proposed a Laplacianfaces approach [7] that is based on Locality Preserving Projection [8]. The Laplacianfaces approach can map face images into a subspace that can optimally preserves the local manifold structure. It is also a well-known way to design a proper learning classifier for face recognition such as the Hidden Markov Model (HMM) [9] and the Support Vector Machine (SVM) [10]. Lee et al. designed facial-trait codes to represent and encode face images, which can provide distinctive facial traits for recognition even if partial occlusion

Cheng-Chieh Chiang is with the Department of Information Technology, Takming University of Science and Technology, Taipei, Taiwan (corr. author to provide phone: +886 226585801; e-mail: kevin@ csie.ntnu.edu.tw).

Yi-Chia Chan and Greg C. Lee are with the Department of Computer Science and Information Engineering, National Taiwan Normal University, Taipei, Taiwan.

occurs [11]. Rama et al. designed an aligned face texture map using nine different views and proposed a Partial Principal Component Analysis (P^2CA) method to learn face models based on these aligned face maps [12]. Mohammed et al. employed two approaches of bidirectional two dimensional principal component analysis (B2DPCA) and extreme learning machine (ELM) to deal with the human face recognition [13].

Video-based face recognition can in principle be regarded as a fusion of recognition results from a set of sequential images. A face recognition method using temporal voting to incorporate results of still images has been proposed for image sequences [14]. Visual features extracted from images of a human face in a continuous video sequence could form a manifold in high-dimensional feature space, and hence the problem of face recognition can be converted into a matching problem between the corresponding manifolds [15]-[17]. Many personal photographs have been shared on the social network sites, so Stone et al. argued that social network context may be the key for large-scale face recognition to succeed [18]. Therefore, the resources and structure of such social networks can be used for improving face recognition rates on the images shared. Cinbis et al. designed an unsupervised approach of metric learning to improve results for identification, recognition, and clustering of face tracks automatically extracted from uncontrolled TV video [19]. Haar and Veltkamp treated face images from video frames as 3D models, and converted the recognition problem into the problem of matching and searching 3D models for face images [20].

Head pose understanding may also be useful for improving face recognition in a video-based environment. Zhang and Gao classified techniques of face recognition across head poses into three categories: general algorithms, 2D techniques, and 3D approaches [21]. Our work, which may belong to the category of 2D techniques, constructs a probabilistic graphical model containing likelihood and transition measures to identify which person appears in a video. Kim et al. designed a HMM model for building a face model of face/poses in consecutive video frames [22]. Arandjelovic and Cipolla modeled the motion of the face manifold by combining person-specific face motion appearance manifolds with generic pose-specific illumination manifolds [23]. Their approach can achieve a robust recognition to changes in illumination, pose and the motion pattern of the user.

III. State-Space Model

A state-space model is based on Bayesian network to analyze dynamic systems, which estimate the states of systems changing over time from a sequence of noisy measurements [1], [2]. Here, we only provide a brief summary of how the posterior probability of a state-space model is inferred.

A state-space model in general contains two types of nodes at time t: (i) x_t for the system state and (ii) z_t for the observation measurement, whose probabilistic graphical structure is shown as Fig. 1. To simply express the equations, we use the notations $X_t=\{x_1, ..., x_t\}$ and $Z_t=\{z_1, ..., z_t\}$ for all states and observations, respectively, over time *t*.

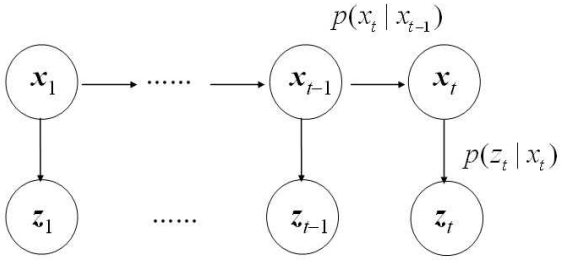

Fig. 1 The basic graphical structure of a state-space model

There are two basic assumptions in the model, which can be available by use of the d-separation property [2] of Bayesian Network. The first is the first-order Markov property, i.e.,

$$p(x_t \mid X_{t-1}) = p(x_t \mid x_{t-1}), \tag{1}$$

and the second is that the observations are mutually independent:

$$p(z_t \mid X_t, Z_{t-1}) = p(z_t \mid x_t). \tag{2}$$

According to the above two assumptions and Bayes' rule, the posterior probability of a state x_t given the past observations Z_t can be inferred as:

$$p(x_t \mid Z_t) = \frac{p(z_t \mid x_t)\, p(x_{t-1} \mid Z_{t-1})}{p(z_t \mid Z_{t-1})}, \tag{3}$$

where

$$p(x_t \mid Z_{t-1}) = \int p(x_t \mid x_{t-1})\, p(x_{t-1} \mid Z_{t-1}) dx_{t-1}. \tag{4}$$

Thus, the posterior probability can be computed by:

$$\begin{aligned} p(x_t \mid Z_t) &= \frac{p(z_t \mid x_t)}{p(z_t \mid Z_{t-1})} \int (x_t \mid x_{t-1})\, p(x_{t-1} \mid Z_{t-1}) dx_{t-1} \\ &\propto p(z_t \mid x_t) \int p(x_t \mid x_{t-1})\, p(x_{t-1} \mid Z_{t-1}) dx_{t-1} \end{aligned} \tag{5}$$

Hence, the posterior probability $p(x_t|Z_t)$ in a state-space model can be recursively computed by: (i) a likelihood model, $p(z_t|x_t)$, which relates the observation and noise to the state, and (ii) a transition model, $p(x_t|x_{t-1})$, which describes the possibility of the state change over time. Besides, it is also necessary to define the prior probability of state $p(x_1)$ at the beginning of the recursion.

IV. Formulation

People may appear different poses in different frames of a video, which generally makes the recognition problem more difficult. We attempted to overcome this problem by including additional pose information in a basic state-space system. A head can in general appear in different poses such as rotation and skew, but the range of head poses is restricted by

articulation limitations produced by the connection with the neck. Moreover, biomorphic features (e.g., eyes and nose) are similar when different people adopt the same head poses. For example, we may see only one eye of a person in a view of the right side of the face. In our past experiences, different people with similar head poses are likely recognized as similar, but one people with different head poses are sometimes difficult to be successfully identified. Head pose is an important factor disturbing the face recognition. Hence, our idea is to incorporate head poses with facial appearance features into our proposed model stated in the follows.

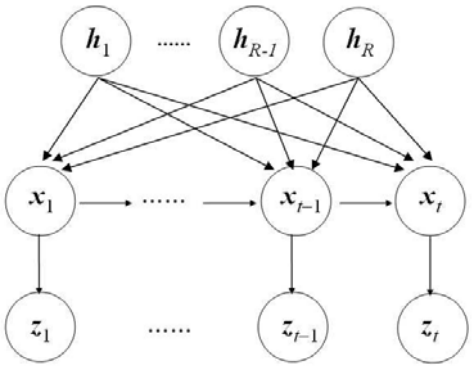

Fig. 2 Our proposed state-space model with additional pose nodes for video-based face recognition

Fig. 2 shows our proposed probabilistic graphical model for a revised state-space system. Assume that there are Q head poses, denoted as $H=\{h_1, ..., h_Q\}$, in the video frames. Since all possible poses can appear with most of people, we assume that people do not favor any special head pose. Hence, all Q pose nodes connect to each of system states $X_t=\{x_1, ..., x_t\}$ in Fig. 2. These Q extra nodes associated with head poses are then appended as the prior information to the state-space system. Note that the number of head poses in face recognition does not need to equal that in face detection and tracking, denoted as symbols Q and P, respectively.

Assume that K persons appear in a set of M consecutive video frames denoted as $\{I_1, ..., I_M\}$. We link up time t in a state-space system with changes in the video frames; that is, frame I_t is observed at time t in the system within the prior pose information. Hence, we summarize the formulation using the state-space model for video-based face recognition as follows:

- State vector x_t: to indicate which person (1 to K) is observed at time t.
- Observation z_t: the video frame I_t at time t.
- Prior pose information $H=\{h_1, ..., h_Q\}$: prior information for Q head poses.
- Goal: to estimate $p(x_t|Z_t, H)$ in order to identify which person appears at time t according to all (current and past) observed video frames and the available pose information.

Lemma.

Given a set of pose information $H=\{h_1, ..., h_Q\}$ and a set of observations $Z_t=\{z_1, ..., z_t\}$ at time t for the probabilistic graphical model shown in Fig. 2, the posterior probability of state x_t can be computed as

$$p(x_t \mid Z_t, H) \propto p(z_t \mid x_t) \int p(x_t \mid x_{t-1}, H) p(x_{t-1} \mid Z_{t-1}, H) dx_{t-1} \quad (6)$$

Proof.

According to the two assumptions, the first-order Markov property and the mutually independence, of the state-space model and using d-separation property [2] of Bayesian network for the probabilistic graphical model shown as Fig. 2, we could have the following four properties for conditional independence:

$$p(x_t \mid X_{t-1}, H) = p(x_t \mid x_{t-1}, H)$$
$$p(z_t \mid x_t, Z_{t-1}, H) = p(z_t \mid x_t)$$
$$p(x_t \mid X_{t-1}, Z_{t-1}) = p(x_t \mid X_{t-1})$$
$$p(H \mid X_t, Z_t) = p(H \mid X_t).$$

We then have

$$\begin{aligned} p(x_t \mid Z_t, H) &\propto \int p(X_t, Z_t, H) dX_{t-1} \\ &= \int p(H \mid X_t, Z_t) p(X_t, Z_t) dX_{t-1} \\ &= \int p(H \mid X_t) p(X_t, Z_t) dX_{t-1} \\ &= \int p(H \mid X_t) p(z_t \mid x_t) p(x_t \mid X_{t-1}) p(X_{t-1}, Z_{t-1}) dX_{t-1}. \end{aligned} \quad (7)$$

Since

$$\begin{aligned} p(H \mid X_t) &= \frac{p(H, X_{t-1}, x_t)}{p(x_t, X_{t-1})} \\ &= \frac{p(x_t \mid X_{t-1}, H) p(H \mid X_{t-1}) p(X_{t-1})}{p(x_t \mid X_{t-1}) p(X_{t-1})} \\ &= \frac{p(x_t \mid X_{t-1}, H) p(H \mid X_{t-1})}{p(x_t \mid X_{t-1})}, \end{aligned}$$

(7) becomes

$$\begin{aligned} & p(z_t \mid x_t) \int p(x_t \mid X_{t-1}, H) p(H \mid X_{t-1}) p(X_{t-1}, Z_{t-1}) dX_{t-1} \\ &= p(z_t \mid x_t) \int p(x_t \mid X_{t-1}, H) p(H \mid X_{t-1}, Z_{t-1}) p(X_{t-1}, Z_{t-1}) dX_{t-1} \\ &= p(z_t \mid x_t) \int p(x_t \mid X_{t-1}, H) p(H, X_{t-1}, Z_{t-1}) dX_{t-1} \\ &\propto p(z_t \mid x_t) \int p(x_t \mid x_{t-1}, H) p(x_{t-1} \mid Z_{t-1}, H) dx_{t-1}, \end{aligned}$$

and the proof is done.

Thus, there are three factors to determine the state x_t: (i) $p(z_t|x_t)$ is the likelihood measure for current observations, (ii) $p(x_t|x_{t-1}, H)$ is the transition measure between two consecutive states based on prior pose information, and (iii) $p(x_{t-1}|Z_{t-1}, H)$ is the recursive result at the previous iteration. Likelihood measure $p(z_t|x_t)$ can be achieved by traditional face recognition in a single image, for which 2DLDA [5] was adopted in this

study. We also propose a transition measure covering persons and poses to make the system flexible when incorrect recognition occurs. The details of our proposed likelihood and transition measures are presented in Sections V and VI, respectively. Another problem for the proposed model in (6) is the initialization of the state vector, $p(x_0|z_0)$. We apply the 2DLDA method to recognize the first face image for initialization.

V. Likelihood Measure

The likelihood term in (6), $p(z_t|x_t)$, measures the probability of the current observations given a certain state (i.e., a known person). It can be estimated from a similarity measure between the face image of the current observation and the training images of the given person or by applying face recognition to still images.

LDA [4] has been well studied in face recognition applications. Given a set of training face images, each associated with a known person, LDA can learn the plane that is best for discriminating these persons when these data are projected into the plane. A test face image can then be projected into the same plane to determine which person is most closely associated with the mean of the training images. In general, pixels of an image are arranged into a column vector for learning an LDA plane. Yang et al. proposed the IMLDA (uncorrelated image matrix-based LDA) technique [24] to preserve the 2D-matrix feature from an image for LDA plane learning. Furthermore, they performed IMLDA twice in the horizontal and vertical directions to implement 2DLDA [5]. The basic concept of 2DLDA is shown in Fig. 3, which can be considered as compressing an original image into a compact representation in the upper-left corner. Yang et al. also suggested a strategy to select the most discriminative features from the compressed corner [5]. In our work, we simply reduce the dimension of an image into a $d\times d$-dimensional vector.

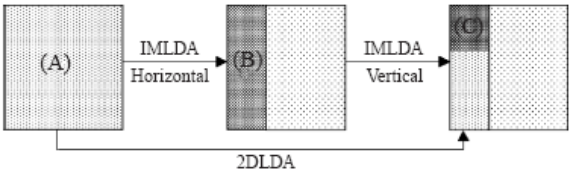

Fig. 3 The basic concept of 2DLDA

Given a training set D of face images including K persons, a 2DLDA projection could be trained, and a video frame of observation can be transformed into a $d\times d$-dimensional vector. Suppose that m_i is the mean of the projected points for training face images associated with person M_i (i=1 to K). Also, let z'_t be the projected point of an observation z_t; that is, a video frame I_t at time t. We can compute (z'_t-m_i), which estimates the difference between the observed face image and a known person in the 2DLDA plane, and normalize it for the approximated likelihood term,

$$p(z_t \mid M_i) = (2\pi)^{-d/2} \mid C \mid^{-1/2} \exp(-\frac{1}{2}(z_t^{'} - m_i)^t C^{-1}(z_t^{'} - m_i)), \quad (8)$$

where C is the covariance matrix of training images associated with person M_i in the 2DLDA plane. Thus, the likelihood term $p(z_t|x_t)$ in (6) can be approximated by $p(z_t|x_t=M_i)$, or simplifying $p(z_t|M_i)$, for each person M_i.

VI. Transition Measure

The transition term in (6), $p(x_t|x_{t-1}, H)$, measures the transitive probability from the previous to the current state in the system. This measure makes correction possible when the system performs an incorrect recognition. The transition measure is a static table that is built in before the system begins evolving. According to

$$\begin{aligned} p(x_t \mid x_{t-1}, H) &= \frac{p(H \mid x_t, x_{t-1})\, p(x_t \mid x_{t-1})\, p(x_{t-1})}{p(H \mid x_{t-1})\, p(x_{t-1})} \\ &= p(x_t \mid x_{t-1}) \frac{p(H \mid x_t, x_{t-1})}{p(H \mid x_{t-1})}, \end{aligned} \quad (9)$$

The transition measure can be divided into two parts as follows:

- $p(x_t|x_{t-1})$, which measures the transition probability of two consecutive states; that is, the transition among persons. This part is independent of the person's head poses.
- $p(H|x_t, x_{t-1})/p(H|x_{t-1})$, which measures the pose-transition likelihood of two consecutive states. This part is dependent on changes in head poses.

A. Transition among Persons

The first part of the transition measure, $p(x_t|x_{t-1})$, only depends on the recognition results of states at each iteration. Our idea is to compute the similarity measures between any two persons according to their training face images in the 2DLDA plane. That is, we estimate the transition measure of two persons based on how similar the two persons are in the 2DLDA plane, which is also used for the likelihood measure. A higher similarity of two persons in the 2DLDA plane increases the likelihood of our observation measure incorrectly recognizing them, and hence their transition probability should be higher.

Simply following the notations in Section V, let D_i be the data set of projected points in the 2DLDA plane for the training face images associated with persons M_i. The similarity of these two persons M_i and M_j can be defined as

$$sim(M_i, M_j) = \frac{1}{\mid D_i \mid}(\sum_{r \in D_i} (r - m_j)^t (r - m_j))^{1/2} \quad (10)$$

where m_j are means of projected points in the 2DLDA plane for the training images associated with person M_j. These similarity measures are also normalized by a Gaussian distribution. Note that $sim(M_i, M_j)$ is not symmetric, so we define

$$p(x_t \mid x_{t-1}) = (sim(M_i, M_j) + sim(M_j, M_i))/2 \quad (11)$$

for the transition measure between any two persons. The transition measure among persons depicts the between-class similarity of training face images according to the same scheme as that used for the likelihood measure.

B. Transition among Poses

The second part of the transition measure, $p(H|x_t, x_{t-1})/p(H|x_{t-1})$, only depends on head poses of persons at two consecutive iterations. Unfortunately, it seems difficult to deduce a closed form for this term. The numerator of the term, i.e., $p(H|x_t, x_{t-1})$, expresses the likelihood of head poses given two consecutive states, and the denominator, i.e., $p(H|x_{t-1})$, expresses the likelihood of head poses of the previous state. Hence, the entire measure can be approximated as the probability of changes in head poses between consecutive iterations t and t-1.

According to the approximation, there are two tasks to estimate the term for the transition among poses. We first identify the poses of the observed face images in the current and previous stages. We also build a 2DLDA classifier for recognizing head poses of the observed images. The 2DLDA classifier for face-pose recognition is similar to the face classifier described in Section V. Next, the probability of the pose changing from the current to the previous stage should be determined. We collected short videos containing different kinds of face movements and then counted the actual times that the pose changed between consecutive frames in order to compute the probabilities of changing from one pose to another pose. The approach used to count the number of pose transitions followed that described previously [15] except that our counting was based on all of the persons present rather than only certain individuals, due to the assumption of the independence between head poses and the person being observed stated in Section IV.

VII. Experimental Results

A. Data Set

In our experiments, we used a public data set, the Honda/UCSD Video Database [15], [25], to evaluate the performance of our proposed approach. This data set contains a set of 52 videos of 20 different persons. Each person appears in at least two videos: one for training and the other for testing. Thus, this data set contains a training part (20 videos) and a test part (the other 32 videos), with the former for learning tasks and the latter for test evaluations in the following experiments. The videos were recorded at 15 frames per second, with each frame comprising 640×480 pixels. The subjects in the data set rotate and turn their heads according to their own preferred order and speed, and hence the data set contains a wide range of different poses [15]. Some individuals in the testing videos appear in special poses that are not present in the training videos. Fig. 4 illustrates examples of video frames in the data set.

Fig. 4 Snapshots of videos in the Honda/UCSD dataset

B. Evaluation

This section describes several experiments that we performed to quantitatively evaluate our proposed method of video-based face recognition. The first task is to learn a 2DLDA classifier for still-image-based face recognition using the training part of the Honda/UCSD data set. The main question is what dimension is feasible for the 2DLDA classifier. We transformed images into d×d-dimensional feature vectors, as described in Section V, using several values of d; the average rates of face recognition for still images are listed in Table I. It may be due to overfitting with higher dimensions so that larger values of d cannot reach good performances. From these results, we adopted d=5 for the highest rate in the subsequent experiments.

TABLE I
Recognition Rates Using 2DLDA for Still Images of Video Frames with Different Dimensions

dim (d×d)	3×3	4×4	5×5	6×6	7×7	8×8	9×9
reg rate (%)	63.80	74.23	80.17	78.43	76.86	73.91	69.44

The next experiment tested the efficiency of the proposed method of video-based face recognition. For each test video of persons in the Honda/UCSD data set, we generated 10 subvideos by randomly capturing 100 consecutive frames (about 6 seconds). Thus, a total of 200 test videos with 20 different persons were employed in the experiments. Table II lists the average recognition rates for our proposed approach for three transition cases: without transition, with transitions between persons only, and with transitions between persons and between head poses. In the absence of transitions the face recognition in video frames was performed only according to the trained 2DLDA classifier. Table II indicates that there was a significant improvement for our approach when the transition information (either on persons or head poses) was incorporated in the model. For comparison, Table III lists the average rates of face recognition using different well-known methods based on this same data set.

TABLE II
THE AVERAGE RATES OF FACE RECOGNITION WITH/WITHOUT DIFFERENT TRANSITION APPROACHES

	without transition	transition among persons	transition among persons and poses
reg rate (%)	80.17	87.33	90.67

TABLE III
THE AVERAGE RATES OF FACE RECOGNITION USING DIFFERENT METHODS

	Eigne-Face	Fisher-Face	Nearest Neighbor	2DLDA	Our Approach
reg rate (%)	69.3	74.5	81.6	80.7	90.67

Now consider the convergence process with the likelihood and transition measures over time. Fig. 5 illustrates an example of face recognition at times 8, 14, 23, and 28. Note that the person in this the example corresponded to index "4" in the plots. His head poses changed from front to left in this example. For simplicity, this example only gives the probability values for three persons. There are five plots at each row. The first plot shows the likelihood measure of the current observation according to (8). Plots 2 to 4 display the probabilities of head poses for different persons given the observation. The last plot shows the final probability of person recognition given the observed face image according to (6). He was initially identified incorrectly (at t=8) but finally recognized correctly (at t=28). In general, it is difficult to avoid incorrect decisions about either face or pose recognition. However, our method makes it possible to converge to the correct decision by aggregating the recognitions in the likelihood and transition measures such as illustrated in the last two iterations.

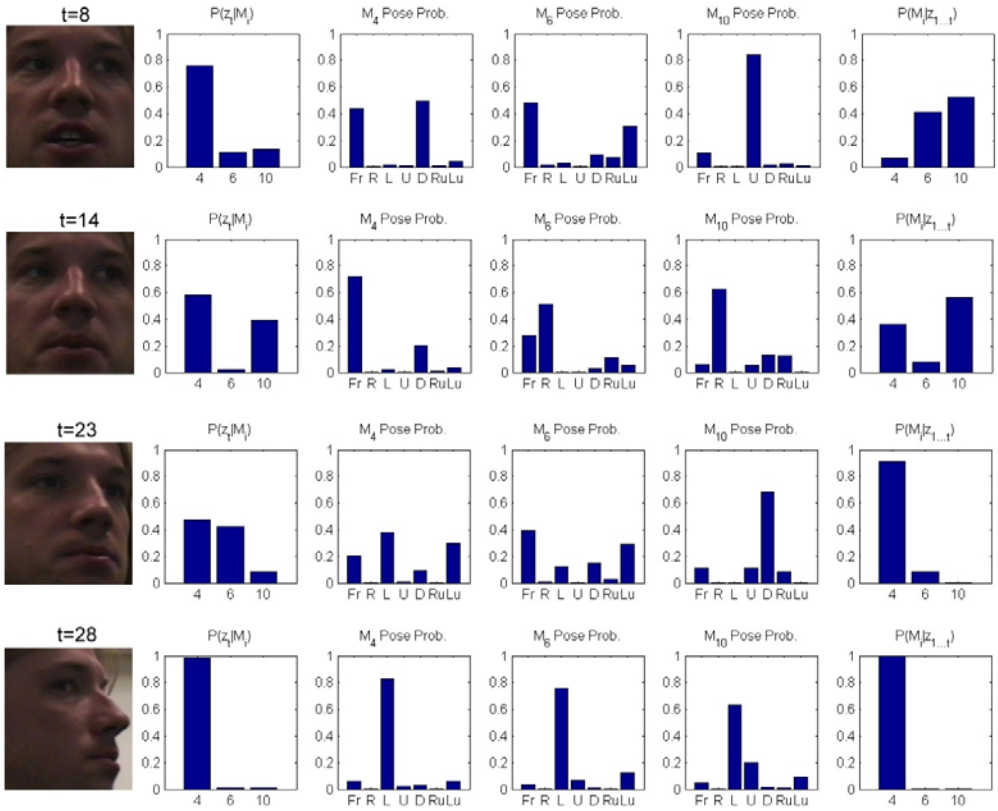

Fig. 5 An illustration of the convergence for the recognition process over time

VIII. CONCLUSION AND FUTURE WORK

This paper presents a face recognition approach that is based on a state-space model with extra nodes associated with head poses in video sequences. Our formulation integrates the likelihood measure for the still-image-based face recognition in video frames and the transition measures for modeling possible changes both between persons and between head poses. Moreover, the transition measure makes it possible to change recognition results from an incorrect decision to the correction one. Our experiments demonstrate that the proposed method can achieve a good performance for face recognition.

Future research is to extend this work in a real application such as a roll-call system in a classroom. Many complex issues should be carefully considered in a real system, e.g., how to track or locate the face areas in video? How to involve cosmetic, glasses, and hair styles in our recognition model? Besides, another important task is to improve the approximation for the transition measure among persons and poses in (9). We also plan to design an incremental learning algorithm with our state-space model and thereby make our approach even more robust.

ACKNOWLEDGMENT

This work was supported by National Science Council,

Taiwan, under Grant No. NSC 100-2631-S-003-006 and by Ministry of Economic Affairs, Taiwan, under Grant No. 100-EC-17-A-02-S1-032.

REFERENCES

[1] Z. Ghahramani, "An introduction to hidden Markov models and Bayesian networks," International Journal of Pattern Recognition and Artificial Intelligence, vol. 15, no. 1, pp. 9–42, 2001.
[2] K. P. Murphy, "Dynamic Bayesian networks: representation, inference and learning", U. C. Berkeley, PhD. Thesis, 2002.
[3] M. Turk and A. Pentland, "Eigenface for recognition," Journal of Cognitive Neuroscience, vol. 3, no. 1, pp. 71–86, 1991.
[4] R.O. Duda, P.E. Hart, and D.G. Stork. Pattern classification. In second ed., John Wiley and Sons, Inc, 2001.
[5] J. Yang, D. Zhang, X. Yong, and J. Yang, "Two-dimensional discriminant transform for face recognition," Pattern Recognition, vol. 38, no. 7, 2005.
[6] B. Li, C.-H. Zheng, and D.-S. Huang, "Locally linear discriminant embedding: An efficient method for face recognition," Pattern Recognition, vol. 41, pp. 3813–3821, 2008.
[7] X.-F. He, S.-C. Yan, Y.X. Hu, P. Niyogi, and H.-J. Zhang, "Face recognition using Laplacianfaces," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 27, no. 3, pp. 328–340, March 2005.
[8] X. F. He and P. Niyogi, "Locality preserving projections," in Proceedings of 17th Annual Conference on Neural Information Processing Systems, NIPS, 2003.
[9] X. Liu and T. Chen, "Video-based face recognition using adaptive hidden markov models," in Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition, CVPR, vol. 1, pp. 340–345, 2003.
[10] B. Heisele, P. Ho, and T. Poggio, "Face recognition with support vector machines: Global versus component-based approach," in Proceedings of International Conference on Computer Vision, ICCV, 2001.
[11] P.-H. Lee, G.-S. Hsu, and Y.-P. Hung, "Face verification and identification using facial trait code," in Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition, CVPR, 2009.
[12] A. Rama, F. Tarrs, and J. Rurainsky, "Aligned texture map creation for pose invariant face recognition," Multimedia Tools and Applications, vol. 24, no. 2, pp. 321–335, 2010.
[13] A.A. Mohammed, R. Minhas, Q.M. Jonathan Wu, and M.A. Sid-Ahmed, "Human face recognition based on multidimensional PCA and extreme learning machine," Pattern Recognition, vol. 44, no. 10–11, pp. 2588–2597, 2011.
[14] G. Shakhnarovich, J. W. Fisher, and T. Darrell, "Face recognition from long-term observations," In Proceedings of European Conference on Computer Vision, ECCV, pp. 851–865, 2002.
[15] K.-C. Lee, J. Ho, M.-H. Yang, and D. Kriegman, "Visual tracking and recognition using probabilistic appearance manifolds," Computer Vision and Image Understanding, vol. 99, no. 3, pp. 303–331, 2005.
[16] C.-Y. Lu, J.-M. Jiang, and G.-C. Feng, "A boosted manifold learning for automatic face recognition," International Journal of Pattern Recognition and Artificial Intelligence, vol. 24, no. 2, pp. 321–335, 2010.
[17] R.-P. Wang, S.-G. Shan, X.-L. Chen, and W. Gao, "Manifold-manifold distance with application to face recognition based on image set," in Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition, CVPR, 2008.
[18] Z. Stone, T. Zickler, and T. Darrell, "Toward large-scale face recognition using social network context," in the Proceedings of the IEEE, vol. 98, no. 8, 2010.
[19] R.G. Cinbis, J. Verbeek, and C. Schmid, "Unsupervised metric learning for face identification in TV video," in Proceedings of IEEE International Conference on Computer Vision, ICCV, 2011.
[20] F. B. ter Haar and R. C. Veltkamp, "3d face model fitting for recognition," in Proceedings of European Conference on Computer Vision, ECCV, 2008.
[21] X.-Z Zhang and Y.-S. Gao, "Face recognition across pose: a review, " Pattern Recognition, vol. 42, pp. 2876–2896, 2009.
[22] M. Kim, S. Kumar, V. Pavlovic, and H. Rowley, "Face tracking and recognition with visual constraints in real-world videos," in Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition, CVPR, 2008.
[23] O. Arandjelovic and R. Cipolla, "A pose-wise linear illumination manifold model for face recognition using video," Computer Vision and Image Understanding, vol. 113, pp. 113–125, 2009.
[24] J. Yang, J.-Y. Yang, A.F. Frangi, and D. Zhang, "Uncorrelated projection discriminant analysis and its application to face image feature extraction," International Journal of Pattern Recognition and Artificial Intelligence, vol. 17, no. 8, pp. 1325–1347, 2003.
[25] The honda/ucsd video database, http://vision.ucsd.edu/leekc/ hondaucsdvideodatabase/hondaucsd.html.