Integration of Likelihood and Transition Measures for Video-Based Face Recognition

Cheng-Chieh Chiang
Dept. of Information Tech.,
Takming University of Science & Tech.
Email: kevin@csie.ntnu.edu.tw

Yi-Chia Chan
Dept. of Comp. Science and Info. Eng.,
National Taiwan Normal University
yichia0427@yahoo.com.tw

Greg C. Lee
Dept. of Comp. Science and Info. Eng.,
National Taiwan Normal University
leeg@csie.ntnu.edu.tw

***Abstract*—This paper presents a probabilistic graphical model to formulate video-based face recognition. There are two main parts involving in our approach: one for likelihood measure and the other for transition measure. The likelihood measure can be viewed as a traditional task of face recognition within a still image, i.e., to recognize who the current observing face image is. Two-dimensional linear discriminant analysis (2DLDA) is employed to judge the likelihood measure. Moreover, the transition measure estimates the probability of the change from the recognized state at the previous stage to each of possible states at the current stage. Our approach for the transition measure can consider both the visual difference of persons according to the training face images and the pose change over time in video frames. We also provide several experiments to show the efficiency of our proposed approach in this paper.**

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

I. Introduction

Video data has been widely used for many kinds of applications in our life, e.g., camera in a hand device for capturing our life, web camera in a laptop for sending video messages, or surveillance cameras in city streets for security goals. Many kinds of technologies for video data have been developed for their goals of applications. It is an important task to know who appears in a video when we design a video-based application. Thus, face recognition is often a key technology in many video-based applications, which aims to recognize which persons appear in a video sequence.

In traditional, face recognition is treated as a supervised learning, i.e., classifiers are trained by a set of prepared face images associated with persons and then new face images are recognized by use of the classifiers. It is a long history to develop technologies of face recognition in still images [1][15][20]. Different methods of classifier learning, e.g., eigenface [16], PCA (Principal Component Analysis) and LDA (Linear Discriminant Analysis) [3], LPP (Locality Preserving Projection) [5], and SVM (Support Vector Machine) [6], have been proposed to deal with the problem.

In recent, researchers have paid more attention to the task of video-based face recognition that recognizes who appears in a video stream. In principle, video-based face recognition can be regarded as a fusion of recognition results in a set of sequential and consecutive images. However, there are, in fact, more information hidden in video frames. For example, a face may keep moving in a video so that different face poses should be involved. Incorporating all of recognition results for different face poses appeared in a video could overcome some difficult cases of recognition for special face poses. A face recognition method using temporal voting to incorporate results of still images is proposed for image sequences in [14]. Considering in continuous video frames, visual features extracted from face images could form a manifold in high-dimensional feature space. Thus, we can convert the problem of face recognition to a matching problem between the corresponding manifolds [2][10][17]. Regarding the relationship of faces/poses in consecutive video frames, HMM (Hidden Markov Model) is often used for building a face model [9][11]. Another approaches treat face images from video frames as 3D models and the recognition problem is converted to match and search 3D models for face images [4][7][12].

This work deals with video-based face recognition in a static environment such as a classroom that the members are fixed. We assume there are K persons in the system. In a video, these persons may appear with different face poses, or not appear. Similarly, we have their face and pose images for training. Our goal is to build a model to recognize whose face in a video is.

Our basic idea is like a tracking task: to track the selection in the K candidates over time according to the observations of visual features in video frames. That motivates us to employ the state-space model to construct a probabilistic graphical model for video-based face recognition. Our formulation divides video-based face recognition into two parts: likelihood and transition measures. The former is like a traditional task of face recognition in a still image to make a decision who the current observing face image is. The latter estimates the probability of the change from the recognized state at the previous stage to each of possible states at the current stage. The transition measure can make it possible to change recognition results from a false to the correction decision.

The rest of this paper are organized as the follows. Section 2 formulates the problem of video-based face recognition based on a probabilistic graphical model by revising a basic state-space model. Next, in Section 3, we describe how to perform the face recognition in still images using 2DLDA for the likelihood measure in our formulation. Therefore, how to measure the transition probabilities among persons and face poses are presented in Section 4. Section 5 provides several experimental results to show the performance of our proposed approach, and Section 6 draws our conclusion and future works.

II. Formulation

A. 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 [8][13]. Here, we only provide a brief summary of how the posterior probability of a state-space model is inferred.

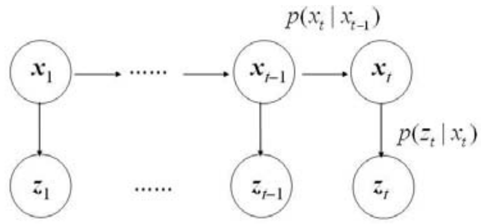

Figure 1: The graphical structure of a state-space model.

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 Figure 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.

There are two basic assumptions in the model, which can be available by use of the d-separation property [13] of Bayesian Network. The first is the first-order Markov property, i.e.,

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

and the second is that the observations are mutually independent:

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

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:

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

where

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

Thus, the posterior probability can be computed by:

$$\begin{aligned} p(x_t \mid Z_t) &= \frac{p(z_t \mid x_t)}{p(z_t \mid Z_{t-1})} \int (x_t \mid x_{t-1})\, p(x_{t-1} \mid Z_{t-1})\, dx_{t-1} \\ &\propto p(z_t \mid x_t) \int p(x_t \mid x_{t-1})\, p(x_{t-1} \mid Z_{t-1})\, dx_{t-1} \end{aligned} \quad (4)$$

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.

B. Formulation for video-based face recognition

A video in general consists of a set of consecutive video frames. Hence, to recognize who appears in a video could be considered the recognition in the set of video frames. That means a task of video-based face recognition could be regarded as a collection of traditional face recognition in many still images from a video. Moreover, there are some temporal relationships among these still images in a video. The situation motivated us to design a state-space model for video-based face recognition to involve the traditional recognition task in a set of video frames and the temporal information over video frames.

Given a set of consecutive video frames denoted as $\{I_1, ..., I_N\}$ with N images, assume there are K persons appeared in the system. We link up the time t in a state-space system with the change of video frames, i.e., frame I_t is observed at time t in the system. Following the notation stated in the previous section, hence, a state-space model for video-based face recognition could be formulated as the follows.

- state vector x_t: to indicate which person (from 1 to K) observed at time t.
- observation z_t: the video frame I_t at time t.
- goal: to estimate $p(x_t|Z_t)$ that recognizes which person appears at time t according to all (including current and past) observing video frames.

Summarily, the observation set $Z_t=\{z_1, ..., z_t\}$ collects the face images in video frames, and $X_t=\{x_1, ..., x_t\}$ shows the recognition results of face images of these observations.

However, the basic state-space model shown in Figure 1 could not reach an accurate recognition while people are changing their poses in video frames. In order to overcome the change of head poses for face recognition in video frames, our approach, in this paper, is to insert additional pose nodes which express head poses appeared.

Assume there are R head poses, denoted as $H=\{h_1, ..., h_R\}$, for moving head in video frames. Then, R extra nodes associated with head poses are appended to our proposed probabilistic model shown as Figure 2. In general, a head can appear with different poses such as rotation and skew, but head poses are limited by the articulation connected with the neck. Similar biomorphic features, e.g., eyes and nose, could be observed with the same head poses even for different people. For example, we may see only one eye of a person through the view of the right-side face. Hence, we can assume these R pose nodes are prior information to the state-space system.

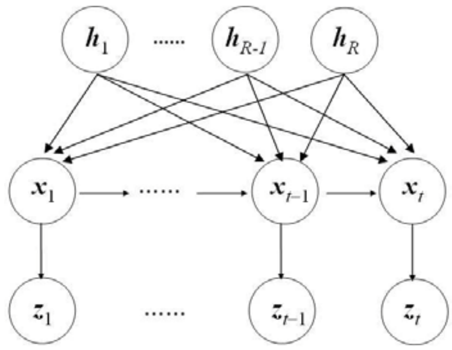

Figure 2: The probabilistic structure of the state-space model for face recognition with pose nodes.

Lemma:
Given the pose information $H=\{h_1, ..., h_R\}$ and the set of observations $Z_t=\{z_1, ..., z_t\}$ at time t for the Bayesian network in Figure 2, the posterior probability of the state x_t can be computed as:

$$p(x_t \mid Z_t, H) \propto p(z_t \mid x_t)\int p(x_t \mid x_{t-1}, H)p(x_{t-1} \mid Z_{t-1}, H)dx_{t-1} \quad (5)$$

Proof:

According to the two assumptions in Eq. (1) and (2), and using the d-separation property [13] of Bayesian network for Figure 2, we could have the following four properties of conditional independence:

$$p(x_t \mid X_{t-1}, H) = p(x_t \mid x_{t-1}, H), \quad (6)$$

$$p(z_t \mid x_t, Z_{t-1}, H) = p(z_t \mid x_t), \quad (7)$$

$$p(x_t \mid X_{t-1}, Z_{t-1}) = p(x_t \mid X_{t-1}), \quad (8)$$

and

$$p(H \mid X_t, Z_t) = p(H \mid X_t). \tag{9}$$

Then,

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

and the proof is done. □

That shows there are three factors to determine which person the state x_t is: (i) $p(z_t|x_t)$ means the likelihood measure for the current observation, (ii) $p(x_t|x_{t-1}, H)$ means the transition measure based on pose information for the previous state, and (iii) $p(x_{t-1}|Z_{t-1}, H)$ is the recursive result at the previous iteration. Besides, the system needs an initial recognition result, denoted $p(x_1)$, of the first face image in a video for indicating the prior probability.

In order to more simply achieve face recognition according to Eq. (5) in practice, two assumptions are held in this paper. The first is to assume that face images have been properly cropped in video frames. That can be performed by face detection. The second is to assume that poses of face images are aligned. That is to say, we define R poses for face images and each of training face images can be categorized into a pose. In our work, we apply k-means clustering to roughly divide training face images into R subsets and manually check whether face images are the same pose in the same subset.

III. Face Recognition for Still Images

The likelihood term, denoted as $p(z_t|x_t)$, in Eq. (5) measures the possibility of the current observations given a state (i.e., a known person). That can be estimated by the similarity measure between the face image of the current observation and training images of the given person. Thus, the computation of the likelihood measure for face images in video frames associated with each of time t can be regarded as face recognition for still images.

In this work, we adopt 2DLDA (two-dimensional linear discriminant analysis) [19] for performing face recognition in a still image. 2DLDA employs IMLDA (uncorrelated image matrix-based linear discriminant analysis) [18] twice: one for the horizontal and the other for the vertical direction shown as Figure 3 which is taken from [19]. In principle, 2DLDA can select most discriminative features learned from training images. We roughly describe the procedure of 2DLDA as the follows.

Assume the image set D consists of K categories of face images, associated with K persons, and suppose the image size, with loss of generality, is $m \times n$. According to the computation of IMLDA presented in [18], a transformation $\boldsymbol{U}$ with size $n \times d_1$ can be learned. For each image $\boldsymbol{A}$ in D, we can compute $\boldsymbol{B}=\boldsymbol{AU}$ and collect all matrices $\boldsymbol{B}$ with size $m \times d_1$ as the set D'. Similarly, another transformation $\boldsymbol{U}'$ with size $m \times d_2$ could be learned by use of IMLDA on the dataset D'. Compute $\boldsymbol{B}^t\boldsymbol{U}$ with size $d_1 \times d_2$ for each $\boldsymbol{B} \in D'$ and in final an original $m \times n$ matrix in the image set D can be converted to a $d_1 \times d_2$ matrix.

The basic concept of 2DLDA could be shown as Figure 3; that could be viewed as to compress an original image into a compact representation in the up-left corner. Yang et al. [19] also suggested a feature selection strategy to select the most discriminative features from the compressed corner. In this work, we simply set $d=d_1=d_2$ for reducing the dimension of an image as a $d \times d$- dimensional vector and treat it as $d^2 \times 1$-dimensional column vector.

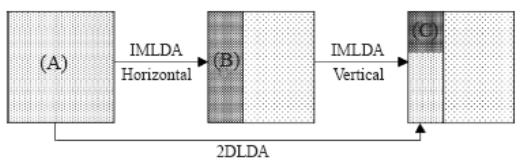

Figure 3: Illustration of 2DLDA.

In the 2DLDA plane, suppose that m_i is the mean of 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, i.e., a video frame at time t. We can compute (z'_t-m_i) to estimate the difference between the observing face image and a known person in the 2DLDA plane, and normalize it for the approximated likelihood term,

$$p(z_t \mid M_i) = (2\pi)^{-\frac{d}{2}} \mid C \mid^{-\frac{1}{2}} \exp(-\frac{1}{2}(z_t^{'} - m_i)^t C^{-1} (z_t^{'} - m_i)) \quad (10)$$

where C is the covariance matrix of training images associated with the person M_i in the 2DLDA plane. Thus, the likelihood term $p(z_t|x_t)$ in Eq. 5 can be approximated by $p(z_t|x_t=M_i)$, or simplifying $p(z_t|M_i)$, for each person M_i.

IV. Transition Measure in Video-based Face Recognition

The transition term, denoted as $p(x_t|x_{t-1}, H)$, in Eq. 5 measures the transitive possibility from the previous to the current state in the system. That measure can make correction possible while the system sometimes has a false recognition. According to the following equation,

$$\begin{aligned} p(x_t \mid x_{t-1}, H) &= \frac{p(H \mid x_t, x_{t-1}) p(x_t \mid x_{t-1}) p(x_{t-1})}{p(H \mid x_{t-1}) p(x_{t-1})} \\ &= p(x_t \mid x_{t-1}) \frac{p(H \mid x_t, x_{t-1})}{p(H \mid x_{t-1})} \end{aligned} \quad (11)$$

the transition measure can be divided into two parts described as the follows.

- $p(x_t|x_{t-1})$. That term measures the transition probability of two consecutive states. This part is independent of the persons' head poses.
- $p(H|x_t, x_{t-1})/p(H|x_{t-1})$. That term measures the pose-transition likelihood of two consecutive states.

These two terms are described in detail in the follows.

A. Transition among persons

Regarding the first term, $p(x_t|x_{t-1})$, of the transition measure, it only depends on the recognition results of states at each iteration. The transition measure among persons is a fixed table that is built in prior before the system begins evolving. Moreover, the design must provide the ability of correction for false recognition using 2DLDA. Our idea is to compute the similarity measures between any two persons according to their training face images in the 2DLDA plane. That is to say, we estimate the transition measure of any two persons by use of how similar the two persons are in the 2DLDA plane which is also used for the likelihood measure. While two persons are similar observed in the 2DLDA plane, which means it is more possible to false recognize them in our observation measure, and their transition probability should be higher.

Simply following the notations in Section 3, let D_i be the dataset of projected points in the 2DLDA plane for the training face images associated with persons M_i. Then, the similarity of these two persons can be defined as

$$sim(M_i, M_j) = \frac{1}{\mid D_i \mid} (\sum_{r \in D_i} (r - m_j)^t (r - m_j))^{1/2} \quad (12)$$

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

$$p(x_t \mid x_{t-1}) = (sim(M_i, M_j) + sim(M_j, M_i))/2 \quad (13)$$

for the symmetric property.

B. Transition among poses

Regarding the second term $p(H|x_t, x_{t-1})/p(H|x_{t-1})$, it is difficult to induce a closed-form for explaining the term. Observing the sub-terms, $p(H|x_t, x_{t-1})$ and $p(H|x_{t-1})$, they can be roughly considered the face poses at the current and the previous stages, respectively. Hence, the second term $p(H|x_t, x_{t-1})/p(H|x_{t-1})$, in this work, is approximated to the possibility of the change of the face poses in successive iterations t and t-1.

According to the approximation, there are two tasks to estimate the term of the transition among poses. We first recognize which poses the observing face images of the current and the previous stages are. We also build a 2DLDA classifier for recognizing face poses of observing images. The 2DLDA classifier for face-pose recognition is similar to the face classifier described in Section 3. Next, the probabilities of the pose change from the current to the previous stage should be determined. We collect a little of videos which contain different kinds of face moving and then count the actual times of the pose change in each of two consecutive frames to compute the probabilities from one to another poses. Note that the counting approach of the pose transition is referred to [10] except our counting is based on all of persons, not on individuals.

V. Experimental Results

In our experiments, the Honda/UCSD Video Database [10][21] is adopted for our training and test dataset to evaluate the performance of our proposed approach. This dataset contains 20 different persons. For each person, there are two videos: one for training and the other for testing. In each video, the person rotates and turns his/her head in his/her own preferred order and speed, and typically in about 15 seconds, the individual is able to provide a wide range of different poses [10]. Sometimes an individual in the testing video may show some special poses which are not appeared in the training video.

We employ the training part of the dataset to learn a 2DLDA classifier for still-image face recognition and to determine the transition probabilities for persons and face poses. The first question is what dimension is feasible in the 2DLDA classifier. We transform each of images to $d \times d$- dimensional feature vectors, stated as Section 3, by trying several values of d, showing the average rates of face recognition for still images in Table 1. Then, we adopted the d=5 for the highest rate in the following experiments.

Table 2 shows the average of recognition rates which are based on our proposed approach with three transition cases: without transition, with transition only among persons, and with transition among both persons and face poses. The case of without transition means the face recognition in video frames is performed only according to the trained 2DLDA classifier. Table 2 presents a significant improvement for our approach when the transition information, either on persons or face poses, is incorporated in the model. Moreover, we also list the average rates of face recognition using different well-known methods in Table 3 for comparison.

Table 1: Recognition rates using 2DLDA for still images of video frames with different dimensions *d*.

dim $d \times d$	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

Table 2: The average rates of the 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 3: The average rates of the face recognition using different well-known methods.

	Eigen-Face	Fisher-Face	Nearest Neighbor	2DLDA	Our Approach
reg rate (%)	69.3	74.5	81.6	80.17	90.67

Next, let us discuss the convergence process with the likelihood and the transition measure over time. Figure 4 illustrates an example of face recognition at time 8, 14, 23, 28. Note that the person of the example is with index “4” in plots. His face poses changed from front to left in this example. The observing person is identified incorrectly in initial (t=1), but he is recognized correctly in final (t=28). There are five plots at each row. This example only displays the probability values for three persons for simplicity. The first plot shows the likelihood measure of the current observation according to Eq. (10). The second to fourth plots display the probabilities of face poses for different persons given the observation. The last plot shows the final probability of persons given the observing face image. In general, it is difficult to avoid false decision either for face or pose recognition. However, our method makes a possibility converging to the correct decision by aggregating the recognitions in the likelihood and the transition measures such as illustrated in the last two iterations.

VI. Conclusion and Future Works

This paper deals with video-based face recognition to determine which persons appear in a video sequence. We formulate the problem using a probabilistic graphical model which is based on a state-space model to integrate both the likelihood and the transition measures in recognition. We employ 2DLDA to perform the face recognition in still images to compute the likelihood measure. Also, we design a transition measure to cover the change of persons and face poses appeared in a video. The integration of the likelihood and transition measures in our formulation can improve the performance of video-based face recognition, as stated in our experiments.

Regarding the future extension of this work, we are performing some detailed experiments on more kinds of datasets to evaluate the abilities of our approach. We also try to revise our model more accurately to formulate the problem. Another possible way is to design an incremental learning algorithm with our probabilistic model. That could improve face or pose recognition in the likelihood and transition measures and make our proposed model more robust.

ACKNOWLEDGMENT

This work was supported by National Science Council, Taiwan, under Grant No. NSC 97-2511-S-003- 007-MY3 and NSC 97-2631-S-003-003 and by Ministry of Economic Affairs, Taiwan, under Grant No. 97-EC-17-A -02-S1-032.

REFERENCE

[1] A. F. Abate, M. Nappi, D. Riccio, and G. Sabatino , “2D and 3D Face Recognition: a Survey”, Pattern Recognition Letters, Vol. 28, No. 14, pp. 1885-1906, October 2007.

[2] O. Arandjelovic, G. Shakhnarovich, J. Fisher, R. Cipolla, and T. Darrell, “Face Recognition with Image Sets Using Manifold Density Divergence,” in Proceedings of CVPR, 2005.

[3] P. Belhumeur, J. Hespanha, and D. Kriegman, “Eigenfaces vs. Fisherfaces: Recognition Using Class Specific Linear Projection,” IEEE Trans. on PAMI, Vol. 19, No. 1, pp. 711 – 720, 1997.

[4] F. B. ter Haar and R. C. Veltkamp, “3D Face Model Fitting for Recognition,” in Proceedings of ECCV, 2008.

[5] X. F. He, and P. Niyogi, “Locality Preserving Projections.”, In Proceedings of NIPS, 2003.

[6] B. Heisele, P. Ho, and T. Poggio, “Face Recognition with Support Vector Machines: Global versus Component-Based Approach,” in Proceedings of ICCV, 2001.

[7] Y.-X. Hu, D.L. Jiang, S.-C. Yan, L. Zhang, and H.-J. Zhang, “Automatic 3D Reconstruction for Face Recognition,” in Proceedings of International Conference on Automatic Face and Gesture Recognition, 2004.

[8] Z. Ghahramani, “An Introduction to Hidden Markov Models and Bayesian Networks”, International Journal of Pattern Recognition and Artificial Intelligence, 15(1): 9-42, 2001.

[9] M. Kim, S. Kumar, V. Pavlovic, and H. Rowley, “Face Tracking and Recognition with Visual Constraints in Real-World Videos”, in Proceeding of CVPR, 2008.

[10] K.C. Lee and J. Ho and 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.

[11] X. Liu and T. Chen, “Video-Based Face Recognition Using Adaptive Hidden Markov Models”, in Proceedings of CVPR, 2003.

[12] X.-G. Lu, A.K. Jain, and D. Colbry, "Matching 2.5D Face Scans to 3D Models,", IEEE Trans. on PAMI, Vol. 28, No. 1, pp. 31-43, 2006.

[13] K. P. Murphy, "Dynamic Bayesian Networks: Representation, Inference and Learning", U. C. Berkeley, PhD. Thesis, 2002.

[14] G. Shakhnarovich, J. W. Fisher, and T. Darrell, "Face Recognition from Long-Term Observations," in Proceedings of ECCV, pp. 851-865, 2002.

[15] A. S. Tolba, A.H. El-Baz, and A.A. El-Harby, "Face Recognition: a Literature Review", International Journal of Signal Processing, Vol. 2, No. 1, pp. 88-103, 2005.

[16] M. Turk and A. Pentland, "Eigenfaces for Recognition", Journal of Cognitive Neuroscience, Vol. 3, pp. 72-86, 1991.

[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 CVPR, 2008.

[18] 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.

[19] J. Yang, D. Zhang X. Yong, and J. Yang, "Two-Dimensional Discriminant Transform for Face Recognition," Pattern Recognition, Vol. 38, No. 7, July 2005.

[20] W. Zhao, R. Chellappa, P.J. Phillips, and A. Rosenfeld, "Face Recognition: a Literature Survey", ACM Computing Surveys, 35(4): 399-458, 2003.

[21] The Honda/UCSD Video Database, http://vision.ucsd.edu/ ~leekc/HondaUCSDVideoDatabase/HondaUCSD.html.

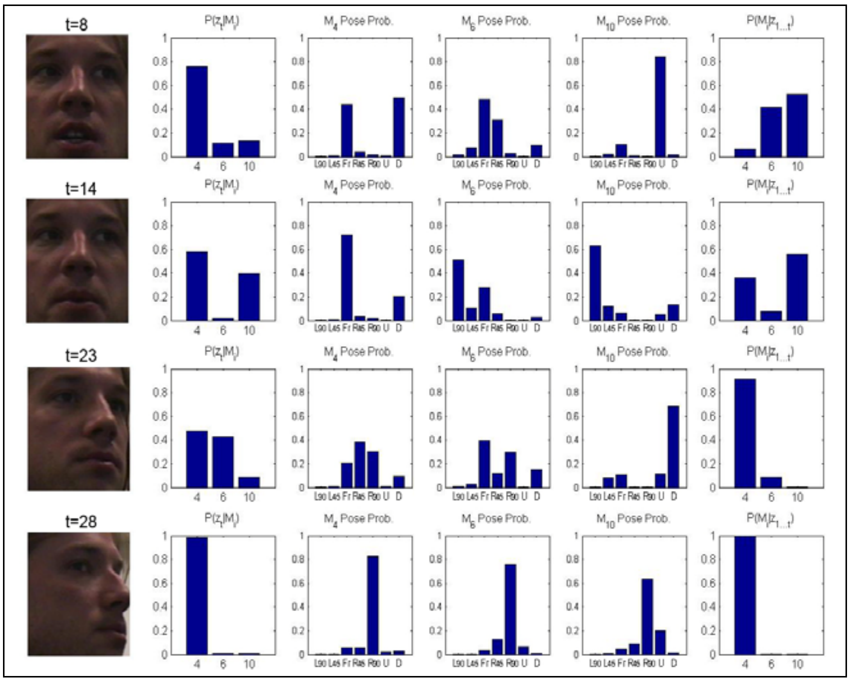

Figure 4: Illustration of the recognition process over time.