

# Multi-Eigenspace Learning for Video-Based Face Recognition

Liang Liu<sup>1</sup>, Yunhong Wang<sup>2</sup>, and Tieniu Tan<sup>1</sup>

<sup>1</sup>National Laboratory of Pattern Recognition  
Institute of Automation, Chinese Academy of Sciences, Beijing, China  
{lliu, tnt}@nlpr.ia.ac.cn

<sup>2</sup>School of Computer Science and Engineering  
Beihang University, Beijing, China  
yhwang@buaa.edu.cn

**Abstract.** In this paper, we propose a novel online learning method called Multi-Eigenspace Learning which can learn appearance models incrementally from a given video stream. For each subject, we try to learn a few eigenspace models using IPCA (Incremental Principal Component Analysis). In the process of Multi-Eigenspace Learning, each eigenspace generally contains more and more samples except one eigenspace which contains the least number of samples. Then, these learnt eigenspace models are used for video-based face recognition. Experimental results show that the proposed method can achieve high recognition rate.

**Keywords:** face recognition, online learning, incremental principal component analysis.

## 1 Introduction

For video-based face recognition, most state-of-the-art algorithms [1, 2, 5, 9, 10, 12] can perform the recognition task in real time. However, the training process commonly runs off-line in a batch mode. Though several online learning algorithms were proposed recently, those algorithms generally need to perform online learning based on a pre-trained model. A pre-trained model is typically trained from a data set which is manually collected and labeled. The task of collecting and labeling data is usually tedious and boring. In addition, a pre-trained model may not be flexible enough to cope with various conditions online. Therefore, how to learn appearance models online without a pre-trained model is valuable to be considered.

Compared with batch training, there are several advantages to use online learning. First, it is convenient to add new training samples. Every time we add new samples, the computational complexity is roughly the same because we only need to update the models using new samples. Second, huge data sets can be easily handled by sequential processing. Third, online learning algorithms are suitable for real time online training. These properties are quite desirable for video-based face recognition.

In this paper, we propose an online learning algorithm which can learn appearance manifolds without a pre-trained model. Similar to [7], several eigenspace models are used to represent sub-manifolds. However, in our method, the eigenspace models are learnt completely online, while Lee and Kriegman’s method [7] learnt appearance manifolds based on a pre-trained model.

The remainder of this paper is organized as follows. In Section 2, we will introduce some previous work on video-based face recognition. In Section 3, the proposed method will be described in detail. Some experimental results will be shown in Section 4. Finally, conclusions will be drawn in Section 5.

## 2 Previous Work

General reviews of recent face recognition literature can be found in [18, 4]. Here, we only briefly review the literature which deals specifically with video-based face recognition.

In [17], Mutual Subspace Method (MSM) is proposed in which the similarity is defined by the angle between the input and the reference subspaces. Shakhnarovich *et al.* [12] defines the similarity as the KL divergence between the estimated density of two image sequences. Krüeger and Zhou [5] proposed an exemplar-based method which selects representative face images from training face videos. These representative face images are used for tracking and recognition. Liu and Chen [10] proposed to use adaptive Hidden Markov Models (HMM) which can learn temporal dynamics from each sequence. In [14] and [9], frame synchronization was used to align different sequences. They also exploit audio information in videos. In [1], a generic shape-illumination manifold is learnt off-line using Probabilistic PCA (PPCA) [15]. A generic framework for both tracking and recognition is proposed by Zhou *et al.* [19, 20]. They estimate the joint posterior probability distribution of motion vector and identity variable at each time instant and then propagate it to the next time instant.

Lee *et al.* [6] proposed a method using probabilistic appearance manifolds to model and recognize faces in video sequences. An appearance manifold is approximated by several eigenspaces and a transition matrix learnt from an image sequence. They also proposed an online learning algorithm for constructing a probabilistic appearance manifold [7]. In this online learning algorithm, an appearance model is incrementally learnt online using a prior generic model and successive frames from the video. Both the generic and individual appearances are represented as an appearance manifold that is approximated by several eigenspaces and the connectivity between them. One obvious limitation is that their method requires a generic prior model which should be learnt off-line.

In [8], an online learning method using eigenspace merging and splitting is proposed. For each subject, a fixed number of eigenspace models and a transition

matrix are learnt to approximately construct the face appearance manifold. Then, these eigenspace models are used for video-based face recognition.

Different from our previous work [8], we avoid using eigenspace merging, eigenspace splitting and transition matrix. There are several reasons. First, eigenspace splitting may not be faithful to the original samples. Second, for eigenspace merging, two eigenspaces which correspond to different poses may be merged into one eigenspace. Third, the transition probability between two eigenspaces relies too much on the training samples. We also avoid setting a fixed number of eigenspace models to be learnt. Instead, we just set an upper bound of that number, which makes the process of learning more flexible.

### 3 Multi-Eigenspace Learning

In Section 3.1, we give a description of the eigenspace models to be learnt online. In Section 3.2, a framework of online appearance model learning method called Multi-Eigenspace Learning is presented. In our method, eigenspace update using IPCA (Incremental Principal Component Analysis) is a critical part and is discussed in Section 3.3. In Section 3.4, we discuss a desirable property of the algorithm of Multi-Eigenspace Learning.

#### 3.1 Model Description

Given the training face video stream of each subject in the data set, we aim to construct up to  $K$  eigenspaces,  $\Omega^{(1)}, \Omega^{(2)}, \dots, \Omega^{(k)}$  ( $k \leq K$ ,  $K$  is decided empirically), to approximately represent the appearance manifold of that subject. There are four parameters for each eigenspace [3]:

$$\Omega = (\bar{\mathbf{x}}, \mathbf{U}, \mathbf{\Lambda}, N), \quad (1)$$

where  $\bar{\mathbf{x}}$  is the center of the eigenspace.  $\mathbf{U}$  is a matrix whose columns are orthonormal bases of the eigenspace, namely eigenvectors.  $\mathbf{\Lambda}$  is a diagonal matrix. Elements along the diagonal are variances for each principal axis, namely eigenvalues. They are arranged in descending order.  $N$  is the number of samples to construct the eigenspace.

How to learn these eigenspace models online is what we focus on in this paper. Our method will be described in Section 3.2.

#### 3.2 The Algorithm of Multi-Eigenspace Learning

For each subject, suppose there are  $n$  frames which arrive sequentially. Initially,  $k$  is assigned as 1 and the center of  $\Omega^{(1)}$  is assigned as the first frame. Then for each incoming frame  $\mathbf{x}_t$ , compute the distance between  $\mathbf{x}_t$  and each eigenspace center. If the smallest distance is not greater than a given threshold, the corresponding eigenspace is updated using IPCA. Otherwise, a new eigenspace is

established and its center is assigned as  $\mathbf{x}_t$ . If  $k = K$ , an eigenspace which is constructed from the least number of samples is discarded before the new eigenspace is established. Our algorithm is summarized as follows.

### Algorithm 1. Multi-Eigenspace Learning

**Input:**

$\{\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n\}$ : a consecutive face image sequence from a subject, where  $\mathbf{x}_i$  is the feature vector extracted from the  $i$ th image. ( $1 \leq i \leq n$ )

$K$ : upper limit of the number of eigenspaces.  $K$  can be decided empirically.

$T$ : a threshold which can be decided empirically.

**Output:**

$\{\Omega^{(i)} | \Omega^{(i)} = (\bar{\mathbf{x}}^{(i)}, \mathbf{U}^{(i)}, \mathbf{\Lambda}^{(i)}, N^{(i)}), i = 1, \dots, k, k \leq K\}$ .

**Method:**

1.  $k \leftarrow 1$
2.  $\Omega^{(1)} \leftarrow (\mathbf{x}_1, \{\}, \{\}, 1)$
3. **for**  $j \leftarrow 2$  to  $n$
4.  $p^* \leftarrow \arg \min_{1 \leq p \leq k} d(\mathbf{x}_j, \Omega^{(p)})$
5. **if**  $d(\mathbf{x}_j, \Omega^{(p^*)}) \leq T$
6.     IPCA( $\Omega^{(p^*)}, \mathbf{x}_j$ )
7. **else if**  $k < K$
8.      $k \leftarrow k + 1$
9.      $\Omega^{(k)} \leftarrow (\mathbf{x}_j, \{\}, \{\}, 1)$
10. **else**
11.      $p^* \leftarrow \arg \min_{1 \leq p \leq K} N^{(p)}$
12.      $\Omega^{(p^*)} \leftarrow (\mathbf{x}_j, \{\}, \{\}, 1)$
13. **end if**
14. **end for**

In Line 4,  $d(\mathbf{x}_j, \Omega^{(p)})$  computes the  $l_1$ -distance between  $\mathbf{x}_j$  and  $\bar{\mathbf{x}}^{(p)}$ . In Line 6,  $\mathbf{x}_j$  is incorporated into  $\Omega^{(p^*)}$  using IPCA. Generally, an eigenspace is more likely to be an outlier if it is constructed from fewer samples. In Lines 11-12, we remove an eigenspace which contains the least number of samples.

In Section 3.3, we will show how IPCA works without knowing both the original samples and the covariance matrix.

### 3.3 IPCA

In this section, we discuss more details about the computation of IPCA. An algorithm of IPCA was proposed by Hall *et al.* [3], but their method is complicated to some extent. A concise method is described as follows [8].

**Algorithm 2. IPCA****Input:**

$\Omega = (\bar{\mathbf{x}}, \mathbf{U}, \mathbf{\Lambda}, N)$ : constructed from  $\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_N$ .

$\mathbf{x}$ : a new sample.

**Output:**

$\Omega' = (\bar{\mathbf{x}}', \mathbf{U}', \mathbf{\Lambda}', N')$ : constructed from  $\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_N, \mathbf{x}$ .

**Method:**

1.  $N' \leftarrow N + 1$ .
2.  $\alpha_1 \leftarrow N/N', \alpha_2 \leftarrow 1 - \alpha_1$ .
3.  $\bar{\mathbf{x}}' \leftarrow \alpha_1 \bar{\mathbf{x}} + \alpha_2 \mathbf{x}$ .
4. Generate artificial data:  $\mathbf{Y} \leftarrow [\sqrt{\alpha_1} \mathbf{U} \mathbf{\Lambda}^{1/2}, \sqrt{\alpha_1 \alpha_2} (\bar{\mathbf{x}} - \mathbf{x})]$ .
5. Compute the eigenvectors and eigenvalues of  $\mathbf{Y}^T \mathbf{Y}$ :  $\mathbf{Y}^T \mathbf{Y} = \mathbf{V} \mathbf{\Lambda}' \mathbf{V}^T$ .
6.  $\mathbf{U}' \leftarrow \mathbf{Y} \mathbf{V} \mathbf{\Lambda}'^{-1/2}$ .

This IPCA algorithm can be viewed as a special case of the algorithm of eigenspace merging [13]. For eigenspace merging, if one of the two eigenspaces to be merged has zero dimensions, the problem becomes IPCA.

In Step 5 and Step 6, it is enough to retain only a few relatively larger eigenvalues and corresponding eigenvectors.

**3.4 A Desirable Property of Multi-Eigenspace Learning**

**Property (of Algorithm 1).** *Suppose that after the iteration of  $j = h$  ( $h < n$ ),  $K$  eigenspace models are learnt and the ascending permutation of  $N^{(1)}, N^{(2)}, \dots, N^{(K)}$  is  $N_{h1} \leq N_{h2} \leq \dots \leq N_{hK}$ . Then for an arbitrary integer  $l$  subjected to  $h < l \leq n$ , the following inequality holds.*

$$N_{hi} \leq N_{li}, \quad i = 2, 3, \dots, K, \quad (2)$$

where  $N_{l1} \leq N_{l2} \leq \dots \leq N_{lK}$  is the ascending permutation of  $N^{(1)}, N^{(2)}, \dots, N^{(K)}$  after the iteration of  $j = l$ .

*Proof.* To prove this property, we only need to prove that

$$N_{l-1,i} \leq N_{li}, \quad i = 2, 3, \dots, K, \quad h < l \leq n. \quad (3)$$

Note that when  $j \geq h$ ,  $k = K$  holds. So the condition in Line 7 will be false. There are only the following two cases in the iteration of  $j = l$ .

Case 1:  $d(\mathbf{x}_l, \Omega^{(p^*)}) \leq T$ . Line 6 is executed. So the update of  $N^{(i)}$  ( $i = 1, 2, \dots, K$ ) is as follows.

$$N_{li} \leftarrow \begin{cases} N_{l-1,i} + 1 & \text{when } i = p^*, \\ N_{l-1,i} & \text{otherwise.} \end{cases} \quad (4)$$

Case 2:  $d(\mathbf{x}_l, \Omega^{(p^*)}) > T$ . Lines 11-12 are executed. The update of  $N^{(i)}$  ( $i = 1, 2, \dots, K$ ) is as follows.

$$N_{li} \leftarrow \begin{cases} 1 & \text{when } i = 1, \\ N_{l-1,i} & \text{otherwise.} \end{cases} \quad (5)$$

Obviously, (3) holds in both Case 1 and Case 2. From (3), we can get

$$N_{hi} \leq N_{h+1,i} \leq \dots \leq N_{li}, \quad i = 2, 3, \dots, K. \quad (6)$$

□

This property shows that after  $K$  eigenspaces are learnt, each eigenspace generally contains more and more samples except one eigenspace which contains the least number of samples. This is a desirable property for online learning.

Moreover, after  $K$  eigenspaces are learnt, the algorithm tends to establish new eigenspaces while discarding old eigenspaces which contain the least number of samples. This conforms with human experience.

## 4 Experimental Results

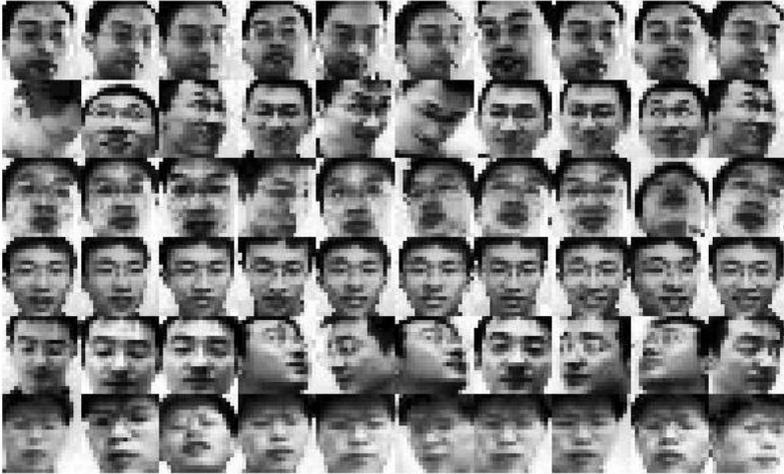
We conduct some experiments on a 36-subject face video data set which bears large pose variation collected by our lab. In this data set, there are 36 videos which correspond to 36 subjects respectively. These video sequences are captured indoors at 30 frames per second. The sequences contain large 2-D (in plane) and 3-D (out of plane) head rotation, with slight expression and illumination changes. In each video sequence, the number of frames ranges from 236 to 1270. The faces in the videos are cropped automatically using a boosted cascade face detector [16], or cropped manually when the detection results are not good enough. Then, all the cropped images are transformed into gray-level images and resized to a standard size of  $20 \times 20$ . Then a histogram equalization step is applied to eliminate the illumination impact to some extent. Some samples are shown in Figure 1.

We use the first half of each video sequence, ranging from 118 to 635 frames, for online learning. The second half of each video sequence is used for the recognition task. We do some experiments using the following two methods.

-**EMS + transition**, Eigenspace Merging and Splitting with consideration of transition matrix [8].

-**Multi-Eigenspace Learning**, the proposed method.

$K$ , namely the number of eigenspace models, is a parameter which is difficult to optimize without experiments. Take this into consideration, we do experiments for  $K = 6, 7, 8, 9$  respectively to evaluate the two methods. The test sets are constructed by randomly sampling from the second half of each video sequence for 10 times with each set containing 50 samples [2]. For each sample in the test



**Fig. 1.** Typical samples of the videos used in our experiments. The images in each row come from a different video sequence.

**Table 1.** Average recognition rates (%) of different methods when choosing  $K = 6, 7, 8, 9$  respectively

Method	$K$			
	6	7	8	9
EMS + transition [8]	90.0	91.9	96.4	91.4
Multi-Eigenspace Learning	97.5	96.7	98.1	96.4

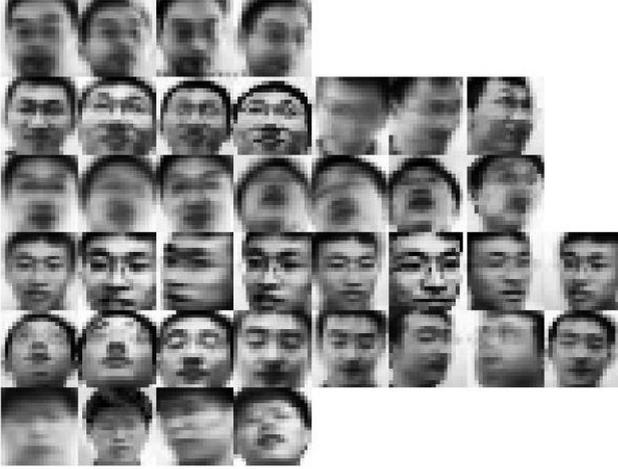
set, we compute the likelihood probability of each eigenspace model and find the maximum to make a classification. The likelihood probability can be computed using the following formula [11]:

$$p(\mathbf{x}_t | \Omega^{(j)}) = \frac{\exp(-\frac{1}{2} \sum_{i=1}^q \frac{y_i^2}{\lambda_i})}{(2\pi)^{q/2} \prod_{i=1}^q \lambda_i^{1/2}} \cdot \frac{\exp(-\frac{\epsilon^2(\mathbf{x})}{2\rho})}{(2\pi\rho)^{(m-q)/2}}, \quad (7)$$

where  $q$  is the dimension of Eigenspace  $\Omega^{(j)}$ , and  $[y_1, y_2, \dots, y_q]^T$  is the projection from  $\mathbf{x}_t$  to  $\Omega^{(j)}$ .  $\epsilon^2(\mathbf{x})$  is the Euclidean distance from  $\mathbf{x}_t$  to  $\Omega^{(j)}$ , and  $\lambda_i$  is the  $i$ th eigenvector in  $\Omega^{(j)}$ .  $m$  is the dimension of the original feature space. The parameter  $\rho$  is chosen as  $0.3\lambda_q$  empirically.

For each test set, majority voting is used to make a final decision. The average recognition rates over all runs are shown in Table 1.

The results show that the proposed method outperforms EMS + transition. This conforms with the analysis at the last paragraph of Section 2. Figure 2 shows some eigenspace centers learnt using Multi-Eigenspace Learning when choosing  $K = 8$ .



**Fig. 2.** Some eigenspace centers learnt using Multi-Eigenspace Learning when choosing  $K = 8$ .

We also implemented the Probabilistic Manifold online learning algorithm [7] for comparison. This algorithm starts with a generic manifold which is trained off-line in a batch mode. There are two steps in the online learning process. The first step is to find the pose manifold to which the current image belongs with the highest probability. The second step is to update the current appearance manifold using IPCA. The result from the first step is used to find a set of pre-training images that are expected to appear similar to the current subject in other poses. Then all of the other eigenspaces in the current appearance manifold are updated with synthetic images.

We use 15 of the 36 sequences for pre-training. The face images in the pre-training sequences are manually classified into 5 different pose clusters [7]. Then a 10-D pose subspace is computed from each cluster using PCA. The remain 21 video sequences are used for online learning and recognition. The first half of each sequence is used for online learning, and the second half is used for recognition. Our proposed algorithm also runs on these 21 sequences for comparison. The average recognition rates and the processing time are shown in Table 2. The recognition rate of Multi-Eigenspace Learning in Table 2 is different from that in Table 1. The reason is that the results in Table 2 are obtained from a 21-subject data set, while the results in Table 1 are obtained from a 36-subject data set.

In Table 2, we can see that our proposed algorithm gives higher recognition rate than that of Probabilistic Manifold algorithm [7], while the processing time is much shorter than the latter. One main reason is that the Probabilistic Manifold algorithm use only 5 pose manifolds which are not flexible enough to represent different poses. Moreover, updating eigenspaces with synthetic face images significantly increases the time complexity. In contrast, the proposed algorithm is more flexible to learn representative eigenspace models online.

**Table 2.** Comparison of the Probabilistic Manifold algorithm and our proposed algorithm. For Multi-Eigenspace Learning, we choose  $K = 8$ . Because Multi-Eigenspace Learning needs no pre-training, the corresponding blank is left empty.

Method	Probabilistic Manifold [7]	Multi-Eigenspace Learning
Recognition rates	92.4%	96.2%
Online learning	34.3s	7.1s
Pre-training	77.3s	-

## 5 Conclusions

In this paper we have presented a novel online appearance model learning algorithm for video-based face recognition. For each subject, we build several eigenspace models to approximately construct the face appearance manifold. The number of eigenspace models corresponding to each subject is variable. The eigenspace models are dynamically adjusted using IPCA. In the process of online learning, each eigenspace generally contains more and more samples except one eigenspace which contains the least number of samples. The learnt eigenspace models can be used for video-based face recognition. In our experiments, the average recognition rate can reach as high as 98.1%.

**Acknowledgments.** This work was supported by Program of New Century Excellent Talents in University, National Natural Science Foundation of China (No. 60575003, 60332010, 60335010, 60121302, 60275003, 69825105, 60605008), Joint Project supported by National Science Foundation of China and Royal Society of UK (60710059), Hi-Tech Research and Development Program of China (2006AA01Z133, 2006AA01Z193), the National Basic Research Program (Grant No.2004CB318110) and the Chinese Academy of Sciences.

## References

1. Arandjelovic, O., Cipolla, R.: Face recognition from video using the generic shape-illumination manifold. In: Proc. European Conf. on Computer Vision, vol. 3594, pp. 27–40 (2006)
2. Fan, W., Yeung, D.-Y.: Face recognition with image sets using hierarchically extracted exemplars from appearance manifolds. In: Proceedings of the 7th International Conf. on Automatic Face and Gesture Recognition, pp. 177–182 (2006)
3. Hall, P.M., Marshall, D., Martin, R.R.: Incremental eigenanalysis for classification. In: The British Machine Vision Conference, pp. 286–295 (1998)
4. Kong, S.G., Heo, J., Abidi, B.R., Paik, J., Abidi, M.A.: Recent advances in visual and infrared face recognition - a review. *Computer Vision and Image Understanding* 97, 103–135 (2005)
5. Krueger, V., Zhou, S.: Exemplars-based face recognition from video. In: Proc. European Conf. on Computer Vision, vol. 4, pp. 732–746 (2002)

6. Lee, K.-C., Ho, J., Yang, M.-H., Kriegman, D.: Videobased face recognition using probabilistic appearance manifolds. In: proceedings of the CVPR, pp. 313–320 (2003)
7. Lee, K.-C., Kriegman, D.: Online learning of probabilistic appearance manifolds for video-based recognition and tracking. In: Proceedings of the CVPR, vol. 1, pp. 852–859 (2005)
8. Liu, L., Wang, Y., Tan, T.: Online appearance model learning for video-based face recognition. In: IEEE Computer Society Workshop on Biometrics. IEEE Computer Society Press, Los Alamitos (2007)
9. Liu, W., Li, Z., Tang, X.: Spatio-temporal embedding for statistical face recognition from video. In: Leonardis, A., Bischof, H., Pinz, A. (eds.) ECCV 2006. LNCS, vol. 3952, pp. 374–388. Springer, Heidelberg (2006)
10. Liu, X., Chen, T.: Video-based face recognition using adaptive hidden markov models. In: Proceedings of the CVPR, pp. 340–345 (2003)
11. Moghaddam, B., Pentland, A.: Probabilistic visual learning for object recognition. IEEE Transactions on Pattern Analysis and Machine Intelligence 19(7), 696–710 (1997)
12. Shakhnarovich, G., Fisher, J.W., Darrell, T.: Face recognition from long-term observations. In: Proc. European Conf. on Computer Vision, vol. 3, pp. 851–865 (2002)
13. Skarbek, W.: Merging subspace models for face recognition. In: Proceedings of the CAIP, pp. 606–613 (2003)
14. Tang, X., Li, Z.: Frame synchronization and multi-level subspace analysis for video based face recognition. In: Proceedings of the CVPR, pp. 902–907 (2004)
15. Tipping, M.E., Bishop, C.M.: Mixtures of probabilistic principal component analysers. Neural Computation 11(2), 443–482 (1999)
16. Viola, P., Jones, M.: Rapid object detection using a boosted cascade of simple features. In: Proceedings of the CVPR, vol. 1, pp. 511–518 (2001)
17. Yamaguchi, O., Fukui, K., ichi Maeda, K.: Face recognition using temporal image sequence. In: Proceedings of International Conf. on Automatic Face and Gesture Recognition, pp. 318–323 (1998)
18. Zhao, W., Chellappa, R., Phillips, P., Rosenfeld, A.: Face recognition: A literature survey. ACM Computing Surveys 35(4), 399–458 (2003)
19. Zhou, S., Krueger, V., Chellappa, R.: Probabilistic recognition of human faces from video. Computer Vision and Image Understanding 91, 214–245 (2003)
20. Zhou, S.K., Chellappa, R.: Probabilistic identity characterization for face recognition. In: Proceedings of the CVPR, vol. 2, pp. 805–812 (2004)