Expert Systems with Applications 42 (2015) 973–979

Contents lists available at ScienceDirect

Expert Systems with Applications

journal homepage: www.elsevier.com/locate/eswa

Face recognition from a single registered image for conference socializing

Yu Zhao [a], Yan Liu [a,*], Yang Liu [a], Shenghua Zhong [b], Kien A. Hua [c]

[a] Department of Computing, The Hong Kong Polytechnic University, Hung Hom, Kowloon, Hong Kong, China
[b] Department of Psychological and Brain Sciences, Johns Hopkins University, Baltimore, MD 21218-2686, USA
[c] Department of Electrical Engineering and Computer Science, University of Central Florida, Orlando, FL 32816-2362, USA

ARTICLE INFO

Article history:
Available online 8 September 2014

Keywords:
Conference socializing
Face recognition
Single registered image
Large pose variation

ABSTRACT

Scientific conferences are primary venues for connecting with and forming relationships with fellow researchers and scientists. Thus, over the course of a conference participants often take advantage of the many opportunities to network. In this setting, it is desirable to quickly recognize the identity of the persons we see and wish to meet. In particular, it could be embarrassing to not recognize a prominent researcher. In this paper, we investigate a novel face recognition framework that is applicable to conference socialization scenarios. In the proposed framework, only frontal images are used as training images; and face recognition is possible from an arbitrary view of a subject. Our system prototype assumes that the conference participants have uploaded a frontal photo during the registration process. At the conference, the identity of a person can be recognized from a picture, taken from an arbitrary angle with a standard mobile phone. Our experimental results indicate that the proposed framework is robust to possible large pose variations between the non-frontal image captured impromptu and the training image of the same person. Experiments based upon standard face dataset and real conference socializing datasets are conducted to test the effectiveness of the proposed techniques.

1. Introduction

As one of the most ubiquitous activities in the modern society, socializing becomes an important and necessary component in human's daily life. For most of the researchers, the academic or scientific conferences are the primary venues for connecting with and forming relationships with fellow researchers and scientists. Thus, over the course of a conference participants often take advantage of the many opportunities to network. In this setting, it is desirable to quickly recognize the identity of the persons we see and wish to meet. In particular, it could be embarrassing to not recognizing a prominent researcher.

Current handheld devices such as mobile phones can help us to take the face pictures of the conference participants. Then the face recognition algorithms could be applied to the face pictures to classify the identities of these participants. As an active research topic in machine learning, human face recognition plays an increasingly important role in a wide range of application, such as criminal identification, credit card verification, and surveillance systems, and many face recognition algorithms have therefore been proposed (Abdullah et al., 2014; Belhumeur, Hespanha, & Kriegman, 1997; Gumus, Kilic, Sertbas, & Ucan, 2010; He, Yan, Hu, Niyogi, & Zhang, 2005; Perlibakas, 2004; Turk & Pentland, 1991; Vignolo, Milone, & Scharcanski, 2013; Xu, Song, Feng, & Zhao, 2010). Although these algorithms have reported good performance in well controlled experiment environments, most of them do not work well under the conference socializing circumstance because of the following reasons. First, conventional face recognition methods usually assume that there are multiple images for each person in the training phase. In a conference situation, however, we can take only one frontal face picture for each participant in registration, which means that we have only one training image for each person. As a consequence, many existing methods (Belhumeur et al., 1997; He et al., 2005) cannot be directly applied due to the lack of samples to calculate the within-class scatter. Second, when we meet the participants during the conference, it is generally not convenient to ask them standing at a certain place for us to take their frontal face pictures. Therefore, the pictures that we can take may have large variations on pose, illumination condition, or expression, which inevitably limits the performance of variance-based or distance-based methods (Perlibakas, 2004; Turk & Pentland, 1991).

* Corresponding author.
E-mail addresses: csyuzhao@comp.polyu.edu.hk (Y. Zhao), csyliu@comp.polyu.edu.hk (Y. Liu), csygliu@comp.polyu.edu.hk (Y. Liu), szhong2@jhu.edu (S. Zhong), kienhua@cs.ucf.edu (K.A. Hua).

http://dx.doi.org/10.1016/j.eswa.2014.08.016
0957-4174/

Recently, some algorithms have been presented to tackle above problems. Zhang, Chen, and Zhou (2005) developed a method called singular value decomposition-based linear discriminant analysis (SVD-LDA), which aims to enrich the information of eigenspace learned by the single training image per person. However, its performance is still limited by the large pose variations between training and test images. Blanz and Romdhani (2002) constructed a 3-D face model for each person using only one image, including parameters representing the pose and illumination. Based on this model, face with different poses or illumination conditions could be estimated by using corresponding parameter settings. However, this kind of methods is computationally expensive, and thus might not be suitable for the circumstance of conference socializing, which generally requires fast processing. Prince, Elder, Warrell, and Felisberti (2008) proposed a linear statistical model that seeks to map the data to a hidden space, in which the representations of different individuals are far away from each others. This method is much faster than those 3D model-based methods, but it requires manually selecting more than twenty feature points for each image, which is a very heavy workload for the users. Lu, Tan, and Wang (2013) introduced a discriminative multi-manifold analysis (DMMA) model, which segments each of the original training images into non-overlapping sub-images and then conducts face recognition using these sub-images as the training data. Although this method partially alleviates the problem of single training image per person, the assumption that non-overlapping sub-images reside in a low-dimensional smooth manifold is too strong and has not been convincingly tested. More details of face recognition technologies for single training image per person and with large pose variations could be found in Tan, Chen, Zhou, and Zhang (2006) and Zhang and Gao (2009), respectively.

In order to recognize the identity of a person under the conference socializing circumstance, the system should be fully automatic, capable of making full use of the only registration picture for each person in training, robust to large pose variations between training and test images, and with fast processing speed and high recognition accuracy. However, each of the aforementioned methods satisfies only one or two of these requirements, and thus might not work well in the circumstance of conference socializing. In this paper, we propose a novel face recognition framework for the application of conference socializing. Given a face image, the proposed framework first automatically detects important local feature points by template matching. A Gaussian filter then acts on the local feature areas in order to emphasize the important parts and weaken the unimportant regions. In the third step, we utilize a statistical model to learn the discriminative feature in the hidden space for each individual. Finally, the recognition decision is made by choosing the class with maximum posterior probability.

From the perspective of problem formulation, our framework can be classified into the category of statistical model. However, unlike the previous statistical model in Prince et al. (2008) that requires the users to select many feature points for each image manually, our method is fully automatic in detecting the local feature points, which largely alleviates the users' workload. Moreover, our framework has one more important step of feature area smoothing, which is able to emphasize the important regions and weaken the unimportant ones simultaneously, and thus improves the performance. Fig. 1 shows the procedure of the proposed framework.

It is worthwhile to highlight several aspects of the proposed approach here. First, the proposed framework automatically detects the local feature points and its formulation contains only a small number of parameters, it is therefore suitable for the circumstance of conference socializing, which generally requires fast processing. Second, since our method utilizes template matching to determine the important feature regions on faces with different angles, it is relatively robust when there are large pose variations between training and test images. Third, the proposed model does not require any special distribution of the training data, and thus is flexible to various kinds of input data. Besides the application in conference socializing, the proposed face recognition framework is directly applicable in many practical scenarios where only one training sample in each class is available, such as the ID card identification. Moreover, our model is within-class variance-tolerance. So it can be used in the tasks where the semantically similar samples and targets might appear quite different, such as image retrieval in search engines and video tracking in surveillance equipments.

The rest of this paper is organized as follows. The proposed framework is presented in Section 2. Section 3 reports the experimental results of the proposed framework on both standard face dataset and real-world conference socializing datasets. We conclude our work in Section 4.

2. Proposed framework

2.1. Feature point detection

Given a face image, the proposed framework first automatically detects important local feature points. In order to make the entire framework computationally efficient, we simply detect five feature points: the left eye, the right eye, the nose, the left corner of the mouth, and the right corner of the mouth. For each pose, we select the image that is closest to the mean of all the images belonging to this pose as the template image. Given a $w \times h$ template $\mathbf{T}$ of a feature point, we aim to find a location (x, y) on image $\mathbf{I}$ that maximizes the following objective function:

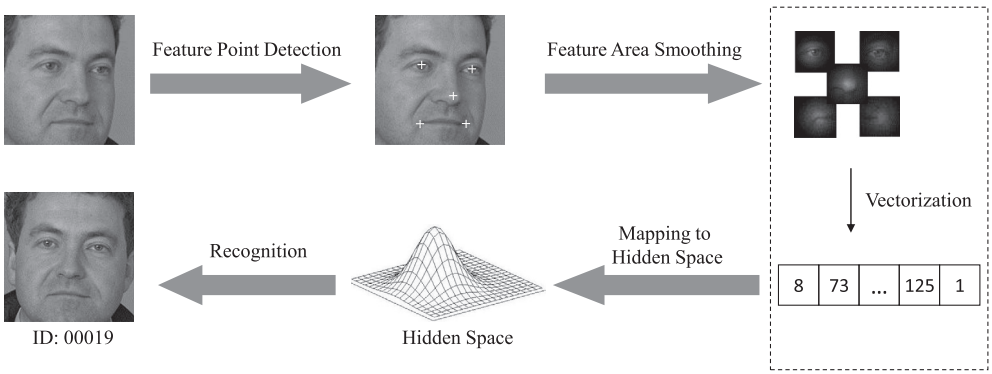

Fig. 1. Procedure of the proposed framework.

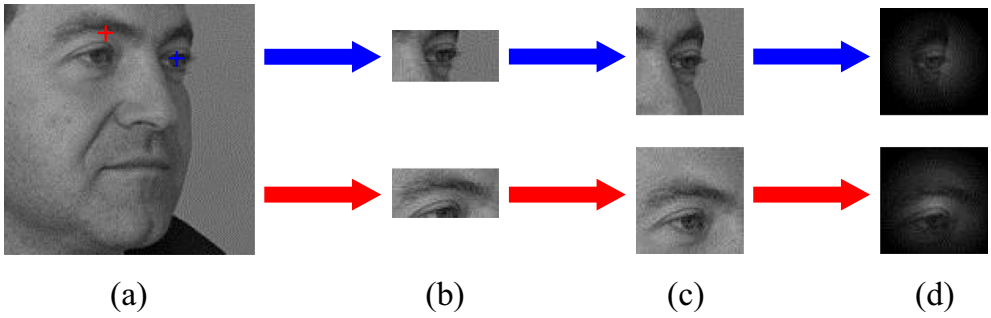

Fig. 2. Procedure of feature area smoothing.

$$\arg\max_{x,y} O(x,y) = Pr(x,y|TrainingSet) \times C(x,y), \tag{1}$$

where (x,y) denotes the location of the left top corner of $\mathbf{T}$, $Pr(x,y|TrainingSet)$ denotes the posterior distribution of (x,y) calculated using the training image set, and $C(x,y)$ is a normalized correlation coefficient defined as follows:

$$C(x,y) = \frac{\sum_{x',y'}(\mathbf{T}'(x',y') \times \mathbf{I}'(x+x',y+y'))}{\sqrt{\sum_{x',y'}\mathbf{T}'(x',y')^2 \times \sum_{x',y'}\mathbf{I}'(x+x',y+y')^2}}, \tag{2}$$

where $\mathbf{T}'(x',y') = \mathbf{T}(x',y') - \overline{\mathbf{T}}$ and $\mathbf{I}'(x+x',y+y') = \mathbf{I}(x+x',y+y') - \overline{\mathbf{I}}$ $(x' = 1,\ldots,w$ and $y' = 1,\ldots,h)$, $\mathbf{T}(x,y)$ and $\mathbf{I}(x,y)$ represent the pixel values at location (x,y) of the template and that of the image, respectively, $\overline{\mathbf{T}}$ and $\overline{\mathbf{I}}$ represent the mean pixel value of $\mathbf{T}$ and $\mathbf{I}$ within the template size.

2.2. Feature area smoothing

For each detected feature point, we locate a corresponding feature area, which takes the feature point as the center. As shown in Fig. 2(a), for each eye, we have detected a feature point. Then we locate the feature areas of the left eye (top of Fig. 2(b)) and the right eye (bottom of Fig. 2(b)), respectively. But if the feature point is not precisely located, some important part will be missing (bottom of Fig. 2(b)). A simple way to overcome this problem is enlarging the feature area as shown in Fig. 2(c), then we can see that the entire eye region has been included into the feature area even the feature point location is not perfectly precise. Under this case, however, some unimportant even useless regions will be introduced into the feature area inevitably. Therefore, we add a Gaussian filter onto the feature area, which aims to emphasize the important parts and weaken the unimportant or useless regions. By doing this, the eye region is well kept and the surroundings are filtered smoothly (Fig. 2(d)).

Let $\mathbf{A}$ be the matrix representation of the selected feature area, we perform the Gaussian filtering as follows:

$$\mathbf{X} = \mathbf{A} \circ \mathbf{G}, \tag{3}$$

where $\circ$ denotes the Hadamard product operation and $\mathbf{G}$ is the Gaussian filtering matrix defined as:

$$\mathbf{G}(k,l) = \frac{1}{2\pi\sigma^2} e^{-\frac{k^2+l^2}{2\sigma^2}}, \tag{4}$$

where k is the number of pixels from the center in the horizontal axis, and l is the number of pixels from the center in the vertical axis. Then we construct the feature vector for each face image by

$$\mathbf{x} = [vec(\mathbf{X}_1)^T \ldots vec(\mathbf{X}_5)^T]^T, \tag{5}$$

where $\mathbf{X}_i$ $(i = 1,\ldots,5)$ denotes the matrix representation of the ith smoothed feature area, and $vec(\mathbf{X})$ denotes the vectorization of the matrix $\mathbf{X}$ formed by stacking the columns of $\mathbf{X}$ into a single column vector.

2.3. Mapping model learning

Given a set of constructed feature vectors, the model assumes that all the representations of the same person with different poses in the feature space can be generated by the same underlying representation from a hidden space. Therefore, the objective of this learning model is to find the mappings between the hidden space and the feature space of different poses. Inspired by Prince et al. (2008), the mapping model can be formulated as follows:

$$\mathbf{x}_{ij} = \mathbf{F}_j\mathbf{h}_i + \mathbf{m}_j + \boldsymbol{\varepsilon}_{ij}, \tag{6}$$

where $\mathbf{x}_{ij}$ $(i = 1,\ldots,I$ and $j = 1,\ldots,J)$ denotes the feature vector of the single image of individual i in the jth pose; $\mathbf{h}_i$ denotes the underlying identity variable of individual i; $\mathbf{F}_j$ and $\mathbf{m}_j$ are the parameters of the mapping function specialized for the jth pose; and $\boldsymbol{\varepsilon}_{ij}$ is a zero-mean multivariate Gaussian noise term with an unknown diagonal covariance matrix Σ_j. Fig. 3 illustrates the idea of mapping model learning.

We can rewrite the model in terms of conditional probabilities:

$$Pr(\mathbf{x}_{ij}|\mathbf{h}_i) = N_{\mathbf{x}}(\mathbf{F}_j\mathbf{h}_i + \mathbf{m}_j, \boldsymbol{\Sigma}_j), \tag{7}$$

$$Pr(\mathbf{h}_i) = N_{\mathbf{h}}(\mathbf{0}, \mathbf{I}), \tag{8}$$

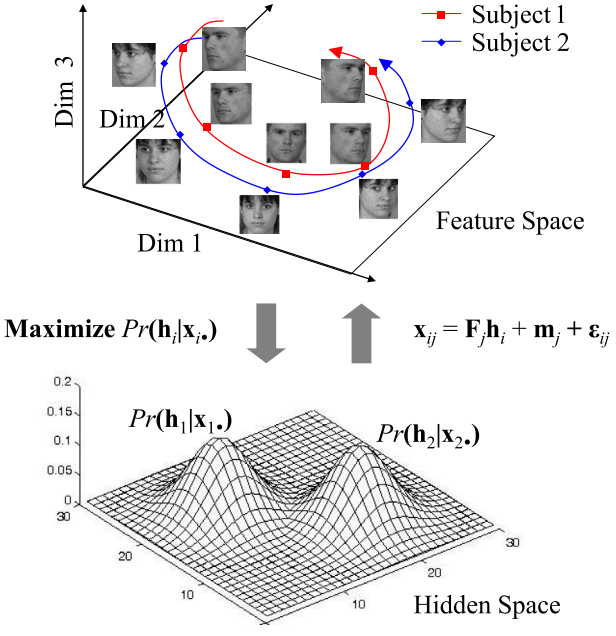

Fig. 3. Illustration of the idea of mapping model learning.

where $N_{\mathbf{x}}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ denotes a Gaussian in $\mathbf{x}$ with mean $\boldsymbol{\mu}$ and covariance $\boldsymbol{\Sigma}$. Here we assume that the prior of $\mathbf{h}_i$ is a zero-mean Gaussian with identity covariance $\mathbf{I}$.

The objective now is to find the parameters $\theta_j = \{\mathbf{F}_j, \mathbf{m}_j, \boldsymbol{\Sigma}_j\}$ $(j = 1, \ldots, J)$ that maximize the joint likelihood $Pr(\mathbf{x}, \mathbf{h}|\boldsymbol{\theta})$. Since both $\mathbf{h}$ and $\boldsymbol{\theta}$ are unknown and they are closely related, we utilize an iterative strategy to update $\mathbf{h}$ and $\boldsymbol{\theta}$ by maximizing the following:

$$Q(\theta_t, \theta_{t-1}) = \sum_{i=1}^{I} \int Pr(\mathbf{h}_i|\mathbf{x}_{i\cdot}, \theta_{t-1}) \times \left(\sum_{j=1}^{J} \log Pr(\mathbf{x}_{ij}|\mathbf{h}_i, \theta_t) + \log Pr(\mathbf{h}_i) \right) d\mathbf{h}_i, \tag{9}$$

where t is the number of iteration and $\mathbf{x}_{i\cdot}$ denotes all the feature vectors with different poses of the individual i.

In the tth iteration, we first fix the parameter set $\theta = \theta_{t-1}$ and estimate the identity vectors $\mathbf{h}_i$ $(i = 1, \ldots, I)$. The posterior distribution of $\mathbf{h}_i$ can be calculated as follows:

$$\begin{aligned} Pr(\mathbf{h}_i|\mathbf{x}_{i\cdot}, \theta) &= \frac{Pr(\mathbf{x}_{i\cdot}|\mathbf{h}_i, \theta) \times Pr(\mathbf{h}_i)}{\int Pr(\mathbf{x}_{i\cdot}|\mathbf{h}_i, \theta) \times Pr(\mathbf{h}_i) d\mathbf{h}_i} \\ &= \frac{\left(\prod_{j=1}^{J} Pr(\mathbf{x}_{ij}|\mathbf{h}_i, \theta)\right) \times Pr(\mathbf{h}_i)}{\int \left(\prod_{j=1}^{J} Pr(\mathbf{x}_{ij}|\mathbf{h}_i, \theta)\right) \times Pr(\mathbf{h}_i) d\mathbf{h}_i} \\ &= \frac{\left(\prod_{j=1}^{J} N_{\mathbf{x}}(\mathbf{F}_j\mathbf{h}_i + \mathbf{m}_j, \boldsymbol{\Sigma}_j)\right) \times N_{\mathbf{h}}(\mathbf{0}, \mathbf{I})}{\int \left(\prod_{j=1}^{J} N_{\mathbf{x}}(\mathbf{F}_j\mathbf{h}_i + \mathbf{m}_j, \boldsymbol{\Sigma}_j)\right) \times N_{\mathbf{h}}(\mathbf{0}, \mathbf{I}) d\mathbf{h}_i}. \end{aligned} \tag{10}$$

The first equality results from the Bayes' rule while the second one is based on the independent and identically distributed (*i.i.d.*) assumption of the feature vectors from the same person. Therefore, $Pr(\mathbf{h}_i|\mathbf{x}_{i\cdot}, \theta)$ is normally distributed and its first two moments can be represented as follows:

$$E(\mathbf{h}_i|\mathbf{x}_{i\cdot}) = \left(\mathbf{I} + \sum_{j=1}^{J} \mathbf{F}_j^T \boldsymbol{\Sigma}_j^{-1} \mathbf{F}_j\right)^{-1} \cdot \left(\sum_{j=1}^{J} \mathbf{F}_j^T \boldsymbol{\Sigma}_j^{-1} (\mathbf{x}_{ij} - \mathbf{m}_j)\right), \tag{11}$$

$$E(\mathbf{h}_i\mathbf{h}_i^T|\mathbf{x}_{i\cdot}) = \left(\mathbf{I} + \sum_{j=1}^{J} \mathbf{F}_j^T \boldsymbol{\Sigma}_j^{-1} \mathbf{F}_j\right)^{-1} + E(\mathbf{h}_i|\mathbf{x}_{i\cdot}) E(\mathbf{h}_i|\mathbf{x}_{i\cdot})^T. \tag{12}$$

In the second step of the tth iteration, we fix $\mathbf{h}_i$ and find the θ that maximizes $Q(\theta_t, \theta_{t-1})$ defined in Eq. (9). Let $\widetilde{\mathbf{F}}_j = [\mathbf{F}_j \mathbf{m}_j]$ and $\widetilde{\mathbf{h}}_i = [\mathbf{h}_i^T 1]^T$, then we have:

$$\frac{\partial Q}{\partial \widetilde{\mathbf{F}}_j} = -\sum_{i=1}^{I} \int Pr(\mathbf{h}_i|\mathbf{x}_{i\cdot}, \theta) \cdot \left(\boldsymbol{\Sigma}_j^{-1} (\mathbf{x}_{ij} - \widetilde{\mathbf{F}}_j \mathbf{h}_i) \mathbf{h}_i^T\right) d\mathbf{h}_i, \tag{13}$$

$$\frac{\partial Q}{\partial \boldsymbol{\Sigma}_j^{-1}} = \frac{1}{2} \sum_{i=1}^{I} \int Pr(\mathbf{h}_i|\mathbf{x}_{i\cdot}, \theta) \cdot \left(\boldsymbol{\Sigma}_j - (\mathbf{x}_{ij} - \widetilde{\mathbf{F}}_j \mathbf{h}_i)(\mathbf{x}_{ij} - \widetilde{\mathbf{F}}_j \mathbf{h}_i)^T\right) d\mathbf{h}_i. \tag{14}$$

Let $\partial Q / \partial \widetilde{\mathbf{F}}_j = 0$ and $\partial Q / \partial \boldsymbol{\Sigma}_j^{-1} = 0$, we obtain:

$$\widetilde{\mathbf{F}}_j = \left(\sum_{i=1}^{I} \mathbf{x}_{ij} E(\mathbf{h}_i|\mathbf{x}_{i\cdot})^T\right) \cdot \left(\sum_{i=1}^{I} E(\mathbf{h}_i\mathbf{h}_i^T|\mathbf{x}_{i\cdot})\right)^{-1}, \tag{15}$$

$$\boldsymbol{\Sigma}_j = \frac{1}{I} \sum_{i=1}^{I} \left(\mathbf{x}_{ij}\mathbf{x}_{ij}^T - \widetilde{\mathbf{F}}_j E(\mathbf{h}_i|\mathbf{x}_{i\cdot}) \mathbf{x}_{ij}^T\right). \tag{16}$$

Above procedure is repeated until convergence.

2.4. Recognition

After we have learned the parameters $\theta_j = \{\mathbf{F}_j, \mathbf{m}_j, \boldsymbol{\Sigma}_j\}$ $(j = 1, \ldots, J)$ of the mapping model, we can use it for face recognition. Given N feature vectors $\mathbf{x}_1, \ldots, \mathbf{x}_N$, where $\mathbf{x}_n$ denotes the representation of frontal face image of the nth person, the objective of our recognition task is to correctly assign each test image $\mathbf{x}_t$ to one of these N classes. To achieve this goal, we aim to find its class label L_t which maximizes the following objective function:

$$\{L_t, j\} = \arg\max_{n,j} Pr(L_t = n|\mathbf{x}_1, \ldots, \mathbf{x}_N, \mathbf{x}_t, \theta_j), \tag{17}$$

where $Pr(L_t = n|\mathbf{x}_1, \ldots, \mathbf{x}_N, \mathbf{x}_t, \theta_j)$ represents the posterior probability that $\mathbf{x}_t$ belongs to the nth class given the frontal face images and the parameter set θ_j. According to the Bayes' rule, we have:

$$\begin{aligned} &Pr(L_t = n|\mathbf{x}_1, \ldots, \mathbf{x}_N, \mathbf{x}_t, \theta_j) \\ &\quad = \frac{Pr(\mathbf{x}_1, \ldots, \mathbf{x}_N, \mathbf{x}_t|L_t = n, \theta_j) \cdot Pr(L_t = n)}{\sum_{n=1}^{N} Pr(\mathbf{x}_1, \ldots, \mathbf{x}_N, \mathbf{x}_t|L_t = n, \theta_j) \cdot Pr(L_t = n)}. \end{aligned} \tag{18}$$

Assume that $Pr(L_t = 1) = \ldots = Pr(L_t = N) = 1/N$, then maximizing the objective function in Eq. (17) is equal to maximizing the following:

$$L_t = \arg\max_{n} Pr(\mathbf{x}_1, \ldots, \mathbf{x}_N, \mathbf{x}_t|L_t = n, \theta_j), \tag{19}$$

where $Pr(\mathbf{x}_1, \ldots, \mathbf{x}_N, \mathbf{x}_t|L_t = n, \theta_j)$ can be calculated as follows:

$$\begin{aligned} &Pr(\mathbf{x}_1, \ldots, \mathbf{x}_N, \mathbf{x}_t|L_t = n, \theta_j) \\ &\quad = \int \cdots \int Pr(\mathbf{x}_{1j}, \ldots, \mathbf{x}_{Nj}, \mathbf{x}_{tj}, \mathbf{h}_1, \ldots, \mathbf{h}_N|\mathbf{h}_t = \mathbf{h}_n) d\mathbf{h}_1 \cdots d\mathbf{h}_N \\ &\quad = \int Pr(\mathbf{x}_{nj}, \mathbf{x}_{tj}, \mathbf{h}_n) d\mathbf{h}_n \cdot \prod_{i=1, i\neq n}^{N} \int Pr(\mathbf{x}_{ij}, \mathbf{h}_i) d\mathbf{h}_i \\ &\quad = \int Pr(\mathbf{x}_{nj}|\mathbf{h}_n) Pr(\mathbf{x}_{tj}|\mathbf{h}_n) Pr(\mathbf{h}_n) d\mathbf{h}_n \cdot \prod_{i=1, i\neq n}^{N} \int Pr(\mathbf{x}_{ij}|\mathbf{h}_i) Pr(\mathbf{h}_i) d\mathbf{h}_i. \end{aligned} \tag{20}$$

Combining Eqs. (7), (8), and (20), and the learned parameter set θ, we can finally obtain the class label L_t that maximizes the objective function in Eq. (17).

3. Experimental results

In this section, we demonstrate the performance of the proposed framework on three datasets: the FERET dataset (Phillips, Moon, Rizvi, & Rauss, 2000) and two self-collected datasets composed of images from real conference socializing environments.

3.1. Experiments on FERET dataset

A subset of FERET dataset is used in our experiments, which consists of a single frontal gallery, and four non-frontal probe sets taken at increasing azimuthal angles (labeled from 'ba' to 'bi') for 200 subjects. For the proposed framework, we randomly choose images of 100 subjects for mapping model learning, and the images of the rest 100 subjects are used for recognition performance evaluation. For each image, the facial part is manually cropped, aligned and resized to 160×160 pixels according to eyes' positions. Due to the symmetry of angles, we only choose the probe images with angle: 'bb', 'bc', 'bd', and 'be' ($+60°$, $+40°$, $+25°$, and $+15°$ respectively).

3.1.1. Automatic feature point detection

In the first experiment, we show the result of automatic feature point detection with different poses. The upper row in each sub image of Fig. 4 shows the automatically detected feature points while the lower row is the ground truth labeled by human. We can see that most of the feature points are correctly detected even under the case of large pose variation. There are two possible reasons. The first is that the feature points we aim to detect are representative and distinguishable points on the face. The second is that we jointly optimize two important items in Eq. (1), which narrows the scope of the template searching.

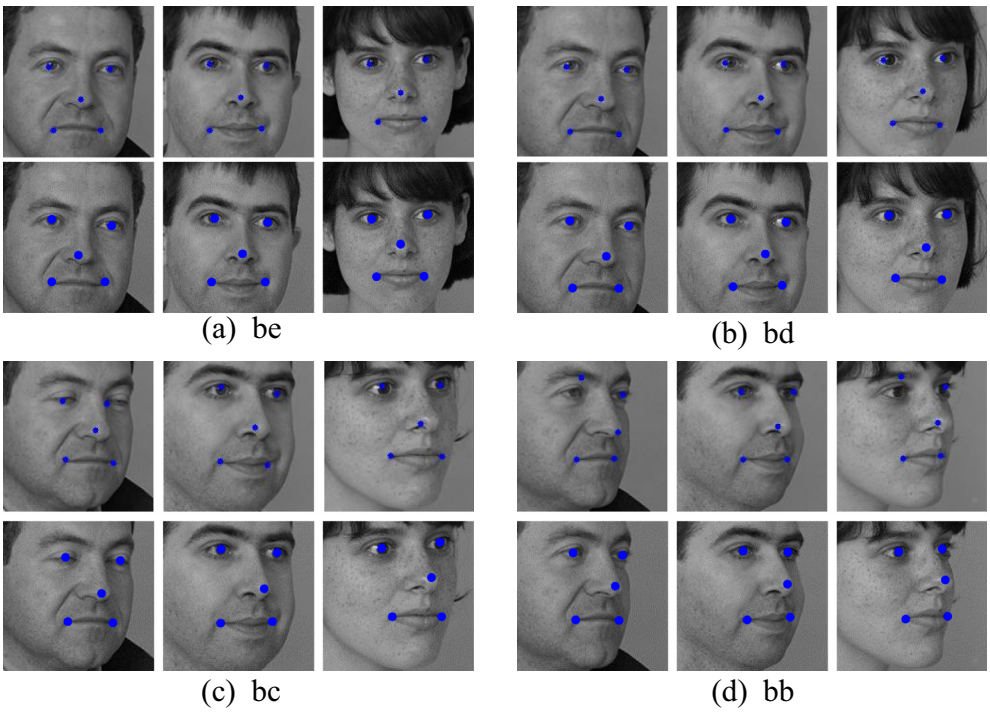

Fig. 4. Results of automatic feature point detection.

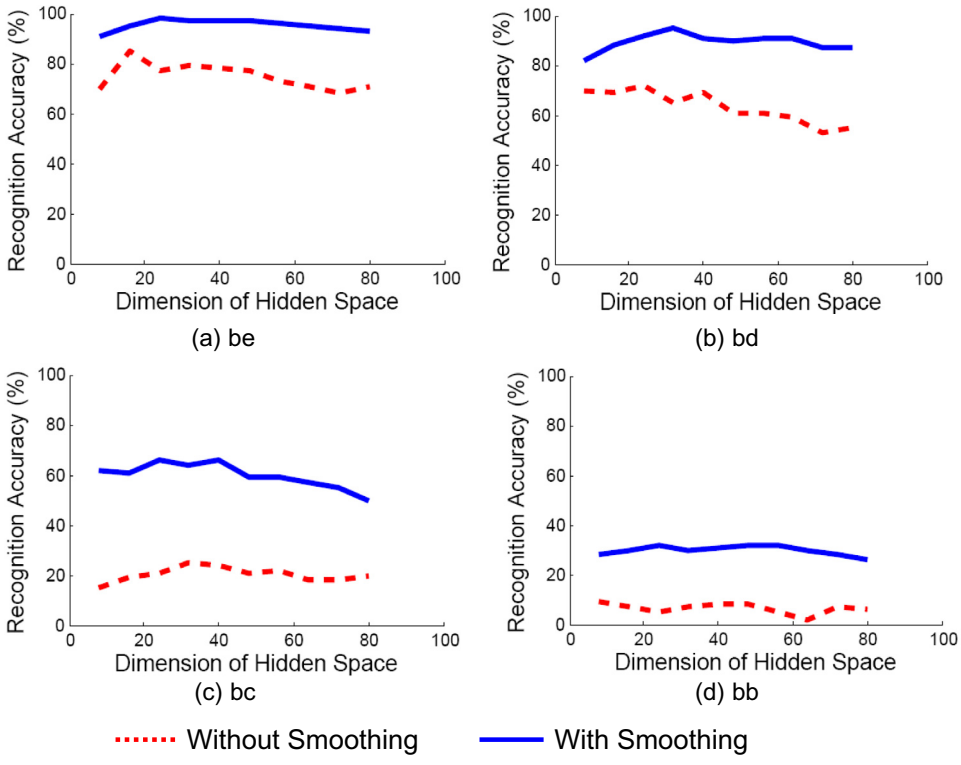

Fig. 5. Results on feature area smoothing.

3.1.2. Feature area smoothing

To demonstrate the effectiveness of feature area smoothing, we perform the face recognition task using the features with the area smoothing and those without smoothing. For each subject, the non-frontal images are used for test, which well matches the conference socialization scenarios. The results are shown in Fig. 5. Obviously, the process of feature area smoothing largely improves the recognition accuracy since it makes the features robust to the inaccuracy of feature point detection and effectively weaken the effects of useless regions.

Table 1
Recognition Accuracy (%) of Different Methods on FERET database.

Methods	be (+60°)	bd (+40°)	bc (+25°)	bb (+15°)
$(PC)^2A$	36.0	15.5	8.5	6.0
$E(PC)^2A$	37.0	16.5	9.0	6.5
2DPCA	37.5	17.5	9.5	6.0
$(2D)^2PCA$	38.0	18.0	10.0	6.5
SOM	38.5	18.5	10.5	7.0
SVD-LDA	38.5	17.5	10.5	6.5
Block PCA	40.5	19.5	13.5	9.5
Block LDA	42.0	21.0	14.5	10.0
UP	41.0	21.0	15.5	11.0
DMMA	45.0	25.0	20.5	17.5
Proposed	81.0	72.0	41.0	22.0

3.1.3. Face recognition

To further evaluate the performance of the proposed framework, we compare it with other competitive face recognition algorithms with automatic feature detection for single image per person, including SVD-LDA (singular value decomposition-based linear discriminant analysis) (Zhang et al., 2005), DMMA (discriminative multi-manifold analysis) (Lu et al., 2013), $(PC)^2A$ (projected combined principle component analysis) (Wu & Zhou, 2002), $E(PC)^2A$ (enhanced projected-combined principal component analysis) (Chen, Zhang, & hua Zhou, 2004), 2DPCA (Yang, Zhang, Frangi, & Yang, 2004), $(2D)^2PCA$ (two-directional 2DPCA) (Zhang & Zhou, 2005), SOM (Tan, Chen, Zhou, & Zhang, 2005), Block PCA (Gottumukkal & Asari, 2004), Block LDA (Chen, Liu, & Zhou, 2004), and UP (uniform persuit) (Deng, Hu, Guo, Cai, & Feng, 2010). For above ten algorithms, the frontal images are used for training and the non-frontal ones are used for test. The facial part of each image is manually cropped, aligned, and resized into 60×60 pixels according to the eyes' positions, and the nearest neighbor classifier with Euclidean distance was applied for recognition. In order to conduct fair comparison, the proposed method also works on 60×60 images in this experiment. Moreover, for the proposed method, we do not pre-assign the pose of the test faces. Instead, it is automatically determined by jointly optimizing L_t and j in Eq. (17). As shown in Table 1, the proposed method achieves higher recognition accuracy than the other algorithms for all poses.

3.2. Experiments on real conference socializing datasets

To further evaluate the performance of proposed method, we test it on two self-collected conference socializing datasets: the G20 dataset and the Oscars dataset. For the G20 dataset, we collect

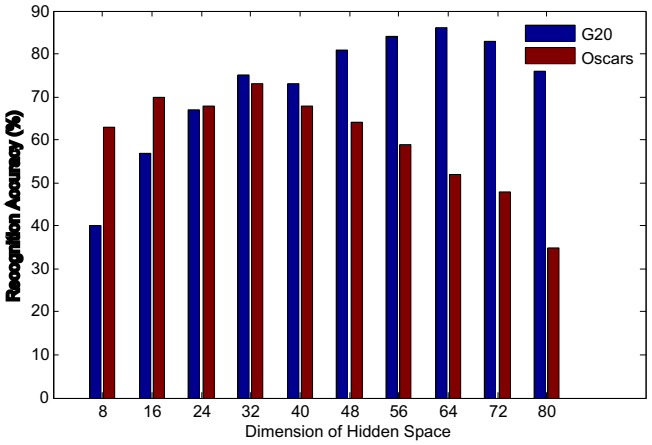

Fig. 7. Recognition accuracy with different dimensions of hidden space on G20 and Oscar dataset.

300 images from the internet, all of which contain various faces of 20 political leaders in G20 summit. Another 300 images of 40 famous movie stars and directors in Oscars are downloaded from the internet to build the Oscars dataset. Both of these two datasets are created under uncontrolled view of subjects. Some sample images are shown in Fig. 6.

In our experiment, the facial part of each image is manually cropped, aligned and resized to 60×60 pixels according to eyes' positions. For each subject in the datasets, one image taken from frontal view with normal expression and illuminations is selected as the registered gallery image for model learning. The other settings are the same as FERET dataset. Fig. 7 shows the recognition accuracy of proposed method on G20 and Oscars with different dimensions of hidden space. For G20, the highest accuracy is 86% when the reduced dimension is 64. For Oscars, this number is 73% achieved on the 32-dimensional space. The results are acceptable for the recognition task in real environment of conference socializing.

From the above results, we observe that the performance on G20 dataset is better than that on Oscars. The possible reason might be that the participants (especially females) in the entertainment activities often put much heavier make-up on the faces than those attending political conferences, which makes the recognition task more challenging. Moreover, in our experiment, the number of individuals in the gallery of Oscars is twice that of G20 (40 for Oscars and 20 for G20), which might cause the degradation of performance.

(a) G20 (b) Oscars

Fig. 6. Sample images from G20 and Oscars datasets.

4. Conclusions

In this paper, we presented a novel face recognition framework, which aims at providing better socializing experience when attending a conference. The main contributions of this work are summarized as follows: (1) An automatic feature point detection scheme is implemented via template matching, which largely alleviates the users' workload; (2) A feature area smoothing strategy is presented to highlight the important feature regions, which benefits the following processing; (3) A feature mapping between the hidden identity space and the feature space is constructed in order to make full use of the only frontal registration image for each person; and (4) A unified face recognition framework is established according to the special requirements in the context of conference socializing.

We have already shown that the proposed face recognition framework is applicable to conference socialization scenarios in our experiments. Actually, it has good impacts and practical implications in a wide range of applications. Since the proposed framework is robust to large within-class variations, it can be used in industry applications such as the image search engines as well as the security devices such as the video surveillance systems. Furthermore, the proposed framework need only the frontal registration picture for each person in training, which makes it potentially suitable for many situations where only one training data sample per class is available in the system, such as law enhancement, e-passport, and smart ID card identification.

Although the proposed face recognition framework has performed better than existing methods in the experiments, there is much room for improvement. From Fig. 5(d) and the last column of Table 1 we can observe that the recognition accuracy of the proposed method is relatively low when the angle of the face is very small (The angle of the frontal face is $+90^{\circ}$). So how to combine the local feature points and holistic information of face images to improve the performance of the proposed framework in such a small-angle situation is the first future work we need to consider. Moreover, the datasets that we have used to evaluate the proposed system are relatively simple. In the future, we need to evaluate the proposed system on more complicated conference scenario. Another meaningful future work is to improve the efficiency of the algorithm in order to transplant the whole system to the portable devices. Besides, how to embed the proposed technology into ambient intelligence applications to improve the personalized recommendations according to user specific needs and preferences is worthwhile to investigate. Last but not least, we are interested in integrating domain-specific knowledge into our technology to adapt the system for more scenarios in practice such as human behavior prediction and medical diagnosis.

Acknowledgments

This work was supported by grant G-UA69: Implicit Learning for Image/Video Retrieval on Large Scale Datasets.

References

Abdullah, M. F. A., Sayeed, M. S., Muthu, K. S., Bashier, H. K., Azman, A., & Ibrahim, S. Z. (2014). Face recognition with symmetric local graph structure (slgs). *Expert Systems with Applications, 41*(14), 6131–6137.

Belhumeur, P. N., Hespanha, J. a. P., & Kriegman, D. J. (1997). Eigenfaces vs. fisherfaces: Recognition using class specific linear projection. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 19*(7), 711–720.

Blanz, V. & Romdhani, S. (2002). Face identification across different poses and illuminations with a 3d morphable model. In *Proceedings of the fifth IEEE international conference on automatic face and gesture recognition* (pp. 192–197).

Chen, S., Liu, J., & Zhou, Z.-H. (2004). Making FLDA applicable to face recognition with one sample per person. *Pattern Recognition, 37*(7), 1553–1555.

Chen, S., Zhang, D., & hua Zhou, Z. (2004). Enhanced $(pc)^2a$ for face recognition with one training image per person. *Pattern Recognition Letters, 25*, 1173–1181.

Deng, W., Hu, J., Guo, J., Cai, W., & Feng, D. (2010). Robust, accurate and efficient face recognition from a single training image: A uniform pursuit approach. *Pattern Recognition, 43*(5), 1748–1762.

Gottumukkal, R., & Asari, V. K. (2004). An improved face recognition technique based on modular pca approach. *Pattern Recognition Letters, 25*(4), 429–436.

Gumus, E., Kilic, N., Sertbas, A., & Ucan, O. N. (2010). Evaluation of face recognition techniques using pca, wavelets and svm. *Expert Systems with Applications, 37*(9), 6404–6408.

He, X., Yan, S., Hu, Y., Niyogi, P., & Zhang, H.-J. (2005). Face recognition using laplacianfaces. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 27*(3), 328–340.

Lu, J., Tan, Y.-P., & Wang, G. (2013). Discriminative multimanifold analysis for face recognition from a single training sample per person. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 35*(1), 39–51.

Perlibakas, V. (2004). Distance measures for pca-based face recognition. *Pattern Recognition Letters, 25*(6), 711–724.

Phillips, P. J., Moon, H., Rizvi, S. A., & Rauss, P. J. (2000). The FERET evaluation methodology for face-recognition algorithms. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 22*(10), 1090–1104.

Prince, S. J. D., Elder, J. H., Warrell, J., & Felisberti, F. M. (2008). Tied factor analysis for face recognition across large pose differences. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 30*(6), 970–984.

Tan, X., Chen, S., Zhou, Z.-H., & Zhang, F. (2005). Recognizing partially occluded, expression variant faces from single training image per person with SOM and soft k-NN ensemble. *IEEE Transactions on Neural Networks, 16*(4), 875–886.

Tan, X., Chen, S., Zhou, Z.-H., & Zhang, F. (2006). Face recognition from a single image per person: A survey. *Pattern Recognition, 39*(9), 1725–1745.

Turk, M., & Pentland, A. (1991). Eigenfaces for recognition. *Journal of Cognitive Neuroscience, 3*(1), 71–86.

Vignolo, L. D., Milone, D. H., & Scharcanski, J. (2013). Feature selection for face recognition based on multi-objective evolutionary wrappers. *Expert Systems with Applications, 40*(13), 5077–5084.

Wu, J., & Zhou, Z.-H. (2002). Face recognition with one training image per person. *Pattern Recognition Letters, 23*(14), 1711–1719.

Xu, Y., Song, F., Feng, G., & Zhao, Y. (2010). A novel local preserving projection scheme for use with face recognition. *Expert Systems with Applications, 37*(9), 6718–6721.

Yang, J., Zhang, D., Frangi, A. F., & Yang, J. (2004). Two-dimensional pca: A new approach to appearance-based face representation and recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 26*(1), 131–137.

Zhang, D., Chen, S., & Zhou, Z.-H. (2005). A new face recognition method based on svd perturbation for single example image per person. *Applied Mathematics and Computation, 163*(2), 895–907.

Zhang, D., & Zhou, Z.-H. (2005). (2d)2pca: Two-directional two-dimensional pca for efficient face representation and recognition. *Neurocomputing, 69*(1-3), 224–231.

Zhang, X., & Gao, Y. (2009). Face recognition across pose: A review. *Pattern Recognition, 42*(11), 2876–2896.