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:

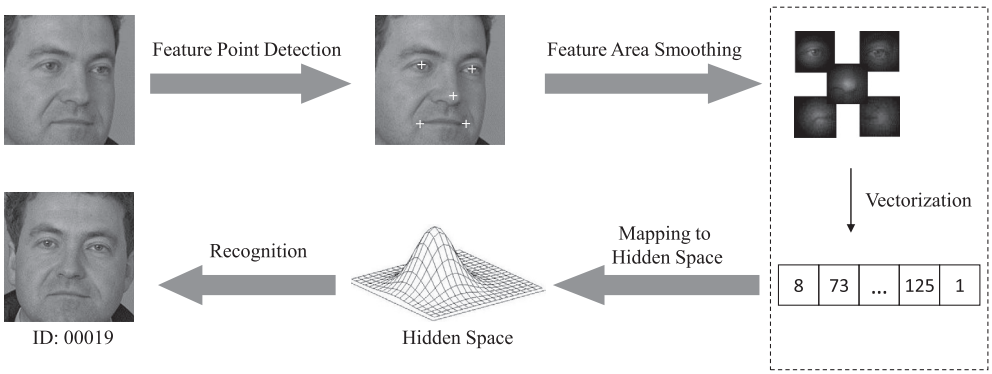

Fig. 1. Procedure of the proposed framework.