ARTICLE IN PRESS

Neurocomputing ∎ (∎∎∎∎) ∎∎∎–∎∎∎

Contents lists available at ScienceDirect

Neurocomputing

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

Latent face model for across-media face recognition

Jiang-Jing Lv, Jia-Shui Huang, Xiang-Dong Zhou*, Xi Zhou, Yong Feng

Intelligent Multimedia Technique Research Center, Chongqing Institute of Green and Intelligent Technology, Chinese Academy of Sciences, Chongqing 400714, PR China

ARTICLE INFO

Article history:
Received 19 January 2016
Received in revised form
22 June 2016
Accepted 9 August 2016
Communicated by Grana Manuel

Keywords:
Across-Media Face Recognition
Latent Face Model
Joint Bayesian model

ABSTRACT

Across-media face recognition refers to recognizing face images from different sources (e.g., face sketch, 3D face model, and low resolution image). In spite of promising processes achieved in face recognition recent years, across-media face recognition is still a challenging problem due to the difficulty of feature matching between different modalities. In this paper, we propose a latent face model that creates mappings from a hidden space to different media space. Images from different media of the same person share the same latent vector in hidden space. A coupled Joint Bayesian model is used to calculate the joint probability of two faces from different media. To verify the effectiveness of our proposed method, extensive experiments conducted on various databases: self-collected low-resolution vs. high-resolution database, sketches vs. photos databases, 3D face model vs. photos on LFW database. Experimental results show that our method boosts the performance of face recognition with images from different sources.
© 2016 Elsevier B.V. All rights reserved.

1. Introduction

Since face recognition systems have achieved high performance under constrained environment, face recognition applications progress toward unconstrained environment, such as security systems, intelligent visual surveillance systems, immigration automated clearance system, etc. However, face images come from different modalities in some applications. For instance, for the police mug-shot retrieval system, a sketch drawing based on the recollection of an eyewitness is usually used for searching suspects through a photo database; for immigration automated clearance system, the image in the identity card or e-passport microchip is used to compare with the face of the individual to verify that the holder of the passport is the rightful owner, while the identity card face images are all frontal with low-resolution. These face recognition applications which match images from different sources are called across-media face recognition [1]. Across-media face recognition is more difficult than traditional face recognition task. First of all, different from traditional face recognition which works on constrained scenarios, across-media face recognition works on unconstrained scenarios. In addition, due to the textural discrepancies between images with different modalities [2], across-media face recognition has to establish a connection between face images from different media for feature matching.

At present, there exist three ways of addressing across-media face recognition: synthesis based methods, common space projection based methods and feature descriptor based methods. Image synthesis based methods [3–8] tend to transform images into the same modality to make all data comparable. However, the synthesis process is difficult and the fidelity of the synthesized images cannot be guaranteed. Common space projection based methods [9–14] project face images from different sources into a common space to make them directly comparable. But the projection needs large number of training data and sometimes the projected features may lose the information of discrimination which impact recognition performance. Feature descriptor based methods [15–18] represent face images by local feature descriptors which is robust to different modalities and only work to specific situations (e.g., photo-sketch). Therefore, across-media face recognition is still a challenge problem.

Compared with previous machine learning methods, faces under different modalities can be easier recognized by human eyes due to our brain is able to abstract discriminative feature shared by faces with different modalities. Inspired from this phenomenon, we believe there exist mappings, which are capable of mapping images from different scenarios to the invariant feature space. Therefore, in this paper, we propose a novel latent face model for across-media face recognition. As shown in Fig. 1, it is assumed that the representations of the same person from different media (e.g., low-resolution images, 3D face models, and sketches) can be generated by the same underlying latent vector from a hidden space. Specifically, for different types of face media, we can find

* Corresponding author.
E-mail addresses: lvjiangjing@cigit.ac.cn (J.-J. Lv), huangjiashui@cigit.ac.cn (J.-S. Huang), zhouxiangdong@cigit.ac.cn (X.-D. Zhou), zhouxi@cigit.ac.cn (X. Zhou), fengyong@cigit.ac.cn (Y. Feng).

http://dx.doi.org/10.1016/j.neucom.2016.08.036
0925-2312/© 2016 Elsevier B.V. All rights reserved.

Please cite this article as: J.-J. Lv, et al., Latent face model for across-media face recognition, Neurocomputing (2016), http://dx.doi.org/10.1016/j.neucom.2016.08.036

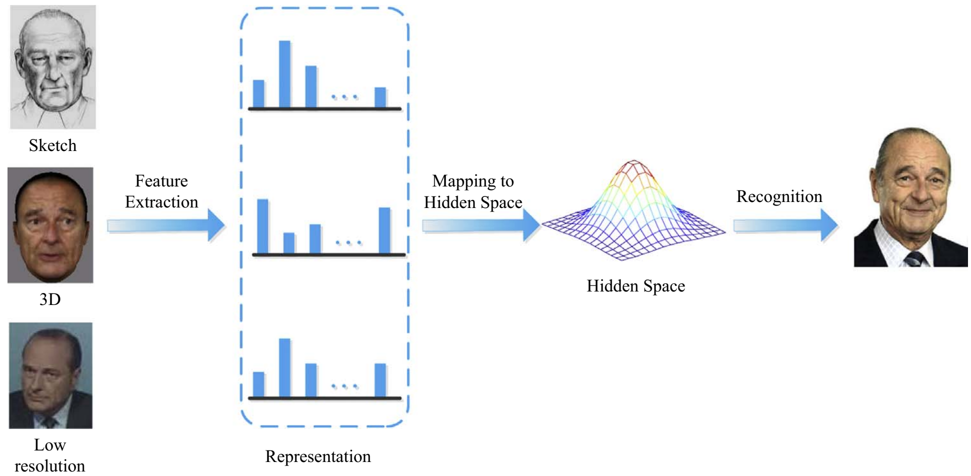

Fig. 1. Procedure of the proposed framework.

the corresponding mappings between the hidden space and different media. After the mapping models between hidden space and media are established, the Bayes' rule and EM algorithm are adopted for solving the parameters. Finally, a Joint Bayesian model is used to calculate joint probability of two faces from different media, meanwhile, inter-personal and intra-personal variations over different media are implicitly learned.

Our method has stronger discrimination ability than Common space projection based methods [9–14] and can be applied to different unconstrained scenarios. The performance of the proposed approach is thoroughly validated on various databases: self-collected low-resolution vs. high-resolution database, sketches vs. photos databases [16,7,19], 3D face model vs. photos on LFW database [20].

The rest of this paper is organized as follows. In Section 2, we briefly review published methods related to across-media face recognition. We detail the proposed latent face model in Section 3. Experimental setup and results are presented in Section 4. We conclude our work in Section 5.

2. Related work

In this section, we briefly review some prior works on across-media face recognition.

Image synthesis was first proposed for across-media recognition, which try to convert one modality to another modality. Tang and Wang proposed eigen-transformation algorithm [5] for holistic mapping, which try to reconstruct face by PCA from training samples using linear mapping. In some cases, holistic mapping cannot be well approximated, therefore nonlinear face mapping between face patches were proposed. In [4], Liu et al. proposed a nonlinear face sketch synthesis and recognition method using local linear embedding between patches. In [3,21], the authors employed embedded hidden Markov model and a selective ensemble strategy to synthesize a sketch. In [7], Wang and Tang proposed a multiscale Markov Random Fields model for face photo-sketch synthesis.

Recently, numerous learning-based approaches [9–14] are used to represent face in different scenes, which aim to minimize the intra-modality difference. In [11], Lin and Tang proposed a common discriminant feature extraction algorithm to map the images from different modality into a common feature space. In [14], Yi et al. employed a canonical correlation analysis correlation mechanism to learn the correlation between NIR and VIS. In [13], Sharma and Jacobs used partial least squares to linearly map heterogeneous faces to a common linear subspace. In [22], Eleftheriadis et al. learned a discriminative manifold shared by multiple views of a facial expression for efficient view-invariant facial expression recognition.

In addition, some feature descriptor based approaches [15–18,23] were also proposed, which are insensitive to changes in modality, but are only suitable to specific situations. In [2], Peng et al. proposed graphical representation heterogeneous face recognition approach, which does not rely on any synthesis or projection procedure but takes spatial information into consideration.

Our approach belongs to learning-based approaches. Contrast to previous methods, which is specific to constrained scenarios with one-versus-one manner, our method is suitable for various unconstrained scenarios. In our model, samples from different media are projected into a discriminant common space, faces with different modalities can be efficiently recognized. Two prior methods [24,25] are similar to our method. In [24], the authors have used factor analysis method to map an idealized identity space to the observed data space where same people with different poses share the same hidden vector, and posterior probabilistic over possible matches is used to calculate the similarity. But it only considered the constrained environment and within-class variations in each scenario that are not efficiently learned. In [25], Chen et al. tried to model the feature representation of face as the sum of inter-personal and intra-personal variations, and the Joint Bayesian was used to calculate joint probability of two faces being the same or different person. However, this method is only suitable to calculate the similarities of faces from the same scenario. In our method, we combine the benefits of these two methods and make some extensions. Firstly, in the generative process, Prince et al. [24] assumed that noise term ε_{ijk} is a zero-mean multivariate Gaussian distribution with an unknown diagonal covariance matrix Σ_k, while in our model ε_{ijk} represents the face variance (e.g., lighting, pose, and expressions) within the same identity and follows Gaussian distribution with a zero-mean and full covariance matrix Σ_k. Secondly, in feature matching process, [25] used images form the same modality to calculate the two joint probabilities, while we extend the Joint Bayesian model for across-media face recognition. Lastly, both prior methods work on constrained scenarios, while our method is suitable for unconstrained scenarios.

3. Latent face model

In this section, we will firstly give a derivation of latent face model, and then illustrate the details about the formulation and optimization, followed by feature matching process.

Download English Version:

https://daneshyari.com/en/article/4948386

Download Persian Version:

https://daneshyari.com/article/4948386

Daneshyari.com