Information Sciences 506 (2020) 19–36

Contents lists available at ScienceDirect

Information Sciences

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

Cross-resolution face recognition with pose variations via multilayer locality-constrained structural orthogonal procrustes regression

Guangwei Gao[a,b,c,*], Yi Yu[b], Meng Yang[d], Heyou Chang[e], Pu Huang[a], Dong Yue[a]

[a] Institute of Advanced Technology, Nanjing University of Posts and Telecommunications, Nanjing, China
[b] Digital Content and Media Sciences Research Division, National Institute of Informatics, Tokyo, Japan
[c] Provincial Key Laboratory for Computer Information Processing Technology, Soochow University, Suzhou, China
[d] School of Data and Computer Science, Sun Yat-Sen University, Guangzhou, China
[e] Key Laboratory of Trusted Cloud Computing and Big Data Analysis, Nanjing XiaoZhuang University, Nanjing, China

ARTICLE INFO

Article history:
Received 24 April 2019
Revised 31 July 2019
Accepted 2 August 2019
Available online 2 August 2019

Keywords:
Face recognition
Low-resolution
Pose variations
Nuclear norm

ABSTRACT

In real video surveillance scenes, the extracted face regions generally have low-resolution (LR) and are sensitive to pose and illumination variations; these flaws undoubtedly degrade the subsequent recognition task. To overcome these challenges, we propose an approach named multilayer locality-constrained structural orthogonal Procrustes regression (MLCSOPR). The proposed MLCSOPR not only learns the pose-robust discriminative representation features to reduce the resolution gap between the LR image space and the high-resolution (HR) one but also strengthens the consistency between the LR and HR image space. In particular, several contributions are made in this paper: (i) Inspired by the orthogonal Procrustes problem (OPP), a matrix approximation is exploited to find an optimal correction between two data matrices. (ii) The nuclear norm constraint is applied to the reconstruction error to maintain the structural property. (iii) Based on the abovementioned learned resolution-robust representation features, a linear regression-based classification strategy is adopted to recognize the LR input face images. Experiments on commonly used face databases have shown the effectiveness of the proposed method on cross-resolution face matching with pose variations.

1. Introduction

Over the past decades, face recognition technology has achieved significant progress [6,7,28–30,32,34,35,38,39]. In particular, pose-robust recognition tasks have been performed on cases where both the gallery and probe faces have the same resolution [32,38]. However, because of the long distance between the object and the surveillance system, the extracted face regions usually have low-resolution (LR), so they lack detailed features that enable recognition. This research scenario is regarded as cross-resolution face recognition, to which previous face recognition methods cannot be directly applied. Generally, existing approaches can be classified into three classes: (i) matching the acquired LR images to the down-sampled high-resolution (HR) galleries; (ii) matching the up-sampled LR images to the HR galleries; (iii) extracting discriminative fea-

* Corresponding author at: Institute of Advanced Technology, No.9 Wenyuan Road, Nanjing 210046, China
E-mail address: csggao@gmail.com (G. Gao).

https://doi.org/10.1016/j.ins.2019.08.004
0020-0255/

tures in the respective manifold to perform the matching task on the obtained feature space. The down-sampling strategy may reduce the resolution gap, but this procedure inevitably loses many useful discriminative facial details for recognition.

Super-resolution based methods: Researchers have exploited extensive learning-based super-resolution (SR) approaches to acquiring the desired HR face image from observed LR image to reduce the resolution discrepancy. Barker and Kanade [4] first introduced the term "face hallucination" and presented a Bayesian formulation based method. Jian et al. [14] presented a singular value decomposition (SVD)-based mapping to preserve the domain information in the HR face image space. Recently, many smaller patches based SR techniques have been proposed. Unlike previous methods adopting a fixed number of neighbors for representation learning, Li and Liang [21,22] introduced sparse representation technology to shape the prior model, which guides the reconstruction process. Lately, since the position prior becomes crucial in face reconstruction, Ma et al. [25] presented a local patch-based face hallucination approach. To alleviate over-fitting in [25], Jung et al. [15] incorporated sparsity before further enhancing the face reconstruction qualities. Then, Jiang et al. [16,17] presented a locality-constrained iterative neighbor embedding scheme to reveal the true topology of the nonlinear manifold. More recently, Jiang et al. [18] introduced smooth prior to obtain stable reconstruction weights. A robust locality-constrained bi-layer representation model is proposed by Liu et al. [23] to hallucinate the desired face images and to remove noise simultaneously. These abovementioned learning-based SR methods aim to obtain the optimal representation on a given training set instead of acquiring the discriminative features in the face super-resolution procedure.

Resolution-robust feature extraction based methods: Compared to the visual quality of the reconstructed faces, resolution-robust feature extraction based methods take face recognition performance into account. Moutafis and Kakadiaris [27] jointly optimized metric learning and representation learning and then used the learned metric for matching. Bhatt et al. [1] improved LR-HR face matching performance by applying ensemble-based co-transfer learning. Yang et al. [40] presented a united face hallucination and recognition framework via sparse representation. Mudunuri and Biswas [28] applied a multidimensional scaling (MDS) scheme to learn a joint transformation function to preserve the distance between LR and HR faces. Inspired by manifold learning, many common space-based approaches have been presented. Wang et al. [36] wanted to use a coupled mapping to depict the non-linearity between LR and HR faces in terms of a low dimensional embedding. Inspired by the supervised approach, Jiang et al. [19] embedded the face pairs into a discriminative feature space. Lu et al. [24] proposed to simultaneously extract the discriminative features and learn the mapping function from the LR to the HR feature space via a semi-coupled representation feature learning. A discriminant correlation analysis method that was introduced by Haghighat and Abdel-Mottaleb [12] projects the features extracted from LR and HR images into a common space. Recently, Banerjee and Das [2] proposed a generative adversarial network-based deep method to reconstruct HR images from LR probe images for recognition. Additionally, Gao et al. [8] proposed obtaining resolution-robust representation features for cross-resolution face recognition with occlusions. However, when learning the representation features, these methods do not consider pose variations, which degrade the stability of the learned features.

In this work, we design a multilayer locality-constrained structural orthogonal Procrustes regression (MLCSOPR) scheme to improve cross-resolution face matching with pose variations. Existing methods usually perform the recognition task on super-resolved faces or perform resolution-robust feature extraction on both LR and HR faces directly. In contrast, our proposed method first uses a structural orthogonal Procrustes regression scheme to learn discriminative representation features that are robust to pose variations in their respective HR and LR face space. Then our method performs recognition based on these learned resolution-robust features. Specifically, a matrix approximation method is used to find an optimal transformation matrix to adjust the pose from one image to that of another. Furthermore, by taking into account the illumination variations and possible structural noises (e.g., disguise or occlusion) in the probe images, the structure constraint (i.e., nuclear norm) is applied to the reconstruction error to maintain the image's appearance properties. Additionally, we incorporate locality-constrained regularization into our method to strengthen the discriminability of the learned features. Based on these learned resolution- and pose-robust discriminative features in their respective image spaces, the sparse representation induced method has a better capability of recognizing the input LR probe. Experimental tests on some commonly used face databases have verified the superiority of our proposed method.

This paper is an extension of our previous conference paper [9] with the following improvements: (i) a MLCSOPR model to further improve performance, (ii) an in-depth analysis of the proposed method, and (iii) extensive experimental evaluations of the method's performance. We organize the rest of our paper as follows: Some related works are briefly reviewed in Section 2. The proposed method is described in Section 3. A relevant analysis of the proposed method is given in Section 4. The performance evaluation is given in Section 5. Future work and conclusions are provided in Section 6.

2. Related work

2.1. Orthogonal procrustes problem

The orthogonal Procrustes problem (OPP) [10,13] is briefly discussed in this section. The OPP desires to find an optimal reflection that adjusts a matrix X to approximate another target matrix B.

Without loss of generality, the two-dimensional data and the orthogonal matrix can be denoted as X and $B\in\Re^{p\times q}$, $Q\in\Re^{p\times q}$, respectively. Then, the OPP can be denoted as

$$\min_{Q} \|XQ-B\|_F^2, \quad s.t. \quad Q^TQ=I_q, \tag{1}$$