ARTICLE IN PRESS

Pattern Recognition (xxxx) xxxx–xxxx

Contents lists available at ScienceDirect

Pattern Recognition

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

Pose-invariant face recognition with homography-based normalization

Changxing Ding[a,*], Dacheng Tao[b]

[a] School of Electronic and Information Engineering, South China University of Technology, Guangzhou 510000, China
[b] Centre for Artificial Intelligence, FEIT, University of Technology Sydney, Ultimo, NSW 2007, Australia

ARTICLE INFO

Keywords:
Pose-invariant face recognition
Face synthesis
Pose normalization
Homography

ABSTRACT

Pose-invariant face recognition (PIFR) refers to the ability that recognizes face images with arbitrary pose variations. Among existing PIFR algorithms, pose normalization has been proved to be an effective approach which preserves texture fidelity, but usually depends on precise 3D face models or at high computational cost. In this paper, we propose an highly efficient PIFR algorithm that effectively handles the main challenges caused by pose variation. First, a dense grid of 3D facial landmarks are projected to each 2D face image, which enables feature extraction in an pose adaptive manner. Second, for the local patch around each landmark, an optimal warp is estimated based on homography to correct texture deformation caused by pose variations. The reconstructed frontal-view patches are then utilized for face recognition with traditional face descriptors. The homography-based normalization is highly efficient and the synthesized frontal face images are of high quality. Finally, we propose an effective approach for occlusion detection, which enables face recognition with visible patches only. Therefore, the proposed algorithm effectively handles the main challenges in PIFR. Experimental results on four popular face databases demonstrate that the proposed approach performs well on both constrained and unconstrained environments.

1. Introduction

Face recognition is one of the most important biometric techniques. It has wide potential in many real-world applications, e.g., video surveillance, access control systems, forensics and security, and social networks [1–9]. The key advantage of face recognition lies in its non-intrusive property, which means it can work in a passive manner. However, the downside of this property is that the appearance of face images is vulnerable to a number of factors, e.g., pose, illumination, occlusion, and expression variations [10]. In particular, pose variation is the primary stumbling block to realizing the full potential of face recognition, as argued in a recent survey [11]. In this paper, we study the pose-invariant face recognition (PIFR) problem, which targets at recognizing face images captured under arbitrary poses.

Pose variation dramatically changes the appearance of face images. The appearance difference caused by pose variations usually exceeds the intrinsic appearance difference between subjects. As illustrated in Fig. 1, pose variation results in displacement of facial components, non-linear texture warping, and self-occlusion. Besides, pose variation is often combined with other factors, e.g, image blur and illumination variation, to jointly affect face recognition, as shown in Fig. 2. To handle these challenges, a number of PIFR approaches have been proposed. Among existing approaches, pose normalization is advantageous as it produces canonical pose faces with high fidelity, and usually requires no training data. Existing pose normalization approaches can be divided into two categories: 2D methods [12–14] and 3D methods [15–18]. As the face is essentially a 3D object, the appearance change caused by pose variation can be modeled more accurately with an ideal 3D face model. However, 3D modeling from a single 2D face image is an ill-posed problem and thus difficult in practice. Another disadvantage of 3D methods is that they depend on complicated computer graphics techniques for face image rendering. In comparison, 2D methods conduct pose normalization within the 2D image domain. Due to the lack of one degree of freedom, accurate pose normalization with 2D methods is difficult. Existing 2D methods usually adopt computationally expensive algorithms, e.g., Markov Random Fields (MRF) [14], Lucas–Kanade algorithm [12], to promote accuracy in pose normalization.

In this paper, we propose a novel pose normalization approach that combines the advantages of both 3D methods and 2D methods. In our approach, a grid of dense 3D facial landmarks are projected to the 2D image by aligning five semantically corresponding facial landmarks between the face image and a generic 3D face model. The grid of facial landmarks efficiently establishes dense correspondence of face images across pose. Next, by assuming the local patch around each facial landmark is a simple planar surface, the transformation of the local

* Corresponding author.
E-mail addresses: chx.ding@gmail.com (C. Ding), dacheng.tao@uts.edu.au (D. Tao).

http://dx.doi.org/10.1016/j.patcog.2016.11.024
Received 15 July 2016; Received in revised form 1 November 2016; Accepted 28 November 2016

Fig. 1. Sample images with dramatic pose variations. Pose variation significantly changes the appearance of face images.

patches across pose is efficiently approximated by homography based on landmarks in the patch. With the estimated transformation, the non-linear texture warping across pose is corrected. Compared with existing 2D pose normalization methods, e.g., Markov Random Fields (MRF) [14], Lucas–Kanade algorithm [12], and optical flow [51], the homography-based method estimates the local warp quite efficiently.

The above method reconstructs frontal-view face image patches from un-occluded facial textures. Existing feature extraction methods, e.g., local descriptors, can be applied on the corrected face patches to compose the face representation. Therefore, occlusion detection is important to distinguish occluded facial textures from visible facial textures. We further propose a method for occlusion detection and a scheme to extract fixed-length face representations from pose varied face images. Based on face symmetry, we extract patch-level features from both the original face image and the horizontally flipped version. For each patch pair of the two images, their features are fused by weighting according to their visibility. The patch-level feature vectors are then concatenated to compose the complete face representation. In this way, we obtain a fixed-length face representation for each face, regardless of their poses. The advantage of this method is that we can make the best of visible facial textures for face recognition.

In this paper, we term the homography-based pose normalization method as HPN. The remainder of the paper is organized as follows: Section 2 briefly reviews related works for PIFR. The proposed HPN method is illustrated in Section 3. Face representation based on HPN is described in Section 4. Experimental results are presented in Section 5, leading to conclusions in Section 6.

2. Related works

A number of approaches have been proposed to solve the PIFR problem from various perspectives. Among existing works, pose-robust feature extraction and pose normalization are the two most important categories of methods. For a comprehensive review of existing methods, we direct readers to a recent survey [11]. In this section, we only review the most relevant works to this paper.

Methods falling in the pose-robust feature extraction category can be further divided into two types: handcrafted features and learning-based features. The handcrafted methods usually extract features from semantically corresponding patches across pose; therefore the face representations are less sensitive to the displacement of facial textures caused by pose variation. For example, Cao et al. [19] extracted local descriptors from image patches cropped around 9 facial components rather than the holistic face image. Chen et al. [20] extracted multi-scale Local Binary Patterns (LBP) features from image patches cropped around 27 2D landmarks, which are detected by a face alignment algorithm. Yi et al. [21] extracted Gabor magnitude features from much denser landmarks located by a 3D morphable face model. The above methods depend heavily on the precision of face alignment. Unfortunately, dense landmark detection in unconstrained face images is still a hard problem. To bypass the difficulty, approaches that are landmark detection-free are proposed. For example, Arashloo and Kittler [22] proposed to detect semantically corresponding patches using MRF. The downside of this approach is that its efficiency is low. The common shortage of the handcrafted methods is that they cannot

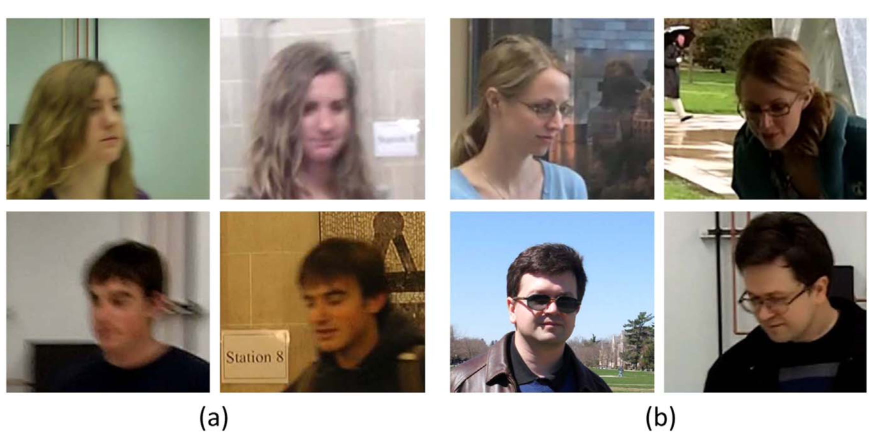

Fig. 2. Pose variation is usually combined with other factors, e.g., low resolution, image blur and illumination variation, to jointly affect face recognition. (a) Combined challenge of image blur and pose variation; (b) combined challenge of illumination and pose variation.