Pattern Recognition 000 (2017) 1–12

Contents lists available at ScienceDirect

Pattern Recognition

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

From one to many: Pose-Aware Metric Learning for single-sample face recognition

Weihong Deng*, Jiani Hu, Zhongjun Wu, Jun Guo

School of Information and Communication Engineering, Beijing University of Posts and Telecommunications, Beijing 100876, China

ARTICLE INFO

Article history:
Received 10 September 2016
Revised 26 September 2017
Accepted 16 October 2017
Available online xxx

Keywords:
Face recognition
Single sample per person
Metric learning
3D generic elastic model
Face re-rendering
3D face construction

ABSTRACT

Pose and illumination variations are very challenging for face recognition with a single sample per person (SSPP). In this paper, we address this issue by a Pose-Aware Metric Learning (PAML) approach. Our primary idea is "*from one to many*": Synthesizing many images of sufficient pose and illumination variability from the single training image, based on which metric learning approach is applied to reduce these "synthesized" variations at each quantified pose. For this purpose, given a single frontal training image, a multi-depth generic elastic model and an extended generic elastic model are developed to synthesize facial images of the target pose with varying 3D shape (depth) and illumination variations respectively. To reduce these "synthesized" variability, Pose-Aware Metric spaces are separately learnt by linear regression analysis at each quantized pose, and pose-invariant recognition is performed in the corresponding metric space. By preserving the detailed texture and reducing the shape variability, the PAML method achieves an 100% accuracy on the Multi-PIE database under the test setting across poses, which is significantly better than the traditional methods that use a large generic image ensemble to learn the cross-pose transformations. On the more challenging setting across both poses and illuminations, PAML outperforms the recent deep learning approaches by over 10% accuracy.

1. Introduction

We consider the pose-invariant face recognition problem with a single training sample per person. This single-sample face recognition (SSFR) problem is one of the major challenges in many real-world applications on law enforcement and homeland security [1]. Theoretically, it is an extreme case of the small sample size problem that deteriorates conventional pattern recognition techniques. As the supervised learning techniques are not applicable without intraclass information, unsupervised techniques, which find the low-dimensional embedding of the gallery data by ICA [2], PCA [3] or their variants [4–6], have been widely applied. However, these methods are suitable only for face representation and effective only for the recognition under constraint variations. Invariant features (e.g. Gabor feature [7,8] and local binary patterns [9]) are effective to increase the robustness to the lighting and expression changes. Unfortunately, since they discard all information about the 3D layout of the face, these feature descriptors are deficient to counteract the unobserved pose variations.

Pose variation is widely regarded as a major challenge in the automatic face recognition application. We envision the typical applications where enrollment of subjects is through frontal images with neutral light and expression (i.e., typical enrollment images for most applications), but test images come from real-world unconstrained scenarios with various poses and illuminations. Since the face images under variable pose reside in a highly nonlinear subspace, conventional subspace learning [8,10], manifold learning [11], sparse representation [12–14] and metric learning [15,16] methods designed for SSFR can not achieve satisfactory performance. Previous SSFR approaches across pose differences mostly rely on a (external) generic training set with multiple samples per person (MSPP) of similar viewpoints to the test samples [17,18]. Recently, deep learning techniques are used to learn the cross-pose transformation for the unconstrained face based on an external multi-view image ensemble [19]. Unfortunately, the performance of these methods depends heavily on the representativeness of the generic training set, though they achieve state-of-the-art performance on the Multi-PIE database [20] with the same training and testing viewpoints. In contrast, the 3D model based methods [21,22] is more flexible, since they do not rely on the similarity of the training and testing viewpoints.

In this paper, we aim to address the pose-invariant SSFR problem by exploring the discriminative information in the gallery images, by the extensions of 3D generic elastic model [23,24]. In general, our primary idea is "*from one to many*": Synthesizing many

* Corresponding author.
E-mail addresses: whdeng@bupt.edu.cn, cvpr_dwh@126.com (W. Deng).

https://doi.org/10.1016/j.patcog.2017.10.020

images of the pose and illumination variability from one single (frontal) gallery image, based on which the metric learning approach further reduces the "synthesized" variations at each quantified pose. Following this idea, the contributions of paper are as follows.

Firstly, we develop a multi-depth 3D generic elastic model (MD-GEM) with variable depth to characterize the uncertainty of the 3D shape [25], when rendering faces of different poses from a single frontal image. To address this problem caused by the depth ambiguity of the face, we make a linear assumption on the depth channel of 3D generic elastic model with a single parameter, instead of a single settled depth map of the conventional GEM [24].

Secondly, we propose an extended generic elastic model (E-GEM) [26] that couples the 3D generic elastic model with the quotient image [27] technique to synthesize faces under different poses and illumination conditions from a single frontal images. Specifically, we develop a shape-free alignment for the quotient image to achieve better face re-rendering results. This adaptive quotient image (AQI) is then used to generate the texture surface of GEM to render varying lightings.

Thirdly, inspired by the "divide and conquer" strategy, we address the pose-invariant face recognition problem by Pose-Aware Metric Learning (PAML) using the synthesized training images of each quantized pose separately. For each quantized pose, PAML applies linear regression analysis technique [16] to transform the synthesized training samples of one subject into a single point of the metric space. In the recognition stage, we first estimates the pose of probe face and then applies the corresponding pose-specific metric to perform classification.

By the virtue of GEM, the proposed PAML method can take advantage of the full texture details of the gallery image under arbitrary poses, which is essential for the highly accurate recognition. Extensive experiments on the Multi-PIE database [20] demonstrates that our method is a superior SSFR solution for variable pose, without using training set of external subjects to learn the pose-invariant transformation. Specifically, on the test setting under variable pose, the PAML method, based on the LBP descriptor of the synthesized images via MD-GEM, achieves 100% accuracy on the MPIE database. On the test setting across both poses and illuminations, PAML, based on the LBP descriptor of the synthesized images via E-GEM, outperforms the recent deep learning methods by over 10% accuracy. Moreover, PAML does not rely on any external data for training, while existing methods use a large generic image ensemble of hundreds of people to learn the pose variations.

It should be noted that this paper is an extended work of our previous conference papers [25,26]. In this paper, we present the comprehensive related works, more technical details on MD-GEM [25] and E-GEM [26], and a combined face synthesis algorithm (Algorithm 1). Moreover, we also integrate them into the proposed PAML framework to obtain much better performance than our previous work. In particular, we demonstrate the 100% accuracy of MD-GEM on the additional pose-invariant recognition experiments of MPIE database.

Algorithm 1 Image synthesis via MD-GEM and E-GEM.

Input: A frontal facial image, a 3D generic model with depth parameter α, a bootstrap set of images with a unified shape, target pose angles
Output: The ensemble of synthesized images at target poses
1: Locate the 77 feature points by off-the-shelf face alignment method or manual labeling.
2: Compute dense correspondence between the input image and 3D-GEM according to the delaunay triangulation of feature points, and then allocate the depth according to the parameter α.
3: Transform the input face to the unified shape of the bootstrap set. Solve Q and light coefficient x_j according to Eq. 4, and then stimulate each illumination by setting lighting coefficient l_j. Finally, transform the re-rendered facial images back to the original shape.
4: Map the texture from the re-rendered images with various illuminations to the 3D-GEMs according to the dense correspondence obtained in Step 2.
5: At each target pose, render the synthesized images with the 3D-GEMs of input depth (α) and lighting (l_j) parameters.

Algorithm 2 Pose-Aware Metric Learning.

Input: The training set with a frontal training image per person, quantized pose angles
Output: The transformation matrix $W^{(p)}$ for each quantized pose
1: Synthesize a predefined number of images at all quantized poses for each frontal gallery image using Algorithm 1.
2: Align all the synthesized images by the predefined landmarks. Extract feature vectors of the aligned faces.
3: At each quantized pose, compute the transformation matrix $W^{(p)}$ in Eq. 7 using the feature matrix of the synthesized images at that pose.

2. Background

Many interesting improvements on face recognition have been reported in the literature to handle pose variation. Robust feature descriptors are expected to be more robust than pixel intensity to counteract the appearance change caused by poses. LGBP [28] is a high-dimensional face descriptor which first convolves the images by a family of Gabor kernels followed by LBP coding of the filtered images. LE+LDA [29] method encodes the micro-structures of the face by a new learning-based encoding method, which can automatically achieve very good tradeoff between discriminative power and invariance. CRBM+LDA [30] learns the local descriptor by local convolutional restricted Boltzmann machines, which exploits the global structure and maintains the robustness to small misalignments.

Several statistical learning methods are proposed by leveraging the correlation among features across poses. CCA [31] aims at projecting the images of different poses onto a common feature space where the correlation between them are maximized. PLS method [32] attempts to project samples from two poses to a common latent subspace, with one pose as regressor and the other pose as response. GMA method [33] is a generalized multi-view analysis method attempting to project the images of all poses to a discriminative common space, where pose variations are minimized. Li et al. [18] represented a test face as a linear combination of training images and utilized the linear regression coefficients as features for face recognition.

3D model based methods provide straightforward solutions for pose-invariant recognition. The 3D morphable model [34] is built using PCA on 3D facial shapes and textures acquired from a laser scanner. The learned 3D face model is reconstructed by fitting the model to the input 2D image. Pose-invariant recognition can be performed by transforming posed face images to the frontal view or comparing the reconstruction coefficients of PCA. However, the PCA subspace may not be accurate enough to characterize the detailed textures of test faces. Besides 3DMM, several 3D model based methods are proposed to rotate the non-frontal face to the frontal one. Different from 3DMM, VAAM method [35] proposes a fully automatic 3D pose normalization method, which can synthesize a frontal view by aligning an average 3D model to the input non-frontal face based on the view-based AAM. We recently