Image and Vision Computing 36 (2015) 9–22

Contents lists available at ScienceDirect

Image and Vision Computing

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

# Unrestricted pose-invariant face recognition by sparse dictionary matrix☆

Ali Moeini [a,*], Hossein Moeini [b], Karim Faez [a]

[a] Electrical Engineering Department, Amirkabir University of Technology, Tehran, Iran
[b] Electrical Engineering Department, Semnan University, Semnan, Iran

ARTICLE INFO

*Article history:*
Received 6 April 2014
Received in revised form 16 September 2014
Accepted 19 January 2015
Available online 19 February 2015

*Keywords:*
Real-world
Facial expression generic elastic models
Face synthesis
Pose-invariant face recognition
Sparse representation
Sparse dictionary matrix

ABSTRACT

In this paper, a novel method is proposed for real-world pose-invariant face recognition from only a single image in a gallery. A 3D Facial Expression Generic Elastic Model (3D FE-GEM) is proposed to reconstruct a 3D model of each human face using only a single 2D frontal image. Then, for each person in the database, a Sparse Dictionary Matrix (SDM) is created from all face poses by rotating the 3D reconstructed models and extracting features in the rotated face. Each SDM is subsequently rendered based on triplet angles of face poses. Before matching to SDM, an initial estimate of triplet angles of face poses is obtained in the probe face image using an automatic head pose estimation approach. Then, an array of the SDM is selected based on the estimated triplet angles for each subject. Finally, the selected arrays from SDMs are compared with the probe image by sparse representation classification. Convincing results were acquired to handle pose changes on the FERET, CMU PIE, LFW and video face databases based on the proposed method compared to several state-of-the-art in pose-invariant face recognition.



## 1. Introduction

Real-world pose-invariant face recognition is one of the most difficult and challenging tasks in computer vision due to large changes in poses of human faces. Typical face recognition methods have been prosperous at working under controlled situations. However, carrying out pose-invariant face recognition is very difficult in real-world situations when alternations in illumination and facial expression exist.

Performances of face recognition systems are great under controlled situations, but expressively reduce for images displaying large pose, illumination and facial expression variations. Handling of head pose change is thus one of the most essential aspects for a concrete real-time face recognition method. Although there have been many advances in the last few years to incorporate robustness into head pose change in face recognition approaches, it is still a challenge to attain this manner under unrestrained situations due to precision and speed problems.

Available pose-invariant face recognition methods can be mostly categorized into two separate types: 1) 2D methods which include subspace-based techniques which present nonlinearities into a 2D shape to extract new poses [21,2], view-based methods which utilize a subset of shapes to demonstrate appearance from a diverse view [3,4] and etc., and 2) 3D model-based methods which employ a 3D model to present new poses of faces [1,5–18]. In this paper, a novel method from the third type is proposed.

A main problem of the above studies is that face images acquired under controlled conditions (e.g., the FERET database) are considered, which are usually frontal, occlusion-free, with a clean background, consistent lighting, and limited facial expressions. However, in real-world applications, face recognition needs to be performed on real-life face images captured in unconstrained scenarios. There are significant appearance variations on real-world faces, which include facial expressions, illumination changes, head pose variations, and etc. Therefore, face recognition in real-world faces is much more challenging compared to the case for faces captured in constrained conditions.

In this paper, a novel approach is proposed for unconstrained face recognition from real-world face image by 3D face reconstruction and sparse representation. Accordingly, a 3D model is initially reconstructed from frontal face images in a real-world gallery. To reconstruct a 3D model from each human face in real-world scenarios, a Facial Expression Generic Elastic Model (FE-GEM) is proposed. The FE-GEM method for facial expression-invariant 3D face reconstruction is the extension of the Generic Elastic Model (GEM) [6] to resolve the drawback of handling facial expression in 3D face reconstruction. Then, each 3D reconstructed face in the gallery is synthesized to all possible views and a Sparse Dictionary Matrix (SDM) is generated based on triplet angles of face pose for each person. On the other hand, automatic head pose estimation by the Constrained Local Model (CLM) [19] is used to extract the triplet angles (including yaw, pitch and roll) of face poses. Therefore, for each person, an array of SDM is selected based on triplet angles of face

☆ This paper has been recommended for acceptance by Ioannis A. Kakadiaris.

* Corresponding author. Tel.: +98 2122871351; fax: +98 2122865548.
*E-mail address:* ali.moeini1989@gmail.com (A. Moeini).

http://dx.doi.org/10.1016/j.imavis.2015.01.007
0262-8856/

pose which were estimated from automatic head pose estimation. Finally, the face recognition is performed by sparse representation classification [20] between the selected arrays of SDM and probe images. The main contribution of this paper is the generation of the SDM matrix based on triplet angles of face pose for each registered input image by sparse representation and 3D reconstructed depth to recognize face across pose changes.

The experimental evaluation was provided to evaluate the proposed method on FERET, CMU PIE and LFW face databases. Promising results were acquired to handle pose changes based on the proposed method compared with several state-of-the-art in pose-invariant face recognition. Also, evaluating the LFW database demonstrated that the proposed method is very rapid and efficient for real-world scenarios in pose-invariant face recognition. Furthermore, to real-time experiments, the proposed method was evaluated on video databases with 30 subjects. Evaluating video database demonstrated that the proposed method is very rapid and real-time for pose-invariant face recognition.

In this paper, the main contributions that are proposed are as follow:

1) The FE-GEM method is presented to handle facial expression variations for 3D face reconstruction in real-world scenarios in which reconstructed models are more accurate than the original GEM method.
2) The manner of generation of the SDM matrix is another contribution of this paper. The SDM is generated based on triplet angles of face pose for each gallery image by 3D face reconstruction and adopting the sparse representation and LBP on synthesized face for carrying out pose-invariant face recognition in real-world scenarios.
3) Speed is a main contribution of this paper that is very rapid and closely real-time in comparison with the state-of-the-art approaches to handle pose changes in face recognition.
4) The presented approach improves the recognition rate of the best 3D model-based across pose on the CMU-PIE and FERET databases. Also, the proposed approach obtained results that outperformed the best approaches on the LFW database. Moreover, the proposed method is evaluated on the video databases to recognize the face in real-time videos.

This paper is organized as follows: Section 2 describes the related works. In Section 3, automatic head pose estimation by the CLM is explained. Section 4 describes the 3D face modeling method from a single frontal face image. In Section 5, the generation manner of SDM and the training method for each human face in the database are proposed for pose-invariant face recognition. Experimental evaluations are given in Section 6 and conclusions are presented in Section 7.

## 2. Related works

In the 2D method context, Sharma et al. [21] proposed the discriminant multiple coupled latent subspace method for pose-invariant face recognition. They find the sets of projection directions for diverse views such that the projected images of the same subject in diverse views are maximally correlated in the latent space. They claimed that the discriminant analysis with artificially simulated pose errors in the latent space makes it robust to small pose errors caused due to a subject's incorrect view estimation. Also, they do a comparative analysis of three common latent space learning methods: Partial Least Squares (PLSs), Bilinear Model (BLM) and Canonical Correlational Analysis (CCA) in their coupled latent subspace method.

Also, Arashloo et al. [22] proposed an MRF-based classification method which employs the energy of the established match between a pair of images as a criterion of goodness-of-match. They claimed that by incorporating an image matching approach as part of the recognition process, the system is made robust to moderate global spatial transformations. Accordingly, by adopting a multi-scale relaxation scheme based on super coupling transform, the inference using sequential tree reweighted the message passing method [23] is accelerated. Then, by taking advantage of a statistical shape prior to the matching, the results are regularized and constrained, making the system robust to spurious structures and outliers. For classification, both textural and structural similarities of the facial images are taken into account.

On the other hand, the surviving 3D model-based approaches can be mostly separated into four categories based on how to utilize the 3D models:

1) Pose Normalization: in this method, the probe face images are normalized to frontal pose based on the 3D model, and after that the normalized probe is matched to the gallery face image [9–12].
2) Pose Synthesis: in this method, several virtual face images are created by synthesis of the 3D model to different poses for the gallery, and then the probe is matched to the virtual face images [5–8].
3) Recognition by Fitting: in this method, the gallery and probe images were fitted by the 3D model. Hence, both texture and shape parameters are utilized for performing face recognition [13–17].
4) Filter Transformation: in this method, the filters are transformed based on the pose and shape of probe face image, and after that the pose adapted filters are used for feature extraction from probe images to be compared with features of the gallery image [18].

Among the attempts to generalize 3D model-based face recognition, Blanz and Vetter [13] presented the 3D Morphable Model (3DMM), where each face was provided as a linear combination of a 3D face reconstructed model. Then, the 3DMM was fitted for each input image and, then the shape and texture parameters were reconstructed based on the analysis-by-synthesis framework. Several methods [14–17] used the 3DMM to carry out the face recognition. The major drawback of these approaches is that it requires high computational complexity to reconstruct image fitting parameters. Also, Asthana et al. [10] proposed an automatic method for pose-invariant face recognition that was a 3D pose normalization approach. This method is completely automatic and leverages the accurate 2D facial landmarks found by the method that can handle 3D pose difference up to $\pm 30°$ in pitch and $\pm 45°$ in yaw angles.

Also, Prabhu et al. [7] suggested a novel approach for real-world pose-invariant face recognition. In this method, a 3D face model was built for each person in a database utilizing only a 2D frontal face image by Generic Elastic Model (GEM). Then, a new 2D pose face was synthesized from these 3D reconstructed face models for the best matching. Also, head pose was estimated in the test images by a linear regression method according to face landmark localization. Then, each 3D reconstructed face model was synthesized at specific estimated poses. Finally, cosine distance similarity was calculated between the synthesized 3D face models and the test image to display usefulness of the pose synthesis approach for real-world scenarios. However, this method is not very fast, real-world and accurate for face recognition. For example, for matching a single test image to all poses of a 3D face model, this method will take 1 s, and for matching a single test image to all poses of 249 face models, it will approximately take 4 min. Also, accuracy and performance of this method are not very high in test images which have a large pose of face and facial expressions.

Recently, Yi et al. [18] proposed an approach for pose robust face recognition towards realistic applications, which is rapid, pose robust and can work well under unrestrained environments. Accordingly, a 3D deformable model is generated and a fast 3D model fitting algorithm is proposed to estimate the pose of the face image. Then, a set of Gabor filters are transformed according to the pose and shape of the face image for feature extraction. Finally, the Principal Component Analysis (PCA) is applied on the pose adaptive Gabor features to eliminate the redundancies and the Cosine metric is utilized to compare the similarity.

## 3. Automatic head pose estimation

The existing methods for automatic head pose estimation still have difficulties in detecting individual independent facial landmarks and in