Pattern Recognition Letters 68 (2015) 83–89

Contents lists available at ScienceDirect

Pattern Recognition Letters

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

Unconstrained pose-invariant face recognition by a triplet collaborative dictionary matrix☆

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

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

ARTICLE INFO

Article history:
Received 24 March 2015
Available online 31 August 2015

Keywords:
Real-world
Face synthesis
Pose-invariant face recognition
Collaborative representation
Collaborative dictionary matrix

ABSTRACT

In this paper, a novel method is proposed for unconstrained pose-invariant face recognition from only an image in a gallery. A 3D face is initially reconstructed using only a 2D frontal image. Then, for each person in the gallery, a Triplet Collaborative Dictionary Matrix (TCDM) is created from all face poses by rotating the 3D reconstructed models and extracting features in rotated face. Each TCDM is subsequently rendered based on triplet angles of face poses. Finally, the classification is performed by Collaborative Representation Classification (CRC) with Regularized Least Square (RLS). Promising results were acquired to handle pose changes on the FERET, LFW and video face databases compared to state-of-the-art methods in pose-invariant face recognition.

1. Introduction

Unconstrained pose-invariant face recognition is one of the most difficult and challenging tasks in computer vision because of large changes in poses of human faces. Typical face recognition methods have been prosperous at working under controlled situations [32,36]. However, carrying out pose-invariant face recognition is very difficult in real-world situations when there are alternations in illumination and expression. In this context, available pose-invariant face recognition 3D model-based methods which employ a 3D model to present new poses of faces can be separated into four categories based on how to utilize the 3D models: (1) Pose Normalization [1–3], (2) Pose Synthesis [4–7,33,34], (3) Recognition by Fitting [8–12], (4) Filter Transformation [13].

Recently, we have proposed some methods for pose-invariant face recognition [14,21,29]. In [14], we proposed a real-world framework for face recognition by Joint Dynamic Sparse Representation Classification (JDSRC) based on pose matrix generation from 3D models. Also, we proposed a Feature Library Matrix (FLM) [29] framework for pose-invariant face recognition through extracting the feature by Dual-Tree Complex Wavelet Transform (DT-CWT) and adopting iterative scoring classification by SVM. Finally, we presented the Sparse Dictionary Matrix (SDM) [21] framework for pose-invariant face recognition by Local Binary Pattern (LBP) and Sparse Representation Classification (SRC). The proposed methods had convincing and fast results on video and LFW databases.

In this paper, a fast method is proposed based on 3D face modeling from real-world face images to recognize faces robust to pose. It is an extension of our methods in [14,21] based on sparse and collaborative representation which has generated the Triplet Collaborative Dictionary Matrix (TCDM) for gallery images. Finally, the face recognition is performed by Collaborative Representation Classification (CRC) with Regularized Least Square (RLS) [16] for classification. The experiments were provided on FERET [18] and LFW [19] face database. Also, for real-time experiments, it was evaluated on video databases. In this paper, we evaluate the impact of CRC on proposed methods in [14,21]. Then it is compared with our previous methods in terms of accuracy and speed. It was demonstrated that proposed method not only is more accurate than our previous methods, but also is faster.

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

(1) TCDM is generated based on triplet angles of face pose for each subject in the gallery by adopting the collaborate representation. (2) The manner of classification by CRC + RLS on the TCDM framework is another contribution of this paper. (3) Speed is the main contribution of this paper that is real-time in comparison with state-of-the-art methods. (4) The presented approach improves the recognition rate on the FERET, LFW databases. Also, it is evaluated on the video databases.

This paper is organized as follows: Section 2 briefly describes automatic head pose estimation and 3D face modeling method used in this work. In Section 3, proposed method for pose-invariant

☆ This paper has been recommended for acceptance by L. Yin.

* Corresponding author. Tel.: +98 21 64543328; fax: +98 21 66406469.
E-mail address: alimoeini@aut.ac.ir, ali.moeini.ir@ieee.org, ali.moeini1989@gmail.com (A. Moeini).

http://dx.doi.org/10.1016/j.patrec.2015.08.012
0167-8655/

face recognition is presented. Experimental evaluations are given in Section 4 and conclusions are presented in Section 5.

2. Head pose estimation and 3D face modeling

In this paper, the CLM [15] method is used to automatically estimate head pose that took full advantage of intensity data for detecting facial landmarks in images and tracking them. There are many methods for head pose estimation [15,20,30,35]. The main reason for using the CLM in this work is that the CLM method is closely real-time for head pose estimation. For further details of the CLM method, refer to [15,20,30]. In order to construct 3D face model from images in real-world scenarios, a Facial Expression Generic Elastic Model (FE-GEM) [17,21] is used which is extending the GEM [4] method. Moeini et al. [21] proposed a method for 3D face modeling robust to facial expression. This method resolved the drawback of GEM approach for 3D reconstruction from images that have facial expressions.

3. Pose-invariant face recognition

In this section, the collaborative representation is represented. Then, the feature extraction method and the manner of generating TCDMs are proposed from the 3D reconstructed face by collaborative representation for pose-invariant face recognition. Finally, unconstrained face recognition method is proposed.

3.1. Collaborative representation

The sparse representation was described by Wright et al. [22] for face recognition. A collection of labeled training samples is given from k discrete classes. Then, the task is done to decide that a new unseen probe sample is belongs to which the class. Let $A_i = [v_{i,1}, v_{i,2},\ldots, v_{i,ni}]$ is an $m \times n_i$ training matrix from the ith class in the gallery in which the n_i training samples are set as columns. Every column $v_{i,j}$ in matrix A_i is vectorized to intensity image or a number of appropriate description of the intensity image. One effective method to use the organization of the matrix A_i for face recognition is to model the samples from a single class as presenting on a linear subspace. Meanwhile, if there are adequate samples from the ith class , any novel test sample y from the same class is estimated to present in the linear length of the columns of matrix A_i, i.e.,

$$y = \sum_{j=1}^{n_i} \alpha_{i,j} v_{i,j} \tag{1}$$

for some scalars $\alpha_{i,j} \in R$, $j = 1, 2,\ldots, n_i$. Because the identity of the probe sample is initially unknown, a novel matrix A was clarified by [23] which is the incorporation of the training samples from all the classes:

$$\mathrm{A}_{m\times n} = [\mathrm{A}_1, \mathrm{A}_2, \ldots, \mathrm{A}_k] = [\mathrm{v}_{1,1}, \ldots, \mathrm{v}_{1,n}] \ldots \left[\mathrm{v}_{k,1}, \ldots, \mathrm{v}_{k,nk}\right] \tag{2}$$

where $n = \sum_{i=1}^{k} n_i$. With this explanation of A, y in Eq. (1) is rewritten as:

$$y = Ax \tag{3}$$

where $x = [0,\ldots, 0,\ldots, \alpha_{i,1},\ldots, \alpha_{i,ni},\ldots, 0,\ldots, 0]^T$ is a coefficient vector whose corresponding arrays to the ith class are not zero and all remaining arrays are zero. The purpose is to acquire x from a novel test sample y and matrix A which is special information about the identity of the test sample to help in the face recognition. To this end, there are numerous decomposition approaches. The sparse representation methods discover the sparsest answer to the linear systems of equations $y = Ax$. This is obvious that test sample y is adequately represented by only the samples from its correct class, which if the total number of classes in A is great, it is obviously guided to sparse x. Whatever the recovered x is more sparse, it demonstrates that improves the identity of the unlabeled test sample, which guides to solve the following optimization problem:

$$\hat{x} = \arg\min_{x\in R^n} ||x||_1 \quad \text{subject to} \quad \mathrm{A}x = y \tag{4}$$

Actually, l^0-minimization instead of l^1-minimization provides the sparsest solution but it guides to a NP-hard trouble even for estimation. Recent researches in compressed sensing [24] demonstrate that if the solution required is sparse sufficient, the solution to the l^0-minimization problem is like to that of the l^1-minimization problem. For more details refer to [22]. The optimization problem in Eq. (4) is acknowledged as Basis Pursuit (BP) and it is solved in polynomial time by standard linear encoding techniques [25].

Recently, Zhang et al. [16] proposed collaborate representation as more efficient and more rapid than sparse representation. They demonstrated that CRC not only is accurate than SRC, but also it is very rapid. In convenient face recognition methods, typically the feature dimensionality will not be located too short in order for an excellent recognition rate. Therefore, we may not need to use l^1-regularization to sparsify x. Thus, collaborative representation leads to solving the following optimization problem:

$$\hat{x} = \arg\min_{x}\{||y - Ax||_2^2 + \lambda||x||_2^2\} \tag{5}$$

where λ is the regularization parameter. The benefit of the regularization term is as follows: (1) it makes the least square solution stable, and (2) it introduces a definite amount of “sparsity” to the solution $\hat{x}$, yet this sparsity is a large amount weaker than that by l^1-norm.

The solution of *CR* with RLS in Eq. (5) can be easily and analytically derived as:

$$\hat{x} = (A^T A + \lambda I)^{-1} A^T y \tag{6}$$

Let $P = (A^T A + \lambda I)^{-1} A^T$. Evidently, P is independent of y so that it can be pre-computed as a projection matrix. Once a query sample y occurs, it can just simply project y onto P via Py. This makes *CR* very fast.

The classification by $\hat{x}$ is similar to the classification in sparse representation classification. In addition to the class specific representation residual $||y - A_i.\hat{x}_i||_2$, where $\hat{x}_i$ is the coefficient vector associated with class i, the l_2-norm “sparsity” $||\hat{x}_i||_2$ can in addition bring some discrimination data for classification. Hence, CRC_RLS algorithm that was proposed by Zhang et al. [16] is summarized as follows:

3.2. Feature extraction by CR

Visual illustration of the proposed method for extracting the feature by collaborative representation (CR) [16] is shown in Fig. 1. Based on the proposed method, the process can be summarized as follows:

(1) Input: input image from subject i.
(2) For subject i, 3D faces are reconstructed by FE-GEM method.
(3) For each subject (subject i), 3D face is synthesized across pose and possible views in all poses of the face are extracted with steps of 5 degrees on either side (at three directions including yaw, pitch and roll) in the subject pose matrix $S_i(Y,P,R)$ in which $S_i(Y,P,R)$ is a cubic matrix with a size of x-by-x-by-x and total array of $S_i(Y,P,R)$ is $x{\times}x{\times}x = x^3$. In fact, arrays of $S_i(Y,P,R)$ matrix for subject i are arranged based on triplet angles of face pose form, where i indicates the number of subject and Y, P and R are the number of array in the corresponding dimension (yaw, pitch and roll) of the $S_i(Y,P,R)$ matrix. To cover all poses of face, range of yaw, pitch and roll angles are between −90 and +90 degrees. Therefore, sizes of $S_i(Y,P,R)$ are $x = 180/s = 180/5 = 36$, in which s is step size and 180 degrees is cover pose angle for each direction of face pose. For example, $S_i(Y = 8, P = 3, R = 4)$ indicates that the pose angle of the face is yaw $= 8 \times 5(s = 5)= 40$, pitch $= 3 \times 5 = 15$ and roll $= 4 \times 5 = 20$ degrees.