Pattern Recognition 46 (2013) 1511–1521

Contents lists available at SciVerse ScienceDirect

Pattern Recognition

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

Pose-robust face recognition via sparse representation

Haichao Zhang [a,*], Yanning Zhang [a,*], Thomas S. Huang [b]

[a] School of Computer Science, Northwestern Polytechnical University (Chang'an Campus), No. 1 Dong Xiang Road, Xi'an 710129, PR China
[b] Beckman Institute, University of Illinois at Urbana-Champaign, IL, USA

ARTICLE INFO

Article history:
Received 3 March 2012
Received in revised form
18 August 2012
Accepted 20 October 2012
Available online 8 November 2012

Keywords:
Pose-robust face recognition
Pose-normalization
Face synthesis
Sparse representation
Latent sparse modeling

ABSTRACT

We propose a pose-robust face recognition method to handle the challenging task of face recognition in the presence of large pose difference between gallery and probe faces. The proposed method exploits the sparse property of the representation coefficients of a face image over its corresponding view-dictionary. By assuming the representation coefficients are invariant to pose, we can synthesize for the probe image a novel face image which has smaller pose difference with the gallery faces. Furthermore, face recognition in the presence of pose variations is achieved based on the synthesized face image again via sparse representation. Extensive experiments on CMU Multi-PIE face database are conducted to verify the efficacy of the proposed method.

1. Introduction

Person identification is of vital importance in many applications such as surveillance. Due to the rich information contained in face images, face recognition is one of the most important approach in person identification. Much work on face recognition has focused on using a single probe face image for identification. One category is the subspace based methods such as Eigenface and Fisherface based method [1–4]. These methods seek for the most representative subspace for dimension reduction and feature extraction. Nearest neighbor (NN) method is one of the most simple and intuitive method for face recognition. NN classifies the probe face based on the best representation using a single training sample, which is essentially a point-to-point classification method. The nearest subspace (NS) method [5] generalizes NN method in the sense that it classifies the test sample based on the best linear representation in terms of all the training samples in each class. The sparse representation-based classification (SRC) method [6] is a further generalization of NS by representing the test sample using the training samples adaptively selected from all possible supports, both within and across different classes. NN, NS, and SRC are all *single* image based face recognition methods. These algorithms perform well when the training and testing faces are with similar poses, but they are limited by their sensitivity to pose variation of the probe face image, which is very common in real-world scenarios. One common problem is that the variation of the pose may cause changes larger than that caused by variation of identity in the observation space. In other words, the face images of different individuals from the same pose appears much closer than the images of the same individual from different poses. A schematic illustration is given in Fig. 1. The observed face images are modeled as vectors in the 'Observed Space' and we assume that there is a 'Latent Space' for explaining the observed images. Each observed face images corresponds to a latent representation in the latent space. Images of the same person observed from different views correspond to the same latent representation. For example, $\mathbf{x}_3$ and $\mathbf{x}_4$ are the images of the same person from different viewpoints and correspond to the same latent representation $\boldsymbol{\alpha}_3$. However, the distance between them is larger than the distance from $\mathbf{x}_4$ to the image of another identity from similar viewpoints, *i.e.*, $\mathbf{x}_1$. Therefore, the conventional distance-based face recognition methods will suffer greatly from this problem and it is desirable to develop face recognition methods that is robust to the pose variations between the probe and gallery faces.

Many different approaches for solving this problem have already been proposed in the literature, which can be roughly summarized in the following categories:

(1) *Multi-view approach*: Expend the testing and (or) training set to encompass more information, which can form a relatively more robust feature set to the undesirable factor, as has been used in the multi-view face recognition approaches.
(2) *Invariant approach*: Perform some particular transformation to eliminate the variations that align with the undesirable

* Corresponding authors. Tel.: +86 29 88431536; fax: +86 29 88491142.
E-mail addresses: hczhang1@gmail.com, hczhang@mail.nwpu.edu.cn (H. Zhang), ynzhang@nwpu.edu.cn (Y. Zhang), huang@ifp.uiuc.edu (T.S. Huang).

0031-3203/$ - see front matter © 2012 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.patcog.2012.10.025

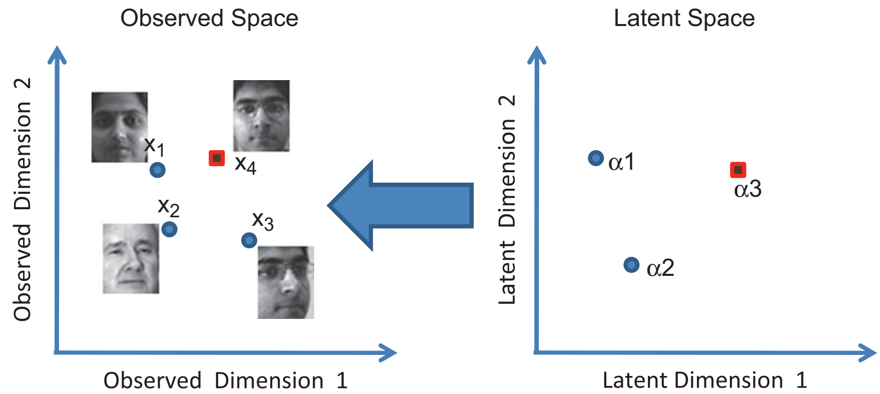

Fig. 1. The latent space representation of the multi-view face images.

factor or to reduce its adverse affect on the final recognition, as has been used in alleviating the illumination problem by removing the largest eigenvector in the eigenspace or by illumination normalization.

These two approaches are discussed briefly in the sequel.

(1) *Multi-view approach*: One category of algorithms for reducing the impact of pose variation on face recognition is using multiple face images of the same subject for recognition, which can potentially improve the robustness of the recognition system to different kinds of variations. Several different schemes have been developed in the literature. In [7], a single view image obtained via the weighted average of the multi-view face images is used for achieving multi-view recognition. Inspired by label propagation [8], a graph-based face recognition method using image set is proposed in [9], which first constructs a similarity graph between the images in the input set and those in the training set and then applies a class-wise graph matching procedure for joint identification. Another set-based face recognition method is proposed in [10], where each face set is modeled by affine hulls (or convex hulls) spanned by the samples in the set and the classification decision is based on the distance from the affine hull of observations to that of training samples from each class. Among all the approaches, one of the most well-known set based face recognition approach is the mutual subspace method (MSM) [11], which models the face set as a point on a Grassmann manifold [12], thus converting the task of comparing set-to-set distance to point-to-point distance on the Grassmann manifold, where distance based classifiers can be applied. More recently, a method exploiting the correlation as well as complementary information among multiple views by exploiting a novel structured sparsity pattern among the representation coefficients has been proposed in [13–15]. Although the multi-view based face recognition approaches have been proved to be effective for handling different factors such as pose variation, it has some limitations. One of the limitations is in the case of providing only a single probe view. In this case, the multi-view approach reduces to the distance-based single view face recognition, thus suffering again from the pose variations. In this sense, the multi-view approach basically just improves its performance by acquiring more data, and is not based on inherent improvement of its robustness to poses.

(2) *Invariant approach*: Apart from the multiple view-based approach, another direction is pose-invariant/robust approach with a single non-frontal view. One representative scheme in this category is pose-normalization. In this approach, given a non-frontal probe face image, its frontal image is first generated from the given non-frontal face image and the synthesized frontal image is then used for recognition via a classifier trained on frontal faces. A pose-normalization method with the help of a 3-D model is presented in [16]. In this method, the 3-D face model is adapted to the given face image. Then images from novel poses can be generated by projecting the 3-D model onto a 2-D plane from a proper viewpoint. With this technique in face recognition, given the non-frontal probe face image, we can first generate its frontal view and then use the generated view for recognition, *e.g.*, [17,18]. Apart from the 3-D model aided approach, methods with example-based learning scheme have also been developed, by treating the problem of pose-normalization as the problem of learning the regression function. In [19], the author proposed an approach by representing each face with templates from multiple views with different faces. The task of recognition is then purely searching for the best match across both identity and poses. Another approach based on linear object class (LOC) is presented in [20]. This approach is based on the fact that for linear object classes, linear transformations can be learned exactly from a basis set of 2-D prototypical views. The successful application of this approach for 'rotating' the face images requires high accuracy of registration. Recently, some regression-based approaches have been developed for learning the mapping for normalizing the pose of the face images. A method based on linear regression for generating the virtual frontal view from a given non-frontal face image is proposed in [21]. This method formulates the task of pose-normalization into a linear regression framework for learning the linear mapping. This approach is further improved by dividing the non-frontal face images into several partitions and then performs linear regression-based mapping for each of them separately. A hybrid-PCA method is developed in [22], which is based on a similar assumption as the linear regression model but uses subspaces learned from the training data via principal component analysis (PCA) in a hybrid way on the concatenated multi-view images. Recently, a partial least squares based method is proposed in [23] which can also handle pose variations in face recognition, by transforming images from different poses into a common space via linear projection. The linear projects are obtained via partial least squares.

In this paper, we propose a novel pose-invariant face recognition method via sparse representation, which lies in the second category as above. A schematic flowchart of the proposed pose-robust face recognition method is shown in Fig. 2. As shown in Fig. 2, the proposed method includes two phases:

(1) *Pose-normalization phase*: The proposed method can transform the pose of the given non-frontal view based on the assumption that the representation for the face images in the latent identity space is invariant to poses. Therefore, we can first recover this latent representation, and then transfer it to other poses for generating novel views. Specifically, the