Applied Mathematics and Computation 320 (2018) 99–115

Contents lists available at ScienceDirect

Applied Mathematics and Computation

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

A single gallery-based face recognition using extended joint sparse representation

Kun Shang [a], Zheng-Hai Huang [b], Wanquan Liu [c], Zhi-Ming Li [d,*]

[a] Center for Applied Mathematics, Tianjin University, Tianjin, PR China
[b] School of Mathematics, Tianjin University, Tianjin, PR China
[c] Department of Computing, Curtin University, Perth, WA, Australia
[d] Faculty of Applied Math, Shanxi University of Finance and Economics, Taiyuan, PR China

ARTICLE INFO

Keywords:
Face recognition
Single image to image set
Customized dictionary
Dictionary learning
Extended joint sparse representation

ABSTRACT

For many practical face recognition problems, such as law enforcement, e-passport, ID card identification, and video surveillance, there is usually only a single sample per person enrolled for training, meanwhile the probe samples can usually be captured on the spot, it is possible to collect multiple face images per person. This is a new face recognition problem with many challenges, and we name it as the single-image-to-image-set face recognition problem (ISFR). In this paper, a customized dictionary-based face recognition approach is proposed to solve this problem using the extended joint sparse representation. We first learn a customized variation dictionary from the on-location probing face images, and then propose the extended joint sparse representation, which utilizes the information of both the customized dictionary and the gallery samples, to classify the probe samples. Finally we compare the proposed method with the related methods on several popular face databases, including Yale, AR, CMU-PIE, Georgia, Multi-PIE and LFW databases. The experimental results show that the proposed method outperforms most of these popular face recognition methods for the ISFR problem.

1. Introduction

Face recognition (FR) is an active research topic in computer vision and pattern recognition [1–6]. Various demands of applications, such as law enforcement, e-passport, ID card identification, video surveillance, access control, social network, photo management, criminal investigation, etc., lead to a wide range of solutions for FR. Over the past decades, many appearance-based methods were proposed to improve the performance of face recognition. With the increasing attention from researchers, many methods have been proposed in the literature, such as principle component analysis (PCA) [7,8], linear discriminant analysis (LDA) [9], independent component analysis (ICA) [10], sparse representation classification (SRC) [11], kernel sparse representation (KSR) [12], linear regression (LR) [13], collaborative representation classification (CRC) [14], locality-constrained collaborative representation (LCCR) [15], manifold constraints transfer (MCT) [16] and so on. All these methods are in one framework that many face samples per person are used for training and a face sample is used for testing. These approaches can achieve state-of-the-art results when the training samples are as large as possible, especially with deep learning technique. We call this category as the image-set-to-image face recognition (SIFR).

* Corresponding author.
E-mail addresses: skun@tju.edu.cn (K. Shang), huangzhenghai@tju.edu.cn (Z.-H. Huang), w.liu@curtin.edu.au (W. Liu), li_zm@sxufe.edu.cn, lizm@tju.edu.cn, 11lizm@163.com (Z.-M. Li).

https://doi.org/10.1016/j.amc.2017.07.058
0096-3003/

With the rapid development of digital imaging and communication technologies, the image-set-to-image-set face recognition (SSFR) becomes a very important research topic for video surveillance and has attracted much intention in research community. Recently, a number of approaches [4,17–23] were proposed to solve the SSFR problem. Different from conventional SIFR where the probe is single, SSFR assumes that the gallery set and the probe set both have multi samples. All the samples are captured with different poses, illuminations and expressions. These face nuisances will affect the classification in the SSFR problem. Therefore, the key issues in SSFR include how to model a set and compute the distance/similarity between probe and gallery sets effectively. Researchers have proposed subspace [24–26], manifold [17,19,22], affine or convex hull [4,20,21] and dictionary learning [27,28] with attempt to achieve a satisfactory solution.

Unfortunately, sometimes there is only a single sample per person (SSPP) for training due to difficulty of collecting the sample with ID information. And in this case, many existing face recognition methods (both SIFR and SSFR methods) may fail to work because there are not sufficient samples for training. For the conventional SSPP face recognition [29,30], there is one sample per person for training and one sample per person for probing. We call this category as single-image-to-single-image face recognition (IIFR). This IIFR problem has attracted much attention in computer vision community due to its difficulty and several kinds of efficient methods [31–38] were proposed in the past. These methods are based on the generic learning, which assumes that the generic training set and the gallery set share similar variation information of both inter-class and intra-class. Especially, the dictionary learning methods, such as extended sparse representation-based classification (ESRC) [33,34], sparse variation dictionary learning (SVDL) [35], sparse illumination learning and transfer (SILT) [36], variational feature representation-based classification (VFRC) [37], would learn a dictionary from an additional generic set to offer the extra information, including illumination, expression, occlusion, and pose.

What may be less obvious is that, in the real world, the probe samples usually can be captured easily on the spot, and it is possible to collect multiple face images per person. The IIFR methods ignore the collection of the multiple probe samples, which would have potentially useful information to improve the performance of FR. This is a special SSPP face recognition problem, named as the single-image-to-image-set face recognition (ISFR). In this case, there are multiple probe samples per person in the testing phase, and only one gallery sample per person in the training phase. This framework is new and more suitable for many practical applications. In the case of ISFR, it arises an essential question for this application scenario: how can we use the multiple testing face images to improve the performance in ISFR? In 2013, Lu et al. [39] proposed the locality repulsion projections and sparse reconstruction-based similarity measure (LRP-SRSM) method to solve it. From the metric learning perspective, Zhu et al. [5] proposed the point-to-set distance metric learning method to learn a proper metric between a single image and image set in Euclidean space with an aim to achieve more accurate classification. From the manifold learning perspective, Huang et al. [40] proposed the learning Euclidean-to-Riemannian metric method. They think the single image is a point lying in Euclidean space, and the image set reside on certain Riemannian manifolds, and build a bridge between them. But all the above methods ignore the specific contents (the various uncontrolled variations, such as pose, illumination and expression) of the face images on the shot. They do not make full use of the information in the observation data.

Inspired by the works of the sparse representation methods and the dictionary learning methods, we propose a new method, named customized dictionary-based face recognition with extended joint sparse representation (CD-EJSR), to solve the ISFR problem. First, each customized dictionary is obtained on the shot by using the samples corresponding the same probe subject. In other words, every probe subject would have a special dictionary. The learned dictionary contains the variation features about the uncontrolled variations (pose, illumination and expression). Then, we propose the extended joint sparse representation (EJSR), which utilizes the information of both the customized dictionary and the gallery samples to classify the probe samples. In summary, we can highlight the contributions of this work as follows:

- Different from the conventional dictionary learning methods [33,35,36] for the SSPP problem (e.g. IIFR), in which the dictionaries are learned from the gallery samples and generic samples, the variation dictionary in CD-EJSR is learned directly from the observed probe samples without identity information.
- The variation dictionary is learned by using a new optimization model, which can be solved by the alternating direction method of multipliers (ADMM) approach. Also the closed-form solution is obtained in each step, which makes the proposed algorithm converge fast.
- We propose the extended joint sparse representation (EJSR) model. The EJSR model not only takes advantage of the learned variation dictionary which represents the intra-class variation between the gallery and probe samples, but also utilizes the group structure to enhance the performance for recognition.

The rest of this paper is organized as follows. Section 2 discusses the proposed customized dictionary learning and the extended joint sparse representation in detail. In Section 3, the experiments on several face databases are presented. The final section gives our conclusions for this paper.

2. Models

For the ISFR problem, there is only one single image for training and multiple samples for testing per person. The great difference (intra-class variation) between one training image and variational testing images becomes a huge barrier to recognize the identity of testing set. In order to reduce the barrier, we firstly propose a novel dictionary learning model to represent the intra-class variations of each probe subject. In fact, the dictionary just relies on the observation data (the