J. Vis. Commun. Image R. 30 (2015) 35–45

Contents lists available at ScienceDirect

J. Vis. Commun. Image R.

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

Variational Feature Representation-based Classification for face recognition with single sample per person ☆

Ru-Xi Ding [a,b], Daniel K. Du [a,c], Zheng-Hai Huang [a,b], Zhi-Ming Li [a,b,*], Kun Shang [a]

[a] Center for Applied Mathematics of Tianjin University, Tianjin 300072, PR China
[b] Department of Mathematics, School of Science, Tianjin University, Tianjin 300072, PR China
[c] School of Electronic Information Engineering, Tianjin University, Tianjin 300072, PR China

ARTICLE INFO

Article history:
Received 6 September 2014
Accepted 4 March 2015
Available online 21 March 2015

Keywords:
Face recognition
Single sample per person
Generic learning
Variational Feature Representation
Face image
Normal feature
Linear regression
Non-ideal conditions

ABSTRACT

The single sample per person (SSPP) problem is of great importance for real-world face recognition systems. In SSPP scenario, there is always a large gap between a normal sample enrolled in the gallery set and the non-ideal probe sample. It is a crucial step for face recognition with SSPP to bridge the gap between the ideal and non-ideal samples. For this purpose, we propose a Variational Feature Representation-based Classification (VFRC) method, which employs the linear regression model to fit the variational information of a non-ideal probe sample with respect to an ideal gallery sample. Thus, a corresponding normal feature, which reserve the identity information of the probe sample, is obtained. A combination of the normal feature and the probe sample is used, which makes VFRC method more robust and effective for SSPP scenario. The experimental results show that VFRC method possesses higher recognition rate than other related face recognition methods.

1. Introduction

In modern society, face recognition (FR) technology has become more and more popular in many application fields [1,2], such as information security, law enforcement and surveillance, smart cards, and access control. With the increased attention from researchers, many methods have been proposed in the literature [3–10]. However, there are still many challenges need to be faced in FR field. One of the challenges is single sample per person (SSPP) problem [11,12], especially for the situation that the probe samples contain large appearance variations caused by illumination, expression, age, pose, and so on. For some FR applications, such as law enhancement, e-passport and ID card, there is usually only a single sample per person recorded in the systems. The main reasons lie in two aspects: on the one hand, it is difficult for those FR systems to collect additional samples under many scenarios; and on the other hand, it is need to be considered to reduce the cost of storage space. As the probe samples may contain various varying information, numerous appearance-based multi-sample methods (e.g., Eigenfaces [13], Fisherfaces [14], Local Binary Pattern [15]) may be noneffective or may fail to work for SSPP issue. Since SSPP problem with non-ideal conditions is a big challenge in FR field, many multi-sample methods have been developed to address SSPP problem. At the same time, a variety of novel specific methods have been created by many researchers [11]. The methods for SSPP problem proposed in the literature can be classified the following three categories.

The first category of methods is based on virtual sample generation. In order to make the discriminative subspace learning methods adjustable for extracting feature with SSPP problem, some additional training samples, which are virtually generated, are added into each class of the gallery set. In [16,17], two singular value decomposition (SVD)-based perturbation algorithms were proposed to make the conventional Linear Discriminant Analysis (LDA) [18] available for SSPP issue by generating multiple images for each person. However, the distinguish information of the virtual samples in each class was not increased with the increasing number of virtual samples. Such an evident shortcoming comes from high correlation among the virtual samples and the corresponding original sample.

The second category of methods is based on image partitioning, as Modular Principal Component Analysis (Block-PCA) [19], Block based Linear Discriminant Analysis (Block-LDA) [20], Discriminative Multimanifold Analysis (DMMA) [21] and Multi-feature Multi-Manifold Learning (M^3L)[22]. In Block-PCA and

☆ This paper has been recommended for acceptance by Prof. M.T. Sun.

* Corresponding author at: Department of Mathematics, School of Science, Tianjin University, Tianjin 300072, PR China. Fax: +86 22 27403615.

E-mail addresses: dingruxi@tju.edu.cn (R.-X. Ding), daniel@tju.edu.cn (D.K. Du), huangzhenghai@tju.edu.cn (Z.-H. Huang), lizm@tju.edu.cn (Z.-M. Li), skun@tju.edu.cn (K. Shang).

http://dx.doi.org/10.1016/j.jvcir.2015.03.001
1047-3203/

Block-LDA, patch skill is used to divided face images into small local patches, on which the discriminant learning techniques are applied. In DMMA [21], Lu et. al. used the manifold theory to divided one person manifold to multi-manifold and compared the similarities between the local patch-manifold per person. Yan et al. proposed M^3L by extending DMMA with respect to the feature representation. In M^3L, not only the raw intensity feature, but also LBP and Gabor features were extracted within each small patch. Nevertheless, the form and size of the patch have big influences on the recognition results.

The last category of methods is based on generic learning [23–25]. The discriminative features, which are learned from an additional generic training set with multi-samples per person, are used to recognize the probe person. The methods with generic learning are based on the assumption that the generic training set and SSPP gallery set share similar variation information of both inter-class and intra-class. Su et al. [25] proposed an adaptive generic learning (AGL) method to successfully apply LDA to solve SSPP problem. Instead of directly employing the discriminatory information (e.g., the mean and covariance of each class) learned from the generic set, AGL method adapts it to predict intra-personal variations and mean for each subject enrolled in the gallery through least square regression. Then, the next step is to estimate the total intra-class and inter-class scatter matrix of all subjects in the gallery set. The classical sparse representation based classification (SRC) [26–28] can represent the probe samples well, when there are enough training samples per person. In addition, the representation ability of SRC is much more limited when it faces SSPP problem. Deng et al. [29,30] presented an effective method, named Extended SRC (ESRC), to overcome the disadvantage of SRC for SSPP. ESRC applies an auxiliary generic variant dictionary to represent the possible intra-class variation between the gallery and probe samples. Though ESRC performs well for SSPP problem, both the extraction of variational feature and the computational time are wildly discussed by researchers. Recently, an effective method, called Sparse Variation Dictionary Learning (SVDL) [31] method, was proposed for FR with SSPP. Instead of learning from the auxiliary generic set independently, SVDL method obtains a projection by learning from both auxiliary generic and gallery sets. The learned sparse variation dictionary plays an important role in handling all kinds of variations in face samples, including illumination, expression, occlusion, and pose. SVDL requires a sufficiently large generic training set, in which each subject should contain multiple face images with all type of variations.

In this paper, we propose a new generic learning method, named Variational Feature Representation-based Classification (briefly noted by VFRC), to solve SSPP problem with various non-ideal conditions (e.g., illumination, expression, occlusion, pose, age and comprehensive situations). Unlike the conventional generic learning methods, such as AGL, ESRC, and SVDL, the proposed VFRC does not directly employ the discriminatory information learned from the generic set. Alternatively, in the proposed VFRC model, the variational feature of probe sample is represented by the joint information of the generic set and the gallery set. As the gallery set can be regarded as a basic reference substance in the representation for variational feature, the rest normal feature of probe sample can be obtained more precisely. For the reason that the normal feature keeps less variational information, it is beneficial to enhance the identity information of the corresponding probe sample and makes a great contribution to the accuracy and robustness of VFRC with complicated, broad changing variations in SSPP scenario.

In order to well verify the effectiveness of the proposed VFRC, two popular image-to-image and image-to-set experiments are implemented. For image-to-image experiments, we compare the proposed VFRC with seven related methods (SRC [26], CRC [32], Block-LDA [20], DMMA [21], AGL [25], ESRC [29], and SVDL [31]) on three public face databases (AR [33], Extended Yale B [34,35], CMU-PIE [36] databases) with various variations, including illumination, expression, pose, and occlusion. For the image-to-set experiments, we choose one of the most challenging face database LFW [37] to verify the effectiveness of our method. The seven methods above and Locality Repulsion Projections and Sparse Reconstruction based Similarity Measure (LRP-SRSM) [38] are chose to compare with VFRC for image-to-set experiments. Experimental results demonstrate that the proposed VFRC achieves state-of-the-art performance for FR under the SSPP scenario with multiple non-ideal conditions.

The remainder of this paper is organized as follows. Section 2 presents our method for SSPP problem. In Section 3, the experimental results on three face databases are presented. The final section is the conclusion of this paper.

2. Proposed approach

It is well known that FR with SSPP problem has two natural characters: the first is that the images enrolled in the gallery set are usually frontal pose with ideal conditions (such as natural expression and illumination without occlusion), and the second is that the images from the probe set can possess kinds of non-ideal conditions (such as varying expression, illumination, pose, and occlusion). Hence, how to build a bridge between the normal gallery image and the non-ideal probe image is a key point for improving the performance of FR methods for solving SSPP problem. In the following subsections, we provide a feasible scheme for this problem.

2.1. Normal feature and variational feature of a face image

The face images we get from real world may contain many variational information, such as illumination, expression, occlusion, and pose. A face image can be represented by the normal face and the variational information, which is written as the following form:

$$I = N + V,$$

where $I \in \mathbb{R}^d$ $(d = m \cdot n)$ is the vectorization of face image with size $m \times n$; $N \in \mathbb{R}^d$ denotes the normal face feature; and $V \in \mathbb{R}^d$ denotes the variational feature. Thus, we divide a probe sample y into two subparts: the normal feature y_n and the variational feature y_v, formally denoted as

$$y = y_n + y_v, \tag{1}$$

where y_v contains some interference factors, which may mislead the identity recognition of the non-ideal probe sample; and y_n possesses the inherent identity information of the probe sample.

In SSPP scenario, $A = [x_1, x_2, \ldots, x_C]$ denotes the gallery set, which contains C different individuals with only one ideal sample per subject, $x_i \in \mathbb{R}^d$ is the vectorization of the i-th enrolled subject. As the most interference factors consist in the variational feature y_v, the normal feature y_n is closer to the corresponding ideal sample in the gallery set than the non-ideal probe sample y. And there is a strong correlation between two samples from the same subject, the normal feature y_n of the probe sample y could be represented as

$$y_n = \alpha_i x_i, \tag{2}$$

where x_i (the i-th sample in the gallery set) possesses the same identity as the probe sample y, and α_i is the representation coefficient of y_n over x_i. Because the identity of the probe sample is initially unknown, the representation of y_n can be rewritten as