Pattern Recognition 50 (2016) 1–16

Contents lists available at ScienceDirect

Pattern Recognition

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

Learning discriminative singular value decomposition representation for face recognition

Ying Tai, Jian Yang *, Lei Luo, Fanlong Zhang, Jianjun Qian

School of Computer Science and Engineering, Nanjing University of Science and Technology, Nanjing 210094, PR China

ARTICLE INFO

Article history:
Received 16 September 2014
Received in revised form
10 July 2015
Accepted 11 August 2015
Available online 21 August 2015

Keywords:
Face representation
Singular value decomposition (SVD)
Learning discriminative singular value decomposition representation (LDSVDR)
Sequential quadratic programming (SQP)

ABSTRACT

Face representation is a critical step in face recognition. Recently, singular value decomposition (SVD) based representation methods have attracted researchers' attentions for their power of alleviating the facial variations. The SVD representation reveals that the SVD basis set is important for the recognition purpose and the corresponding singular values (SVs) are regulated to form a more effective representation image. However, there exists a common problem in the existing SVD based representation methods: they all empirically make a rule to regulate the SVs, which is obviously not optimal in theory. To address this problem, in this paper, we propose a novel method named learning discriminative singular value decomposition representation (LDSVDR) for face recognition. We build an individual SVD basis set for each image and then learn a common set of SVs by taking account of the information in the basis sets according to a discriminant criterion across the training images. The proposed model is solved by sequential quadratic programming (SQP) method. Extensive experiments are conducted on three popular face databases and the results demonstrate the effectiveness of our method when dealing with variations of illumination, occlusion, disguise and face sketch recognition task.

1. Introduction

Face recognition is a classical topic in computer vision and pattern recognition community for its great need in many areas, such as access control, human–machine interaction, law enforcement, surveillance and so on [1,27]. Although great progress has been made by many researchers, it is still a challenging problem because of the large variations existed in the face images, e.g., variations in illumination conditions, poses, facial expressions and various noises (i.e. occlusion, corruption and disguise).

In the past few decades, subspace-based methods [4–10] have been received wide attention. As two most well-known subspace-based methods, Principal Component Analysis (PCA) and Linear Discriminant Analysis (LDA) have achieved great success. As we know, PCA is a method designed to model linear variation in high-dimensional data. The purpose of PCA is to find the optimal linear projection that captures the directions of maximum variance in the data. In [4], Kirby et al. did the first attempt to apply PCA in face recognition and the well-known Eigenfaces method was proposed by Turk and Pentland [6]. Differing from PCA, LDA is a supervised learning method, which follows the criterion that the ratio of the between-class scatter and the within-class scatter is maximized. In [8], Belhumenur et al. proposed the well-known Fisherfaces method, which applied PCA procedure before the LDA procedure to avoid the small sample size problem. There are also some derivative methods [7,30–32], which make a progress over the two classical methods. However, the subspace-based methods are very sensitive to large variations existed in face images.

Recently, several singular value decomposition (SVD) based representation methods have been proposed due to their ability to effectively alleviate the facial variations. Liu et al. [15] proposed a fractional order singular value decomposition representation (FSVDR) method, which comes from the observation that the leading SVD bases are sensitive to the facial variations. They applied a fractional function to deflate the weights of the facial variation sensitive bases and inflate the weights of the facial variation insensitive bases. However, since the performance of FSVDR relies on the fractional order parameter α, which is generated through an exhaustive strategy, it is unsuitable in real-world applications. Additionally, FSVDR considers the reconstructive power of each basis and utilizes all of the bases in face representation. Actually, those lagging bases are insensitive to the facial variations and contain little reconstructive and discriminative information, which may be regarded as the noise components and even affect the recognition performance. From this viewpoint, Lu et al. [20] proposed a dominant singular value decomposition representation (DSVDR) method. Unlike FSVDR that only uses the reconstructive power of bases, DSVDR decomposes the singular value spectrum of each face image into three subspaces and regulates the singular values (SVs) of those important bases according to their

* Corresponding author.

http://dx.doi.org/10.1016/j.patcog.2015.08.010
0031-3203/

discriminative and reconstructive power simultaneously. More recently, Zhang et al. [13] proposed a simple but effective method named nearest orthogonal matrix representation (NOMR). In NOMR, the nearest orthogonal matrix of each image is calculated as its SVD representation, which keeps the original basis set but changes all of the SVs to be 1. The authors consider the individual basis space corresponds to the essence identity information and the singular values associate with illumination variations. Therefore, they think the SVs are not suitable for face recognition and directly advocate the basis set generated via SVD to identify the original face image. Although NOMR achieves some interesting results for alleviating the effect of illumination and heterogeneity, the way replacing all non-zero singular values with 1 not only deflates the variations in the leading sensitive bases but also the discriminant information contained in the leading image. Therefore, NOMR may be unsuitable in some cases. Actually, NOMR can be seen as a special case in FSVDR. When $\alpha = 0$, all of the SVs are regulated as 1, then FSVDR is equivalent to NORM.

Generally speaking, these former SVD representations [13,15,20] empirically make a rule to regulate the SVs without utilizing the information in the SVD basis set, which is obviously not optimal in theory. In fact, SVD representation is composed of the SVs and the SVD basis set. Each SV specifies the luminance of the image layer while the corresponding basis specifies the geometry of the image layer [3]. That is to say, the basis set mostly contains inherent information of the original face image, which is dominant for face recognition. Therefore, in order to get better SVD representation, it seems more reasonable to regulate the SVs according to the information in the basis set. To address this problem, we propose a novel method named learning discriminative singular value decomposition representation (LDSVDR) for face recognition. Specifically, we build an individual SVD basis set for each image and then learn a common set of SVs by taking account of the information in the basis sets according to a discriminant criterion, derived from the fisher criterion [28], across the training images. Since [13,15,20] empirically regulate the SVs, it is difficult for these methods to handle different variations in different databases. However, in our proposed method, according to a discriminant criterion that maximizes the ratio of between-class distance to that of within-class distance, discriminative representations can be learnt via taking account of the variations, such as illumination, occlusion and disguise, in the training process, which brings the robustness to different variations. The proposed LDSVDR is solved via sequential quadratic programming (SQP) method [33,40]. Experiments on variations of illumination, different kinds of occlusions, face sketch recognition task and real disguises are conducted on three public face databases. Our proposed LDSVDR achieves the best results in most cases and is more stable than the other SVD representations under different variations.

The rest of this paper is organized as follows. Section 2 briefly reviews three related works. Section 3 presents the details of LDSVDR. Section 4 conducts extensive experiments to demonstrate the efficacy of our method and Section 5 offers our conclusions.

2. Related work

In this section, we briefly review FSVDR [15], DSVDR [20] and NOMR [13].

Let A be a $m \times n$ ($m \geq n$ without loss of generality) grayscale face image. Seen from Eq. (1), the singular value decomposition (SVD) technique can be defined as

$$A = \overline{U}\overline{S}\overline{V}^T, \tag{1}$$

where $\overline{U} = [u_1, \ldots, u_m] \in R^{m \times m}$ and $\overline{V} = [v_1, \ldots, v_n] \in R^{n \times n}$ are orthogonal matrices, $\overline{S} = [D \quad 0]^T$, $D = \text{diag}(\lambda_1, \ldots, \lambda_n)$, 0 is a $n \times (m-n)$ zero matrix and $\lambda_i, i = 1 : n$ is the SV of matrix A.

Let $k = \text{rank}(A)$, we can further represent A as

$$A = USV^T = \sum_{i=1}^{k} \lambda_i u_i v_i^T, \tag{2}$$

where $U = [u_1, \ldots, u_k] \in R^{m \times k}$, $V = [v_1, \ldots, v_k] \in R^{n \times k}$ and $S = \text{diag}(\lambda_1, \ldots, \lambda_k), \lambda_1 \geq \ldots \geq \lambda_k > 0$.

2.1. FSVDR

According to the observations that the leading bases are sensitive to the facial variations and dominate the composition of the face image, FSVDR [15] subtly deflates the weights of the facial variation sensitive bases via a fractional order function as follows:

$$B = US^{\alpha}V^T = \sum_{i=1}^{k} \lambda_i^{\alpha} u_i v_i^T, \tag{3}$$

where α is a fractional parameter and $0 \leq \alpha \leq 1$. The optimal value for parameter α should be different for different databases, which is obtained via an exhaustive search. Also, the authors offer a heuristic criterion for choosing the parameter α for the LDA-based methods.

2.2. DSVDR

Motivated by the fact that each basis contains different discriminative and reconstructive information for face representation, Lu et al. [20] selected a subset of important bases and regulated their SVs according to their discriminative and reconstructive power simultaneously. Specifically, they decompose the singular value spectrum of each image into three subspaces: the dominant space (D), the complementary space (C), and the null space (N), respectively, as

$$\lambda_i \in \begin{cases} D & \lambda_i \geq 1 \\ C & t_1 \leq \lambda_i < 1 \text{ and } f_i \geq t_2 f_1 \\ N & \lambda_i < t_1 \text{ or } f_i < t_2 f_1 \end{cases}, \tag{4}$$

where λ_i denotes the i-th SV of a face image, f_i denotes the recognition accuracy of the i-th basis, t_1 and t_2 are two parameters to balance the size of the complementary space and the null space. Let the sizes of D, C and N be c_1, c_2 and c_3. Then, the regularization of the SVs of the three subspaces are made as

$$\hat{\lambda}_i \in \begin{cases} f_i / f_{c_1} & 1 \leq i < c_1 \\ 1 & c_1 + 1 \leq i \leq c_1 + c_2 \\ 0 & c_1 + c_2 + 1 \leq i \leq c_1 + c_2 + c_3 \end{cases}. \tag{5}$$

Finally, the new face image is constructed as $B = \sum_{i=1}^{k} \hat{\lambda}_i u_i v_i^T$.

2.3. NOMR

In accordance with the fact that the individual basis space corresponds to the essence identity information and the SVs associate with the illumination variations. Zhang et al. [13] thought the SVs to be unsuitable for face recognition and directly advocated the basis set generated via SVD to identify the original face image. Specifically, the authors replace all of the SVs with 1: $\hat{S} = \text{diag}(1, \ldots, 1)$. Then, the new representation is formulated as $B = U\hat{S}V = \sum_{i=1}^{k} u_i v_i^T$.