J. Vis. Commun. Image R. 46 (2017) 260–268

Contents lists available at ScienceDirect

J. Vis. Commun. Image R.

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

Sparse representation for robust face recognition by dictionary decomposition ☆

Feilong Cao *, Xinshan Feng, Jianwei Zhao

Department of Applied Mathematics, College of Sciences, China Jiliang University, Hangzhou 310018, Zhejiang Province, China

ARTICLE INFO

Article history:
Received 2 May 2016
Revised 27 February 2017
Accepted 27 April 2017
Available online 2 May 2017

Keywords:
Sparse representation
Face recognition
Dictionary decomposition
Projection matrix

ABSTRACT

Sparse representation-based classification (SRC) method has gained great success in face recognition due to its encouraging and impressive performance. However, in SRC the data used to train or test are usually corrupted, and hence the performance is affected. This paper proposes a robust face recognition approach by means of learning a class-specific dictionary and a projection matrix. Firstly, the training data are decomposed into class-specific dictionary, non-class-specific dictionary, and sparse error matrix. Secondly, in order to correct the corrupted test data, the data are projected onto their corresponding underlying subspace, and a projection matrix between the original training data and the class-specific dictionary is learned. Then, the features of the class-specific dictionary and the corrected test data are extracted by using Eigenface method. Finally, the SRC is performed to classify. Extensive experiments conducted on publicly available data sets show that the proposed algorithm performs better than some state-of-the-art methods.

1. Introduction

Face recognition has been a classical and hot research topic in computer vision and pattern recognition due to its wide real-world applications, such as video surveillance, security-related access, and human-computer intelligent interaction [1–7]. However, face recognition is not a simple work. The limited face images, various variations such as facial expressions, poses, and illumination, and images be corrupted, occluded, or disguised, are the main challenges. In addition, handling the problems in the high dimensional feature space or on an undersampled dataset makes the task even more challenging.

To reduce the dimension of the face images and extract feature, some methods such as Eigenfaces [8], Fisherfaces [9], and Laplacianfaces [10], have been proposed. As a result, improved recognition performance is achieved due to the derived feature subspace. However, a common drawback of these methods is that they are not robust to the image corruption caused by outlier pixels, disguise, and occlusions.

Recently, sparse representation-based classification (SRC) method for face recognition has shown very encouraging and impressive performance [11]. The method considers each query image as a sparse linear combination of the training samples, and obtains the classification results by solving an ℓ_1-minimization problem and calculating the minimum residual error. Accordingly, a high recognition accuracy was achieved even if the test image is corrupted. However, while the training images used for SRC are not carefully controlled or corrupted, the method will result in poor performance [12,13].

As SRC method directly uses training images as the dictionary for the sparse representation, a crucial question is how to choose a proper dictionary for the sparse representation [14–16]. Furthermore, some extensions for the SRC have been proposed, which applied dictionary learning to find different optimal dictionaries in representing corresponding data [17,18]. However, even if a good representing dictionary usually does not have the best classification ability. So some methods have addressed a compromise between the data representation and the classification ability via learning a discriminative dictionary [19–22].

In recent years, low-rank matrix recovery theory [23–31], also called robust principal component analysis (RPCA) [32], has been used for image processing. Specifically speaking, the theory can be employed to separate outlier pixels and occlusions from the training data. That is, decomposing the training data of each class into a low-rank matrix and a sparse matrix to obtain a sub-dictionary. Furthermore, a method called low-rank representation (LRR) to remove the sparse corruptions of the training data was proposed in [33]. In [34] Zheng et al. integrated Fisher discriminant

☆ This paper has been recommended for acceptance by Zicheng Liu.

* Corresponding author.
E-mail address: feilongcao@gmail.com (F. Cao).

http://dx.doi.org/10.1016/j.jvcir.2017.04.007
1047-3203/

regularization involved with in-class scatter and between-class scatter, and formulated the face recognition problem under the low rank matrix recovery framework. They regularized the representative basis derived from standard low rank matrix recovery using the class-specific discriminant criterion motivated by Fisher criterion. Lu et al. [35] presented a general framework for solving the low-rank and sparse matrix minimization problems, which may involve multiple non-smooth terms and is used for the face recognition. However, these methods may undesirably remove some important face identity information and need high computational cost, which will affect recognition accuracy or limit the practical applicability.

As we know, there are identity information and other information in a face image. The identity information is crucial to face recognition, while the other information is likely to affect recognition accuracy. So it seems obvious that recognition performance can be improved if the identity information is separated from the original data. For this aim, Jiang and Lai [36] designed a dictionary decomposition algorithm and divided the original data into three parts: class-specific dictionary, non-class-specific dictionary, and sparse error matrix. Although the dictionary decomposition algorithm is cooperated with the sparse and dense hybrid representation model (SDR) proposed in [36], it has shown better performance than RPCA.

Thereout, while this dictionary decomposition algorithm is regarded as a method like the pretreatment of SRC, does it have better performance comparing with SRC or some extensions of SRC? This paper tries to answer the question. We will adopt the dictionary decomposition algorithm to capture the identity information (the class-specific dictionary), and similar to [37], we learn a projection matrix between the original training data and the class-specific dictionary. Then a corrupted test sample can be efficiently corrected by projecting it to the corresponding underlying subspace. Because the face images are usually high dimensional which cost too much calculation, we will reduce the dimension to a series of small numbers via principal component analysis (PCA) method. The experimental results demonstrate the advantages of the proposed approach over some classical algorithms and some state-of-the-art face recognition algorithms.

The remainder of this paper is organized as follows. Section 2 describes the face recognition problem to be solved and reviews the SRC approach. Section 3 introduces our algorithm in detail. Section 4 is devoted to experimental results and corresponding analysis and Section 5 concludes this paper.

2. Related works

Before reviewing SRC, we first describe the problem to be solved. Suppose that there is a training set $D = (d_1, d_2, d_3, \ldots, d_n) \in \mathbb{R}^{m\times n}$which belongs to N classes, the ith class has N_i images, and each element of the set is a training sample, where m is the dimension of each image and n is the total number of training samples. Given a test image $y \in \mathbb{R}^m$, the task is to determine which class y belongs to.

The SRC algorithm proposed by Wright et al. [11] exhibited impressive performance. It considers each query image as a sparse linear combination of the training samples, and obtains the representation coefficients by solving an ℓ_1-minimization problem. Then the query image is classified by evaluating the reconstruction error associated with each class. Some extended works have been proposed to further improve the performance of the SRC algorithm [38–40].

As mentioned above, given a test image $y \in \mathbb{R}^m$, the SRC algorithm calculates the sparse representation coefficients α of y via the following ℓ_1-minimization problem over the entire training image set:

$$\min_{\alpha} \|y - D\alpha\|_2 + \beta\|\alpha\|_1, \tag{1}$$

where β is a regularization constant for a compromise between the sparsity of α and the representation error. Let $\alpha_i \in \mathbb{R}^{N_i\times 1}$ be the entries of α associated with class i, i.e., $\alpha = [\alpha_1; \alpha_2; \alpha_3; \ldots; \alpha_N]$. We can obtain the minimum residual by solving the following problem:

$$\text{identity}(y) = \arg\min_{i} \|y - D_i\alpha_i\|_2. \tag{2}$$

The test image y will be assigned to the class with the minimum reconstruction error. This is because the test image y should lie in the space spanned by the training samples of the corresponding class. Although the SRC method can obtain impressive face recognition results even if the test images are corrupted, the SRC still needs carefully controlled clean training images and sufficient training samples of each class. Or it will result in a poor performance. Our method which will be proposed in the next section is robust for face recognition, even if both training and test data are corrupted.

3. Robust face recognition by dictionary decomposition

3.1. Dictionary decomposition

We know the prior knowledge in sparse representation theory that the signals are sparse or they are sparse in some transform domain. In face recognition, it means that the face images are sparse or they are sparse in some transform domain. As we know, there is some important information in a face image, which contains not only the identity information, but also much other information such as age, gender, expression, illumination, and so on. So the significant coefficients in α may correspond to the same expression or the same illumination between the query image and the training samples, rather than the same identity information which will result in the wrong classification [36]. To separate the class-specific information from others, Jiang and Lai [36] proposed a dictionary decomposition method. We adopt this method and decompose the original training data into a class-specific dictionary, a non-class-specific dictionary, and a sparse error matrix via dictionary decomposition, where the dictionary decomposition is based on the following model:

$$D = A + BX + E. \tag{3}$$

As mentioned above, D is the original training data matrix, A is the class-specific dictionary, B is the non-class-specific dictionary, E is the sparse error matrix, and X is the coefficient matrix of B. We notice that the training data are real face images, and can reasonably assume that D is a full rank matrix because of the existence of the random sparse noise. As D contains only a particular type of images (face images), $D - E$, i.e. $A + BX$ should be a low rank matrix. As $A + BX$ represents the complete clean face images, A represents the class-specific information of it and B represents the non-class-specific information of it(A and B are both part of face images), we can assume that $A + BX, A$ and B are all sparse or they are sparse in some transform domain. For simplicity, we directly assume that they are all sparse in the paper. Consequently, the dictionary decomposition (3) can be represented as the optimization problem:

$$\begin{aligned} &\min_{A,B,X,E} \text{rank}(A) + \lambda\text{rank}(B) + \tau\|X\|_F^2 + \eta\|E\|_0 \\ &\text{s.t. } D = A + BX + E, \end{aligned} \tag{4}$$

where λ, τ, and η are all positive parameters, the $\|X\|_F^2$ is the sum of the squared ℓ_2-norms, $\|x_k\|_2^2$. Both the minimization of rank and the minimization of ℓ_0-norm are NP-hard problems. But it is proved that