J. Vis. Commun. Image R. 25 (2014) 763–773

Contents lists available at ScienceDirect

J. Vis. Commun. Image R.

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

Sparse representation for face recognition by discriminative low-rank matrix recovery

Jie Chen, Zhang Yi *

Machine Intelligence Laboratory, College of Computer Science, Sichuan University, Chengdu 610065, PR China

ARTICLE INFO

Article history:
Received 20 August 2013
Accepted 26 January 2014
Available online 5 February 2014

Keywords:
Sparse representation
Low-rank representation
Matrix recovery
Dictionary learning
Face recognition
A low-rank projection matrix
Subspace
Eigenface

ABSTRACT

This paper proposes a discriminative low-rank representation (DLRR) method for face recognition in which both the training and test samples are corrupted owing to variations in occlusion and disguise. The proposed method extends the sparse representation-based classification algorithm by incorporating the low-rank structure of data representation. The DLRR algorithm recovers a clean dictionary with enhanced discrimination ability from the corrupted training samples for sparse representation. Simultaneously, it learns a low-rank projection matrix to correct corrupted test samples by projecting them onto their corresponding underlying subspaces. The dictionary elements from different classes are encouraged to be as independent as possible by regularizing the structural incoherence of the original training samples. This leads to a compact representation of a corrected test sample by a linear combination of more dictionary elements from the corrected class. The experimental results on benchmark databases show the effectiveness and robustness of our face recognition technique.

1. Introduction

As one of the most challenging research topics in computer vision and pattern recognition, face recognition (FR) [1–8] has been extensively studied over several decades. FR has significant value in developing an understanding of how people recognize each other. In addition, FR is largely motivated by the increasing need to address many challenging real-world applications, including video surveillance, security-related access, and human–computer intelligent interaction. Compared with other behavioral biometric techniques such as fingerprint, iris, or gait recognition, which typically require the cooperation of the individual, facial images can be captured either actively or passively by cameras.

An FR system usually involves two stages: first, features are extracted from given training images, and second, a corresponding classification model is learnt. The recognition of an unseen test image can be accomplished by classifying the features extracted from this face. For face representation and recognition, many representative methods, such as Eigenfaces [1], Fisherfaces [2], and Laplacianfaces [4], have been proposed. These derive a starting set of features from face images. Yang et al. [9] developed a novel criterion, namely Laplacian bidirectional maximum margin criterion (LBMMC), to consider the discriminative information within the local structures of samples and the structural information embedding in the images. Besides, Yang et al. [10,11] propose a Multi-Manifold Discriminant Analysis (MMDA) method for an image feature extraction and pattern recognition. LBMMC and MMDA have shown very promising results on image recognition. Whereas FR in controlled environments (frontal-face images of cooperative subjects with controlled illumination) has already achieved an impressive level of performance, robust FR remains a challenging issue in uncontrolled environments because of various corruptions, such as occlusion, disguise, and the intensity of occluded pixels.

Recently, inspired by advances in l_0-norm and l_1-norm techniques [12–18], there has been a growing trend to incorporate sparsity in various areas of computer vision and pattern recognition [5–8,19–23]. An FR methodology usually considers representation as well as classification problems. Sparse representation is the problem of finding a representation of a signal as a linear combination of only a small number of elementary signals from an over-complete dictionary [24]. Sparse representation, which has proved to be an extremely successful tool for representing and compressing high-dimensional signals, provides a statistical model for finding efficient and compact signal representations. Furthermore, it has been strongly supported by studies of the sparse coding mechanism of human vision [25].

Some recent work, such as that on sparse representation-based algorithms, has gained improved recognition performance. Wright et al. [5] proposed a method that used a sparse representation for robust FR. Such a sparse representation-based classification (SRC)

* Corresponding author.
E-mail address: zyiscu@gmail.com (Z. Yi).

http://dx.doi.org/10.1016/j.jvcir.2014.01.015
1047-3203/

scheme achieved high classification accuracy. Min and Dugelay [6] proposed an FR method that combines local binary pattern (LBP)-based feature extraction with SRC. This method applied the divide-and-conquer strategy to resolve the dimensionality problem, and exhibited an enhanced discriminative ability because of its pyramid architecture. Yang et al. [8] proposed a robust regularized coding (RRC) model to seek an approximate maximum a posterior solution of the sparse coding problem. Although SRC, LBP, and RRC achieve impressive results, these techniques are not robust enough for FR under occlusion and/or disguise. A crucial question when applying the above methods is how to choose a proper dictionary for sparse representation [26–29]. Obviously, a straightforward method is to use a prespecified dictionary that directly involves the original training samples. Unfortunately, gross occlusion and disguise are ubiquitous in practice. Consequently, performance is degraded in the presence of such corrupted samples. As later verified by our experiments, using contaminated training samples in the dictionary can lead to unsatisfactory results. On the contrary, disregarding corrupted training samples does not make full use of the discriminative information hidden in the original training samples, and thus results in poorer performance. In addition, there exists an intractable problem whereby some additional corrupted criteria are involved in removing the corrupted samples from training data in applications.

A number of approaches have been explored and proposed to address this issue. A more generalized method is to learn the dictionary, which has been proven to be an effective scheme for improving the robustness of sparse representation. Aharon et al. [26] presented the K-means singular value decomposition (K-SVD) algorithm to iteratively alternate the update of the sparse coefficients of the training samples, based on the current dictionary, and the dictionary columns in order to better fit the data. Zhang and Li [27] proposed the discriminative K-SVD algorithm, whose goal is to find the dictionary and solve for the classifier using the optimization procedure of the K-SVD algorithm. Ma et al. [28] presented a discriminative low-rank dictionary learning algorithm for sparse representation. These algorithms aim to learn a dictionary and reconstruct signals based on the clean training samples or the corrupted training samples in the presence of low-level noise. However, these techniques do not work well with grossly corrupted observations in real applications.

Recently, some work on robust principal component analysis (PCA) [30–34] has been proposed to alleviate the aforementioned drawbacks. In particular, Wright et al. [31] recently presented the robust PCA (RPCA) method, which aims to recover a low-rank matrix from corrupted observations based on the hypothesis that the underlying data structure is approximately drawn from a single low-rank subspace. Furthermore, Liu et al. [35,36] established a low-rank representation (LRR) method to remove the grossly sparse corruptions of data samples that are approximately drawn from a mixture of multiple low-rank subspaces. These approaches work well to recover a low-rank matrix from an observed data matrix that is low rank and contains sparse errors. However, considering a new corrupted sample, LRR essentially calls for a recalculation over all samples. This can incur a high computational cost, which does not generalize well for online computation. As a result, a high computational cost seriously limits the practical applicability of LRR. All of the above techniques are essentially transductive learning methods, which are inappropriate for classification tasks. Zhang et al. [37] proposed an image classification scheme, which uses the low-rank and sparse matrix decomposition techniques to get more discriminative bases for sparse representation. Zhang and Li [38] introduced a novel image decomposition method, which incorporates low-rank and sparsity constraints into a new representation model, for an ensemble of correlated images. The image decomposition seeks a common image and a low-rank image for each of the subjects in the dataset. Zhang et al. [39] presented a new image classification model, which performs low-rank matrix recovery for contaminated training samples from all classes simultaneously without losing structural information, to learn a structured low-rank representation. However, these methods cannot correct the corrrupted test images for image classification if the test images are corrupted. These methods might not generalize well. Chen et al. [7] proposed a low-rank matrix approximation algorithm with structural incoherence to improve the discriminative ability of SRC. This method decomposes the original training data into a representative basis and a sparse error matrix. The success of such a decomposition is often limited, because all samples are represented approximately with the use of dimension-reduction techniques, such as PCA, if they are not recovered faithfully. Bao et al. [40] presented an inductive robust PCA method to learn the underlying projection matrix. To some extent, this can be applied to recover original corrupted data samples. However, it is inappropriate for classification purposes, as the original training samples are projected on the underlying subspaces without exploring their discriminative capability.

In this paper, we address the problem of sparse representation-based robust FR by providing additional discrimination of the clean dictionary and a low-rank projection matrix, in which both the original training and test samples might be corrupted. As mentioned above, using off-the-shelf training samples in the presence of corrupted observations as the dictionary may bring about poorer FR performance. To overcome this problem, a regularization term is added to the LRR formulation. This promotes incoherence between different classes, meaning we can efficiently and accurately obtain the recovery results as a dictionary from a set of training samples. By introducing a discriminative low-rank representation (DLRR) method (in terms of the low-rank structure of data representation), the dictionary can lead to better representations by providing additional discriminative ability for classification tasks. Furthermore, by solving the nuclear norm minimization problem in a simple and efficient way, we can learn a low-rank projection between the training samples and the recovery results to correct the corrupted test samples. For example, any new corrupted test sample can be efficiently corrected by projecting it onto its underlying subspace. The proposed algorithm not only effectively recovers the training samples used as the dictionary for sparse representation, but also corrects grossly corrupted test samples for the purpose of multi-class classification.

Fig. 1 gives an intuitive example to illustrate our basic idea and its effectiveness. A corrected test image and its dictionary are obtained by correcting an original test image and recovering the original training images, respectively. Fig. 1(a) and (b) shows some examples of the original training images and the corresponding clean dictionary. The sparse representation coefficients and part of the class-wise SRC reconstruction errors are illustrated in Fig. 1(c) and (d), respectively. After introducing a DLRR, as suggested in our scheme, the test image can be represented by a smaller number of elements of a dictionary, compared with SRC, as shown in Fig. 1(c). Clearly, the class-wise reconstruction errors of our method, shown in Fig. 1(d), demonstrate its superior representation ability for classification tasks. Further details will be discussed in Section 3. Extensive experimental results on the benchmark databases demonstrate the effectiveness and robustness of DLRR for FR.

The proposed DLRR method provides several advantages:

(a) It makes the dictionary elements between different classes as independent as possible by promoting the structural incoherence of between-class scatter matrices. As a result, a query sample is represented by more training samples from the corrected class.