Pattern Recognition 47 (2014) 2794–2806

Contents lists available at ScienceDirect

Pattern Recognition

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

Learning locality-constrained collaborative representation for robust face recognition

Xi Peng [a], Lei Zhang [a,*], Zhang Yi [a], Kok Kiong Tan [b]

[a] Machine Intelligence Laboratory, College of Computer Science, Sichuan University, Chengdu 610065, China
[b] Department of Electrical and Computer Engineering, National University of Singapore, Engineering Drive 3, Singapore 117576, Singapore

ARTICLE INFO

Article history:
Received 23 October 2012
Received in revised form
6 September 2013
Accepted 18 March 2014
Available online 27 March 2014

Keywords:
Non-sparse representation
Sparse representation
Local consistency
ℓ_2-minimization
Partial occlusions
Additive noise
Non-additive noise
Robustness

ABSTRACT

The models of low-dimensional manifold and sparse representation are two well-known concise models that suggest that each data can be described by a few characteristics. Manifold learning is usually investigated for dimension reduction by preserving some expected local geometric structures from the original space into a low-dimensional one. The structures are generally determined by using pairwise distance, e.g., Euclidean distance. Alternatively, sparse representation denotes a data point as a linear combination of the points from the same subspace. In practical applications, however, the nearby points in terms of pairwise distance may not belong to the same subspace, and vice versa. Consequently, it is interesting and important to explore how to get a better representation by integrating these two models together. To this end, this paper proposes a novel coding algorithm, called Locality-Constrained Collaborative Representation (LCCR), which introduce a kind of local consistency into coding scheme to improve the discrimination of the representation. The locality term derives from a biologic observation that the similar inputs have similar codes. The objective function of LCCR has an analytical solution, and it does not involve local minima. The empirical studies based on several popular facial databases show that LCCR is promising in recognizing human faces with varying pose, expression and illumination, as well as various corruptions and occlusions.

1. Introduction

Sparse representation has become a powerful method to address problems in pattern recognition and computer version, which assumes that each data point $\mathbf{x} \in \mathbb{R}^m$ can be encoded as a linear combination of other points. In mathematically, $\mathbf{x} = \mathbf{D}\mathbf{a}$, where $\mathbf{D}$ is a dictionary whose columns consist of some data points, and $\mathbf{a}$ is the representation of $\mathbf{x}$ over $\mathbf{D}$. If most entries of $\mathbf{a}$ are zeros, then $\mathbf{a}$ is called a sparse representation. Generally, it can be achieved by solving

$$(P_0): \ \min \|\mathbf{a}\|_0 \quad \text{s.t. } \mathbf{x} = \mathbf{D}\mathbf{a},$$

where $\|\cdot\|_0$ denotes ℓ_0-norm by counting the number of nonzero entries in a vector. P_0 is difficult to solve since it is a NP-hard problem. Recently, compressive sensing theory [1,2] have found that the solution of P_0 is equivalent to that of ℓ_1-minimization problem $(P_{1,1})$ when $\mathbf{a}$ is highly sparse.

$$(P_{1,1}): \ \min \|\mathbf{a}\|_1 \quad \text{s.t. } \mathbf{x} = \mathbf{D}\mathbf{a},$$

where ℓ_1-norm $\|\cdot\|_1$ sums the absolute value of all entries in a vector. $P_{1,1}$ is convex and can be solved by a large amount of convex optimization methods, such as basis pursuit (BP) [3], least angle regression (LARS) [4]. In [5], Yang et al. make a comprehensive survey for some popular optimizers.

Benefiting from the emergence of compressed sensing theory, sparse coding has been widely used for various tasks, e.g., subspace learning [6], spectral clustering [7,8] and matrix factorization [9]. In these works, Wright et al. [10] reported a remarkable method that passes sparse representation through a nearest feature subspace classifier, named sparse representation-based classification (SRC). SRC has achieved attractive performance in robust face recognition and has motivated a large amount of works such as [11–13]. The work implies that sparse representation plays an important role in face recognition under the framework of nearest subspace classification [14].

* Corresponding author.
E-mail addresses: pangsaai@gmail.com (X. Peng), leizhang@scu.edu.cn (L. Zhang), zhangyi@scu.edu.cn (Z. Yi), kktan@nus.edu.sg (K.K. Tan).

http://dx.doi.org/10.1016/j.patcog.2014.03.013
0031-3203/

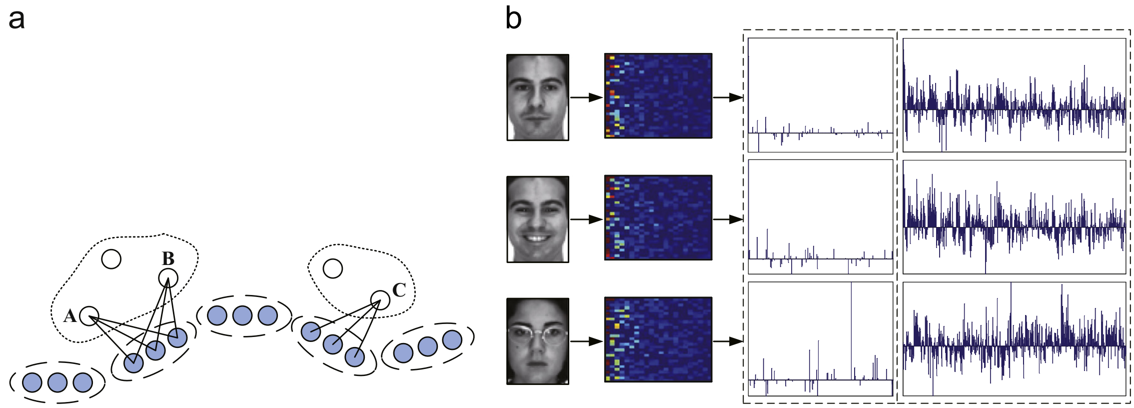

Fig. 1. A key observation. (a) Three face images from two different sub-manifolds are linked to their corresponding neighbors, respectively. (b) The first column includes three images which correspond to the points A, B and C in (a). The second column shows the Eigenface feature [31] matrices for the testing images; The third column includes two parts: the left part is the coefficients of SRC [10], and the right one is of CRC-RLS [18]. From the results, we could see that the representations of nearby points are more similar than that of non-neighboring points, i.e., local consistency could be defined as the similar inputs have similar codes.

However, is ℓ_1-norm-based sparsity really necessary to improve the performance of face recognition? Several recent works directly or indirectly examined this problem. Yang et al. [15] discussed the connections and differences between ℓ_1-optimizer and ℓ_0-optimizer for SRC. They show that the success of SRC should attribute to the mechanism of ℓ_1-optimizer which selects the set of support training samples for the given testing sample by minimizing reconstruction error. Consequently, Yang et al. pointed out that the global similarity derived from ℓ_1-optimizer but sparsity derived from ℓ_0-optimizer is more critical for pattern recognition. Rigamonti et al. [16] compared the discrimination of two different data models. One is the ℓ_1-norm-based sparse representation, and the other model is produced by passing input into a simple convolution filter. Their result showed that two models achieve a similar recognition rate. Therefore, ℓ_1-norm-based sparsity is actually not as essential as it seems in the previous claims. Shi et al. [17] provided a more intuitive approach to investigate this problem by removing the ℓ_1-regularization term from the objective function of SRC. Their experimental results showed that their method achieves a higher recognition rate than SRC if the original data is available. Zhang et al. [18] replaced the ℓ_1-norm by the ℓ_2-norm, and their experimental results again support the views that ℓ_1-norm-based sparsity is not necessary to improve the discrimination of data representation. Moreover, we have noted that Naseem et al. [19,20] proposed Linear Regression Classifier (LRC) which has the same objective function with Shi et al.'s work. The difference is that Shi et al. aimed to explore the role of sparsity while Naseem et al. focused on developing an effective classifier for face recognition.

As another extensively studied concise model, manifold learning (locality preservation model) is usually investigated for dimension reduction by learning and embedding local consistency of original data into a low-dimensional representation [21–23]. Local consistency means that nearby data points share the same properties, which is hardly reflected in linear representation.

Recently, some researchers have explored the possibility of integrating the locality (local consistency) with the sparsity together to produce a better data model. Baraniuk and Wakin [24] successfully bridged the connections between sparse coding and manifold learning, and have founded the theory for random projections of smooth manifold; Majumdar and Ward [25] investigated the effectiveness and robustness of random projection method in classification task. Moreover, Wang et al. [26] proposed a hierarchal images classification method named locality-constrained linear coding (LLC) by introducing dictionary learning into Locally Linear Embedding [27]. Chao et al. [28] presented an approach to unify group sparsity and data locality by introducing the term of ridge regression into LLC; Yang et al. [29] incorporated the prior knowledge into the coding process by iteratively learning a weight matrix of which the entries denotes the similarity between two data points.

In this paper, we proposed and formulated a new kind of local consistency into the linear coding paradigm by enforcing *the similar inputs* (*i.e.*, *neighbors*) *produce similar codes*. The idea is motivated by an observation in biological founds [30] which shows that L2/3 of rat visual cortex activates the same collection of neurons in response to leftward and rightward drifting gratings. Fig. 1 shows an example to illustrate the motivation. There are three face images A, B and C selected from two different individuals, where A and B came from the same person. This means that A and B lie on the same subspace and could represent with each other. Fig. 1(b) is a real example corresponding to Fig. 1(a). Either from the Eigenface [31] matrices or the coefficients of the two coding schemes, we can see that the similarity between A and B is much higher than the similarity between C and either of them.

Based on the observation, we proposed a representation learning method for robust face recognition, named Locality-Constrained Collaborative Representation (LCCR). The algorithm obtains a representation for each data point by enforcing the codes of neighboring points are as similar as possible. Furthermore, the objective function of LCCR has an analytic solution, does not involve local minima. Extensive experiments show that LCCR outperforms SRC [10], LRC [17,19], CRC-RLS [18], CESR [13], LPP [32], and linear SVM with Eigenface [31] in face recognition.

Except in some specified cases, lower-case bold letters represent column vectors and upper-case bold ones represent matrices, $\mathbf{A}^T$ denotes the transpose of the matrix $\mathbf{A}$, $\mathbf{A}^{-1}$ represents the pseudo-inverse of $\mathbf{A}$, and $\mathbf{I}$ is reserved for identity matrix.

The remainder of paper is organized as follows: Section 2 introduces three related approaches for face recognition based on data representation, i.e., SRC [10], LRC [17,19] and CRC-RLS [18]. Section 3 presents our LCCR algorithm. Section 4 reports the experiments on several facial databases. Finally, Section 5 contains the conclusion.

2. Preliminaries

We consider a set of N facial images collected from L subjects. Each training image is denoted as a vector $\mathbf{d}_i \in \mathbb{R}^M$ corresponding to