Journal of Computational Science 27 (2018) 199–205

Contents lists available at ScienceDirect

Journal of Computational Science

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

Feature extraction based on Low-rank affinity matrix for biological recognition

Nan Zhang [a], Yi Chen [b,c,*], Maolong Xi [a], Fangqin Wang [a], Yanwen Qu [d]

[a] Control Technology Institute, Wuxi Institute of Technology, Wuxi 214121, China
[b] Jiangsu Laboratory of Lake Environment Remote Sensing Technologies, Huaiyin Institute of Technology, Huaiyin 223001, China
[c] School of Computer Science and Technology, Nanjing Normal University, Nanjing 210023, China
[d] School of Computer Science and Information Engineering, Jiangxi Normal University, Nanchang 330027, China

ARTICLE INFO

Article history:
Received 2 April 2018
Received in revised form 30 May 2018
Accepted 2 June 2018
Available online 5 June 2018

Keywords:
Low-rank representation
Feature extraction
Discriminative projection
Augmented lagrange multiplier

ABSTRACT

The low-rank representation (LRR) was presented recently and demonstrated its effectiveness for robust subspace segmentation. This paper presents a discriminative projection method based on Low-rank affinity matrix (LRA-DP) for robust feature extraction. The affinity matrix is designed to better preserve the underlying low-rank structure of data representation revealed by LRR. The experiments on the Yale, Extended Yale B, AR face image databases and the PolyU palmprint database showed LRA-DP is always better than or comparable to other state-of-the-art methods, which means underlying low-rank structure of data representation preserved by LRA-DP is helpful for classification problem.

1. Introduction

In recent years, biological recognition has attracted tremendous interests [1–4]. To achieve better results, feature extraction is a very important step in the task of biological recognition. There are a lot of feature extraction methods which were proposed for subspace learning in the last several years [5–12]. Principal component analysis (PCA) [5,6] is a classical algorithm to seek a low-dimensional representation of the high-dimensional data. But it is unable to handle the sparse errors of high magnitude. To solve this problem, Candès [7] and Wright et al. [8] proposed the Robust PCA (RPCA) method. Ran He et al. presented a method for recovering the corrupted low-rank matrices via half-quadratic based nonconvex minimization [9]. Similarly, G. Liu et al. [10] present the low-rank representation (LRR) method based low-rank hypothesis. The two methods both consider that the data X is composed by the original data and the noises. The difference between LRR and Robust PCA is that LRR assumes the representation coefficients of data vectors with respect to a dictionary is low-rank, while the RPCA assumes the data matrix itself is low-rank. LRR has the ability for subspace segmentation, but RPCA does not. Recently, D. Luo et al. proposed a method for multi-subspace representation and discovery [11].

LRR is an effective, robust subspace clustering technique and has found wide applications in machine learning and computer vision, e.g., motion segmentation, face clustering, and temporal segmentation [10,13,14,16,32–34].

LRR, however, expects that the subspaces are independent. In fact, the subspaces are not as independent as expected. Therefore, working without class information, LRR might not obtain desirable clustering result. In addition, LRR performs subspace clustering in the original data vector space and works closely for the training data. For a new test sample, in order to determine its class, we have to retrain by solving the problem (4) after adding the new sample. It is obvious that the retraining process is very time-consuming if there are a large number of test samples.

LRR-based discriminative projection method (LRR-DP) [15] was proposed to address the problem of LRR for robust feature extraction. The more block-tridiagonal the affinity matrix gained by LRR is, the better discriminative information we will obtain. In this paper, we try to enhance the performance of feature extraction by optimizing the affinity matrix. The starting point is the Lagrangian multiplier method which is used to solve the affinity matrix. The only K max singular values are selected, when we take the Inexact ALM algorithm instead of ALM algorithm to calculate the affinity matrix. We perform experiments using the Yale, Extended Yale B,

* Corresponding author at: Jiangsu Laboratory of Lake Environment Remote Sensing Technologies, Huaiyin Institute of Technology, Huaiyin 223001, China.
E-mail address: cystory@qq.com (Y. Chen).

https://doi.org/10.1016/j.jocs.2018.06.001
1877-7503/

AR face image databases and the PolyU palmprint database, and results demonstrate the effectiveness of the proposed method.

2. LRR affinity matrix

2.1. LRR

$X = [x_1, x_2, \ldots, x_n]$ is a set of data vectors in D-dimensional space. The dataset is represented using a dictionary $A = [a_1, a_2, \ldots, a_m]$:

$$X = AW \tag{1}$$

where W is the representation coefficient matrix.

Since the dictionary A is usually overcomplete, LRR seeks an optimal low- rank solution by solving the following problem [10]:

$$\min_W rank(W), s.t. X = AW. \tag{2}$$

But it is hard to solve the above optimization problem (2). Fortunately, Candès et al. [7] indicated that the following convex optimization problem can take the place of the problem (2):

$$\min_W \|W\|_*, s.t. X = AW, \tag{3}$$

where $|\cdot|_*$ denotes the nuclear norm of a matrix [16], i.e. the sum of the singular values of the matrix.

2.2. LRR for noisy data

Suppose the data set X owns c classes $\{S_i\}_{i=1}^{c}$. Let X_i be a set of n_i data points from this subspace, i.e. $X = [X_1, X_2, \ldots, X_c]$.

For subspace clustering, we need to compute an affinity matrix that characterizes the pairwise affinities (similarities) between data vectors. Therefore, the data X itself is used as the dictionary [10], i.e., problem (3) becomes

$$\min_W \|W\|_*, s.t. X = XW. \tag{4}$$

In [10], it is ensured that a good solution liked (5) to problem (4) exists. The solution (5) is considered to be block-diagonal.

$$W^* = \begin{bmatrix} W_1^* & 0 & \cdots & 0 \\ 0 & W_2^* & \cdots & 0 \\ & & & \\ 0 & 0 & \ddots & 0 \\ 0 & 0 & 0 & W_c^* \end{bmatrix}. \tag{5}$$

Liu and Lin [10] give the Inexact ALM algorithm to calculate the optimal solution W^* to problem (4).

2.3. LRR affinity matrix for corrupted data

It is considered that the corrupted data can be composed by original part and noise part, i.e. X = XW + E. The affinity matrix W can be obtained by solving the following optimal problem [10]:

$$\min_{W,E} \|W\|_* + \lambda\|E\|_{2,1}, s.t. X = XW + E, \tag{6}$$

where $\|E\|_{2,1} = \sum_{j=1}^{n}\sqrt{\sum_{i=1}^{n}\left([E]_{ij}\right)^2}$ [17], $[E]_l$ is the *l-th* column of E and $\lambda > 0$.

Refer to [18], a reasonable strategy is simply to relax the equality constraint in (4). The problem (6) is converted to the following equivalent problem:

$$\min_{W,E} \|J\|_* + \lambda\|E\|_{2,1}, s.t. X = XW + E, W = J, \tag{7}$$

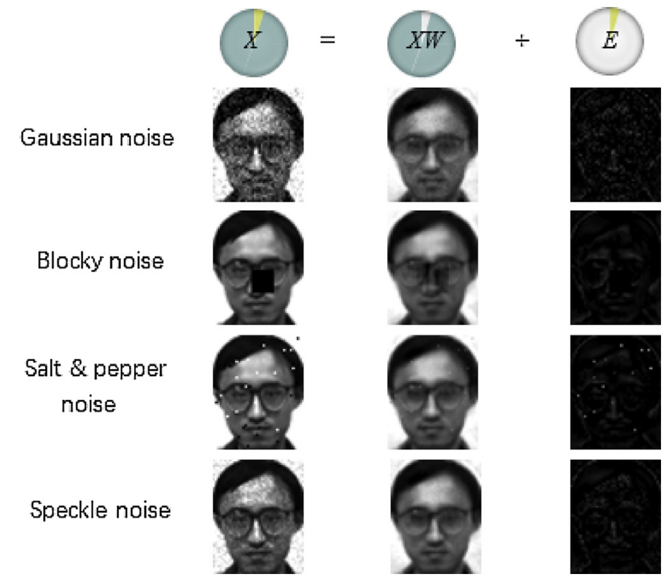

Fig. 1. The example using LRR to correct the noise in face, Left: the original data (X); Middle: the corrected data (XW); Right: the error (E).

which can be obtained by solving the following Augmented Lagrange Multiplier (ALM) problem [19]:

$$\min_{W,E,J,Y_1,Y_2} \|J\|_* + \lambda\|E\|_{2,1} + tr\left[Y_1^t(X - XW - E)\right] + tr\left[Y_2^t(W - J)\right] + \frac{\mu}{2}(\|X - XW - E\|_F^2 + W - J\|_F^2), \tag{8}$$

where Y_1 and Y_2 are Lagrange multipliers and >0 is a penalty parameter.

3. Low-rank affinity matrix based Discriminative Projection (LRA-DP)

The LRR-DP [15] try to find a linear projection by virtue of the low-rank representation coefficient matrix. In this paper, we aim to improve the method by optimizing the affinity matrix W^* of LRR. Because it is considered that the more block-diagonal the affinity matrix is, the better discriminative projection we will obtain. We will just select the only K max singular values for each iteration, when we take the Inexact ALM algorithm instead of ALM algorithm to calculate the affinity matrix of LRR.

The corrupted data can be separated to two parts, i.e. X = XW + E. For four kinds of noise, LRR perform well for denoising on the Yale dataset, as shown in Fig. 1.

In the process of calculating the affinity matrix W, we need solve the singular value decomposition problem:

$$(U, S, V) = svd\left(X - E_k - \mu_k^{-1}Y_k\right), \tag{9}$$

where k is the number of iteration.

Here, we just select the only K max *singular values* S' to take place of S. The iterative update of the affinity matrix is obtained by formula (10):

$$J_{k+1} = US_{\mu_k^{-1}}\lfloor S'\rfloor SV^T, \tag{10}$$

where $S_{\mu_k^{-1}}S' = \begin{cases} S' - \mu_k^{-1}, \text{if } S' > \mu_k^{-1} \\ S' + \mu_k^{-1}, \text{if } S' < -\mu_k^{-1} \\ 0, \textit{otherwise} \end{cases}$

We obtain the affinity matrix W by the improve method above mentioned. Let x_{ij} be the *j-th* sample of class i, w_{ij} be the affinity vector corresponding to the representation weights of x_{ij}, and w_{ij}^s