Low Rank Representation on Grassmann Manifolds

Boyue Wang[1], Yongli Hu[1], Junbin Gao[2,*], Yanfeng Sun[1] and Baocai Yin[1]

[1]Beijing Key Laboratory of Multimedia and Intelligent Software Technology, College of Metropolitan Transportation, Beijing University of Technology, Beijing, China
boyue.wang@gmail.com,{huyongli,yfsun,ybc}@bjut.edu.cn
[2]School of Computing and Mathematics, Charles Sturt University
Bathurst, NSW 2795, Australia
jbgao@csu.edu.au

Abstract. Low-rank representation (LRR) has recently attracted great interest due to its pleasing efficacy in exploring low-dimensional subspace structures embedded in data. One of its successful applications is subspace clustering which means data are clustered according to the subspaces they belong to. In this paper, at a higher level, we intend to cluster subspaces into classes of subspaces. This is naturally described as a clustering problem on Grassmann manifold. The novelty of this paper is to generalize LRR on Euclidean space into the LRR model on Grassmann manifold. The new method has many applications in computer vision tasks. The paper conducts the experiments over two real world examples, clustering handwritten digits and clustering dynamic textures. The experiments show the proposed method outperforms a number of existing methods.

1 Introduction

In recent years, sparse representation and dictionary learning gain much attention in signal processing and machine learning applications. The sparse representation is based on the principle that a signal can normally be represented as a linear combinations of few atoms of a dictionary. And plenty of efforts are dedicated to constructing dictionaries with desired properties [1]. The sparse representation model has achieved great success in various application areas such as face recognition [2], image denoising [3], inpainting [4], image super-resolution reconstruction [5], and so on. In most of sparse representation methods, one mainly focuses on independent sparse representation for data objects, and the relation among data objects or the underlying structure of subspaces that the subsets of group data generated is usually not well considered. While this intrinsic property is very important in some learning tasks, especially in classification and clustering applications.

Some researchers have introduced holistic constraints such as the low rank or nuclear norm $\|\cdot\|_*$ constraints as favoured sparse representation conditions. The good example in the new trend is the Low Rank Representation (LRR)

model [6] which has been successfully used in many applications such as motion segmentation [7], image segmentation [8], and salient object detection [9]. In fact, the low rank criterion as one special type of sparsity measure has long been utilized in matrix completion from corrupted or missing data [10, 11]. Low rank representation tries to reveal the latent sparse property embedded in a data set in high dimensional space. Specifically, it has been proved that when the high-dimensional data set is actually composed of a union of several low dimension subspaces, then the LRR model can reveal this structure through subspace clustering [6].

The current LRR method originates from the subspace clustering [12], based on the hypothesis that the data can be represented by the space spanned by a set of samples. However, this hypothesis may not be always true for many high-dimensional data in practices. It has been proved that many high-dimensional data have their embedded low manifold structures. For example, the human face images are considered as samples from a non-linear submanifold [13]. To deal with this type of data, one has to respect the local geometry existed in the data, i.e., manifold learning, or use a non-linear mapping to "flat" the data, like kernel methods. The classical embedding algorithms such as LLE [14], ISOMAP [15], LLP [16] and LE [17] are the examples of manifold learning from data.

On the other hand, in computer vision, there are many cases where we clearly know what the manifold is, but we want to analyze these manifolds for some practical tasks. For example, a short video clip of dynamic texture can be represented by a subspace, and all such clips together make up the so-called Grassmann manifold. Thus the problem of clustering dynamic textures becomes clustering the points on Grassmann manifold, in other words, to cluster many subspaces into subgroups of subspaces.

Most manifolds can be considered as low dimensional smooth "surfaces" embedded in a higher dimensional Euclidean space. At each point of the manifold, it is locally similar to Euclidean space. In recent years, Grassmann manifold has attracted great interest in research community. Its Riemannian geometry has been recently investigated [18]. Grassmann manifold has a nice property that it can be embedded into the space of symmetric matrices via the projection embedding, referring to section 2.2 below. This property was used in subspace tracking [19], clustering [20], discriminant analysis [21], and sparse coding [22] [23]. Harandi *et.al.* [23] address the problem of kernel sparse coding and dictionary learning within the space of symmetric positive definite matrices. In this paper, we will establish an LRR model on Grassmann manifold based on the similar approach used in the above work and further explore the model performance in clustering subspaces, i.e., grouping a number of subspaces into subgroups of subspaces. The contributions of this work are listed as follows:

- Formulating the LRR model on Grassmann Manifold;
- Exploring the link between the proposed LRR model and kernelization; and
- Providing a practical and effective algorithm for the formulated LRR model.

The rest of the paper is organized as follows. In the next Section, we describe the proposed low-rank representation on Grassmann Manifold in detail. Section

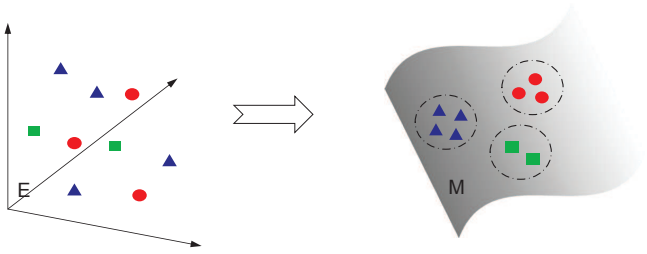


Fig. 1. The points on the Grassmannian manifold and its GLRR representation for clustering.

3 gives the optimization algorithms to resolve the proposed LRR model on on Grassmann Manifold. In Section 4, the performance of the proposed method is evaluated by clustering application on two public databases. Finally, conclusions and suggestions for future work are provided in Section 5.

2 LRR on Grassmann Manifold

2.1 Low-Rank Representation (LRR)

LRR model represents a group of signals on a dictionary with low rank constraint which reveals the intrinsical low rank structure in the signals. The general LRR model can be formulated as the following optimalization problem:

$$\min_{E,Z} \|E\|_l^2 + \lambda \|Z\|_* \quad \text{s.t.} \quad Y = DZ + E \tag{1}$$

where $Y \in \mathbb{R}^{d \times N}$ is a set of N signals with d dimensions and Z is the correspondent low rank representation of Y under the dictionary D, which could be trained or constructed from data samples, and E represents the error between the signals and its reconstructed values on D. $\|\cdot\|_*$ is the nuclear norm which is defined as the sum of singular values of a matrix and is the low envelop of the rank function of a matrix [24]. $\|\cdot\|_l$ is the reconstruction error measurement. We may choose one of different error measurements, depending on the properties of the signals and the aim of applications. For example, in the LRR clustering applications [7] [6], $\|\cdot\|_{2,1}$ is used to cope with columnwise gross errors in signals. Finally $\lambda > 0$ is a penalty parameter to balance the rank term and the reconstruction error.

In the above LRR model, it is critical to use an appropriate dictionary D to represent signals. Generally, a dictionary can be learned from some training data by using one of many dictionary learning methods, such as the K-SVD method [1]. However, a dictionary learning procedure is usually time-consuming and so should be done in an offline manner. So many researchers adopt a simple and direct way to construct a dictionary, i.e. using the original signals themselves

as the dictionary. In practice, if there are sufficient data samples, the dictionary composed of original data also has good performance. For example, as in the subspace clustering method [12] [25], when the number of the data samples in each subspace is sufficient, each data point in a union of subspaces could be reconstructed well by a combination of other points from the same subspace. In fact, the original LRR model adopts the strategy of using data samples themselves as the dictionary. Thus (1) is reformulated as follows:

$$\min_{E,Z} \|E\|_l^2 + \lambda \|Z\|_* \quad \text{s.t.} \quad Y = YZ + E \tag{2}$$

2.2 LRR on Grassmann Manifolds

In most of cases, the reconstruction error of LRR model in (2) is computed in the original data domain. For example, the common form of the reconstruction error is Frobenius norm with Euclidian distance in original data space, i.e. the error term can be chosen as $\|Y - YZ\|_F$. In practice, many high dimension data have its intrinsic manifold structure, for example, the human faces in images are proved to have an underlying manifold structure [26]. In the ideal scenarios, the error should be measured according to the manifold's geometry[1]. So we consider signal representation for the data with manifold structure and formulate the error measurement in LRR model based on the distance defined on manifold spaces. Consequently, the LRR model in (2) can be changed as the following manifold form:

$$\min_{E,Z} \|E\|_{\mathcal{G}}^2 + \lambda \|Z\|_* \quad \text{s.t.} \quad Y = YZ + E \tag{3}$$

where $\|\cdot\|_{\mathcal{G}}$ is the distance (geodesics) on the manifold. Problem (3) is highly nonlinear and it is hard to design a practical algorithm. However when the underlying manifold is Grassmannian, we can use the distance over its embedded space to replace the manifold distance, as detailed below.

Grassmann manifold $\mathcal{G}(p,d)$ [27] is the space of all p-dimensional linear subspaces of $\mathbb{R}^d$ for $0 < p < d$. A point on Grassmann manifold is a subspace of $\mathbb{R}^d$ which can be represented by any of orthonormal basis $X = [\mathbf{x}_1, \mathbf{x}_2, ..., \mathbf{x}_p] \in \mathbb{R}^{d\times p}$. The chosen orthonormal basis is called a representative of a subspace $\mathcal{S}$. Grassmann manifold $\mathcal{G}(p,d)$ has one-to-one correspondence to a quotient manifold of $\mathbb{R}^{d\times p}$, see [27]. On the other hand, we can embed Grassmann manifold $\mathcal{G}(p,d)$ into the space of d order symmetric matrices $\mathrm{Sym}(d)$ by mapping

$$\Pi : \mathcal{G}(p,d) \to \mathrm{Sym}(d), \quad \Pi(X) = XX^T \tag{4}$$

This process can be demonstrated in Fig. 1. The embedding $\Pi(X)$ is diffeomorphism [28] (a one-to-one, continuous, differentiable mapping with a continuous, differentiable inverse). Then it is reasonable to replace the distance on

[1] As the manifold is generally no longer linear, so the linear combination on the manifold should be implemented via exp and log operations on the manifold. We ignore this for the simplicity of presenting our idea.

Grassmann manifold by the following distance defined on the symmetric matrix space under this mapping,

$$\delta(X_1, X_2) = \|\Pi(X_1) - \Pi(X_2)\|_F = \|X_1X_1^T - X_2X_2^T\|_F \tag{5}$$

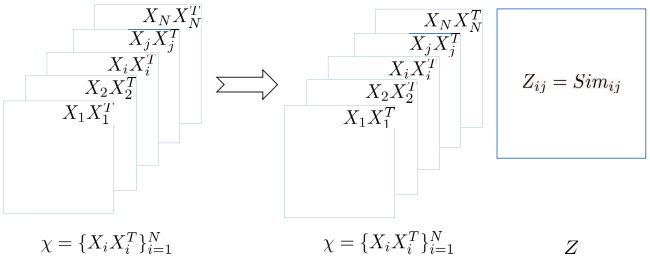


Fig. 2. The GLRR Model. The mapping of the points on Grassmann manifold, the tensor $\mathcal{X}$ with each slice being a symmetric matrix can be linearly represented by itself. The element z_{ij} of Z represents the similarity between slice i and j.

Given a set of data points $\{X_1, X_2, ..., X_N\}$ on Grassmann manifold, i.e., a set of subspaces $\{\mathcal{S}_1, \mathcal{S}_2, ..., \mathcal{S}_N\}$ of dimension p accordingly, we have their mapped symmetric matrices $\{X_1X_1^T, X_2X_2^T, ..., X_NX_N^T\} \subset \mathrm{Sym}(d)$. Similar to the LRR model in (3), we represent these symmetric matrices by itself and use the error measurement defined in (5) to construct the LRR model on Grassmann manifold as follows:

$$\min_{E,Z} \|E\|_F^2 + \lambda\|Z\|_* \ , \ \text{s.t.} \ \ \mathcal{X} = \mathcal{X} \times_3 Z + E \tag{6}$$

where $\mathcal{X}$ is a 3-order tensor by stacking all mapped symmetric matrices $\{X_1X_1^T, X_2X_2^T, ..., X_NX_N^T\}$ along the 3rd mode and $\times_3$ means the mode-3 multiplication of a tensor and a matrix, see [29]. The representation of $\mathcal{X}$ and the 3-order product operation are illustrated in Fig. 2. E is the reconstruction error in the 3-order tensorial form and $\|\cdot\|_F$ is the Frobenius norm of a tensor, which can be defined as the square root of the sum of the absolute squares of all elements in E.

To provide a practical algorithm to the optimization problem in (6), we further investigate the structure of tensor used in the problem. Intuitively, the tensor calculation can be converted to matrix operation by tensorial matricization, see [29]. For example, we can matricize the tensor $\mathcal{X} \in \mathbb{R}^{d\times d\times N}$ in mode-3 and obtain a matrix $\mathcal{X}_{(3)} \in \mathbb{R}^{N\times(d*d)}$ of N data points (in rows). So the problem seems be solved using the method of the standard LRR model. However, as the dimension $d * d$ is often too large in practical problems, the existing LRR algorithm could break down. To avoid this scenario, we carefully analyze the representation of the construction tensor error E and convert the optimization problem to an equivalent and readily solvable optimization model.

Consider the construction error term $\|E\|_F^2 = \|\mathcal{X} - \mathcal{X} \times_3 Z\|_F^2$ in (6). First we write it into a sum of slices of E as follows:

$$\|E\|_F^2 = \sum_{i=1}^{N} \|E_i\|_F^2 \tag{7}$$

where $E_i = X_i X_i^T - \sum_{j=1}^{N} z_{ij}(X_j X_j^T)$ is the i-th slice of E.

Using that $\|M\|_F^2 = \mathrm{tr}(M^T M)$, we further represent the i-th slice $\|E_i\|_F^2$ in the following form:

$$\begin{aligned}\|E_i\|_F^2 &= \mathrm{tr}\left[\left(X_i X_i^T - \sum_{j=1}^{N} z_{ij}(X_j X_j^T)\right)^T \left(X_i X_i^T - \sum_{j=1}^{N} z_{ij}(X_j X_j^T)\right)\right] \\ &= \mathrm{tr}\left[(X_i X_i^T)^T (X_i X_i^T)\right] - 2\sum_{j=1}^{N} z_{ij}\mathrm{tr}\left[(X_i X_i^T)^T (X_j X_j^T)\right] \\ &+ \sum_{j_1=1}^{N}\sum_{j_2=1}^{N} z_{ij_1} z_{ij_2}\mathrm{tr}\left[(X_{j_1} X_{j_1}^T)(X_{j_2} X_{j_2}^T)\right]\end{aligned} \tag{8}$$

It is easy to show that

$$\mathrm{tr}\left[(X_i X_i^T)^T (X_i X_i^T)\right] = \mathrm{tr}\left[(X_i^T X_i)(X_i^T X_i)\right] = \mathrm{tr}[I_p] = p \tag{9}$$

and

$$\mathrm{tr}\left[(X_j X_j^T)^T (X_i X_i^T)\right] = \mathrm{tr}\left[(X_j^T X_i)^T (X_i^T X_j)\right]. \tag{10}$$

Substituting (9) and (10) into (8), we have

$$\begin{aligned}\|E_i\|_F^2 = p &- 2\sum_{j=1}^{N} z_{ij}\mathrm{tr}\left[(X_j^T X_i)(X_i^T X_j)\right] \\ &+ \sum_{j_1=1}^{N}\sum_{j_2=1}^{N} z_{ij_1} z_{ij_2}\mathrm{tr}\left[(X_{j_1}^T X_{j_2})(X_{j_2}^T X_{j_1})\right]\end{aligned} \tag{11}$$

We note that $(X_j^T X_i)$ has a small dimension $p \times p$ which is easy to handle. To simplify expression (11), we denote

$$\Delta_{ij} = \mathrm{tr}\left[(X_j^T X_i)(X_i^T X_j)\right] \tag{12}$$

Clearly $\Delta_{ij} = \Delta_{ji}$. So we construct an $N \times N$ symmetrical matrix

$$\Delta = (\Delta_{ij})_{i=1,j=1}^{N} \tag{13}$$

Substituting (12) and (13) into (11), we have

$$\begin{aligned}\|E_i\|_F^2 &= p - 2\sum_{j=1}^{N} z_{ij}\Delta_{ij} + \sum_{j_1=1}^{N}\sum_{j_2=1}^{N} z_{ij_1}z_{ij_2}\Delta_{j_1 j_2} \\ &= p - 2\sum_{j=1}^{N} z_{ij}\Delta_{ij} + \mathbf{z}_i\Delta\mathbf{z}_i^T\end{aligned} \tag{14}$$

where $\mathbf{z}_i$ is the i-th row of Z. Substituting (14) into (7), we have a simplified reconstruction error representation:

$$\|E\|_F^2 = Np - 2\mathrm{tr}[Z\Delta] + \mathrm{tr}[Z\Delta Z^T] \tag{15}$$

It can be proved that Δ is semi-definite positive (see supplementary materials), so we have $\Delta = LL^T$ by Cholesky decomposition [30], where L is a $N \times N$ matrix. So the above representation can be further written as:

$$\begin{aligned}\|E\|_F^2 &= Np - 2\mathrm{tr}[ZLL^T] + \mathrm{tr}[ZL(ZL)^T] \\ &= \|ZL - L\|_F^2 + \text{const.}\end{aligned} \tag{16}$$

Combining (16) and (6), we have an equivalent problem to the LRR model on Grassmann manifold in (6):

$$\min_Z \|ZL - L\|_F^2 + \lambda\|Z\|_* \tag{17}$$

The final LRR model on Grassmann manifold resembles the original LRR model. We would like to give a remark here. In fact, the LRR model on Grassman manifold (6) can be regarded a kernelized LRR with a kernel feature mapping defined by (4). It is not surprised that Δ semi-definite positive as it serves as a kernel matrix. Finally it is natural that we can further generalize the LRR model on Grassmann manifold based on other kernel functions.

3 Solution to LRR on Grassmann manifold

In this section, we consider an algorithm to solve the optimization problem in (17) with a combination of nuclear and Frobenius norm about Z. We can easily solve it by using linearization technique to deal with the quadratic term $\|ZL - L\|_F^2$. However, to sufficiently take advantage of the existing algorithm for the original LRR, we employ the Augmented Lagrangian Multiplier (ALM) [31]

So we let $J = Z$ to separate the terms of variable Z and the problem in (17) can be formulated as follows:

$$\min_Z \frac{1}{\lambda}\|ZL - L\|_F^2 + \|J\|_* \;\; \text{s.t.} \;\; J = Z \tag{18}$$

Its Augmented Lagrangian Multiplier formulation can be defined as the following unconstrained optimization:

$$\min_{Z,J} \frac{1}{\lambda}\|ZL - L\|_F^2 + \|J\|_* + \langle A, Z - J\rangle + \frac{\mu}{2}\|Z - J\|_F^2 \tag{19}$$

where A is the Lagrangian Multiplier and μ is a weight to tune the error term of $\|Z - J\|_F^2$.

In fact, the above problem can be solved by the following two subproblems in an alternative manner fixing Z or J to optimize the other, respectively.

When fixing Z, the following subproblem is solved to update J:

$$\min_J(||J||_* + \langle A, Z - J\rangle + \frac{\mu}{2}||Z - J||_F^2) \tag{20}$$

When fixing J, the following subproblem should be solved to update Z:

$$\min_Z(\frac{1}{\lambda}||ZL - L||_F^2 + \langle A, Z - J\rangle + \frac{\mu}{2}||Z - J||_F^2) \tag{21}$$

For the subproblem in (20), it can be solved by the following steps. Firstly, the optimization is revised as follows:

$$\min_J(||J||_* + \frac{\mu}{2}||J - (Z + \frac{A}{\mu})||_F^2) \tag{22}$$

(22) has a closed-form solution given by,

$$J^* = \Theta_{\mu^{-1}}(Z + \frac{A}{\mu})$$

where $\Theta(\cdot)$ denotes the singular value thresholding operator (SVT), see [32].

The subproblem in (21) is a quadratic optimization problem about Z. The closed-form solution is given by

$$Z = (\lambda\mu J - \lambda A + 2LL^T)(2LL^T + \lambda\mu I)^{-1} \tag{23}$$

Once solving the former two subproblems about J and Z respectively, we achieve a complete solution to LRR on Grassmann manifold. The whole procedure of LRR on Grassmann manifold is summarized in Algorithm 1.

4 Experiments

4.1 Data Preparation and Experiment Settings

To evaluate the proposed LRR model on Grassmann Manifold, we apply it to image signals representation and then the representation results are used for image clustering by Ncut method [33]. We choose two widely used public databases to test our method. One is the MNIST handwriting digits [34] image set, in which there are more than 70000 digit images written by different persons. The other is the DynTex++ database [35]. This database is a collection of videos from different classes.

To apply our method for image clustering, we use three steps to set up experiments: (a) Mapping the raw signals onto the points on Grassmann Manifold; (b)

Algorithm 1 Low-Rank Representation on Grassmann Manifold.

Input: The Grassmann sample set $\{X_i\}_{i=1}^N, X_i \in \mathcal{G}(p,d)$, the cluster number k and the balancing parameter λ.

Output: The Low-Rank Representation Z

Initialize:$J = Z = 0, A = B = 0, \mu = 10^{-6}, \mu_{max} = 10^{10}$ and $\varepsilon = 10^{-8}$
for i=1:N **do**
  for j=1:N **do**
    $\Delta_{ij} \leftarrow \mathrm{tr}[(X_j^T X_i)(X_i^T X_j)]$;
  end for
end for
Computing L by Cholesky Decomposition $\Delta = LL^T$;
while not converged **do**
  fix Z and update J by
  $J \leftarrow \min_J(||J||_* + < A, Z - J > + \frac{\mu}{2}||Z - J||_F^2)$;
  fix J and update Z by
  $Z \leftarrow (\lambda\mu J - \lambda A + 2LL^T)(2LL^T + \lambda\mu I)^{-1}$;
  update the multipliers:
  $A \leftarrow A + \mu(Z - J)$
  update the parameter μ by $\mu \leftarrow \min(\rho\mu, \mu_{\max})$
  check the convergence condition:
  $\|Z - J\|_\infty < \varepsilon$
end while

Applying the low-rank representation on Grassmann Manifold and (c) Conducting Ncut over the representation results from LRR. Note that our LRR model is designed to cluster points on Grassmann manifold (i.e., clustering subspaces), while most existing clustering algorithms are designed for clustering raw object/signal points. In essence, all the existing algorithms are only considered as benchmarks in assessing our new method.

The raw data in our experiments are image sets with huge volume, for example, the digit image sets for clustering have 3495 image sets and each set is formulated as a high-dimensional vector of size $28 \times 28 \times 20 = 15680$. It is difficult to process this high-dimensional data set on a common machine. So we use PCA (Principal Component Analysis) to reduce the dimension of raw data. Then the data with reduced dimension is represented by LRR model and used for final clustering. For the purpose of clustering comparison, the Ncut method is compared with K-means method for different data representation. Table 1 shows the methods to be compared with our method in our experiments. Under the LRR model, the performance of K-means is bad, so we give up the 4th and the last experiments. All the algorithms are coded by matlab R2011b and implemented on an Xeon-X5675 3.06GHz CPU machine with 12G RAM.

Mapping a sub-group of raw signals onto the points on Grassmann Manifold is a key step in our method. As a point on Grassmann Manifold is represented by an orthonormal basis of a subspace, given the samples from a subspace we should construct its basis representation. According to the work of Harandi [22] [36], we simply adopt Singular Value Decomposition(SVD) to construct the subspace ba-

Data processing method	Data representation method	Clustering method	Combining method
PCA	-	Ncut	PNcut
PCA	-	K-mean	PK-mean
PCA	LRR	Ncut	PLRRNcut
PCA	LRR	K-mean	-
Grassmann Manifold	-	Ncut	GNcut
Grassmann Manifold	-	K-mean	GKmean
Grassmann Manifold	LRR	Ncut	GLRRNcut (our method)
Grassmann Manifold	LRR	K-mean	-

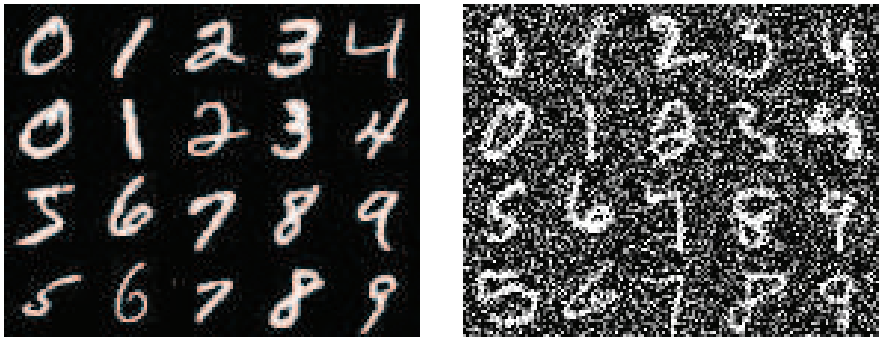
Table 1. Different combining clustering methods with variety in data processing, data representation and clustering methods.

sis. Concretely, given a subset of images from a class, e.g., the same digits written by the same person, denoted by $\{Y_i\}_{i=1}^{M}$ and each Y_i is a grey-scale image with dimension $m \times n$, we can construct a matrix $\Gamma = [\text{vec}(Y_1), \text{vec}(Y_2), ..., \text{vec}(Y_m)]$ of size $(m \times n) \times M$ by vectorizing each image Y_i. Then Γ is decomposed by SVD as $\Gamma = U\Sigma V$. We can pick the first p singular-vectors of U as the representative of a point on Grassmann manifold $\mathcal{G}(p, m \times n)$ with size $(m \times n) \times p$. In the following experiments, the points on Grassmann Manifold of MNIST handwriting digits images are constructed by the above method.

4.2 MNIST Handwritten digits Clustering

The MNIST database [34] has been widely used in pattern recognition field. The digit images in this database are written by about 250 volunteers and for each digit there are different number of samples. In recognition application, 60,000 digit images in the database are often used as training data and 10,000 images are used as testing data. All the digit images in this databse have been size-normalized and centered in a fixed size of 28×28, so it does not need much efforts for preprocessing and formatting. Fig. 3(a) shows some digit samples of this database. As the proposed LRR model on Grassmann Manifold dose not need training, we merge the training set and testing set into a single sample set to do clustering experiment.

In our experiment, we created subgroups randomly according to their classes so that the each subgroup consists of 20 images, i.e. $M = 20$. The subspace generated by each subgroup will be considered as a point on Grassmann manifold. Our task is to cluster all the available $N = 3495$ image subgroups into ten categories. Please note that we are clustering subgroups, not single digits. As mentioned above, for each subgroup of 20 images, we form them as vectors and then stack them into a matrix Γ. Then SVD decomposition is applied on Γ and the front $p = 10$ singular-vectors of its left singular matrix are used as the subspace basis, the point on Grassmann Manifold [22] [36]. Finally the algorithm of the proposed LRR model on Grassmann Manifold is conducted on these points



(a) The origin digit images (b) The digit images with noise

Fig. 3. The MNIST digit samples for experiments

on Grassmann manifold and the result of Z is pipelined to the Ncut algorithm. As a benchmark comparison, we mimic a point in Euclidean for each subgroup of 20 images by stacking them into a vector of dimension $28 \times 28 \times 20$. For PCA based method, the size $28 \times 28 \times 20 = 15680$ of each vector is reduced to 1566 by PCA.

The experimental results are reported in Table 2. It is shown that our proposed algorithm has the highest accuracy of 99%, outperforming other methods more than 10 percents. The two worse results are for GNcut and GK-means methods. This can be explained that, after the Grassmann Manifold mapping, it is not a proper way to measure the Enclidean distance between the points on the manifold space. As our LRR model use the Grassmann distance measurement in a transformed space, the clustering results of our method are better than that of PNcut or PLRRNcut, in which the Euclidean distance is used to measure the relation of data reduced by PCA. The manifold mapping extracts more useful information about the differences among sample data. Then the combination of Grassmann distance and LRR model brings good accuracy for Ncut clustering.

	PK-means	PNcut	PLRRNcut	GK-means	GNcut	our method
Accuracy	0.8638	0.9013	0.8552	0.4103	0.3339	0.99

Table 2. Subspace clustering results on the MINST database.

We further tested the robustness of the proposed algorithm by adding noise to the digit images. We added a Gaussian noise $N(0, \sigma^2)$ onto all the digit images. Fig 3(b) shows the digit images with noise $\sigma = 0.45$. Generally, the noises will effect the performance of the clustering algorithm, especially when the noise is heavy. Fig 4 shows the clustering performance of different methods with the noise standard deviation σ ranging from 0.05 to 0.55. It indicates that our algorithm keeps over 99% accuracy for the standard deviation up to 0.45, while the accuracy

of other methods is generally lower than our method and behaves unstable when the noise standard deviation varies. This indicates that our proposed algorithm is robust for certain level of noises.

λ is an important parameter for balancing the error term and the low-rank term of the LRR model on Grassmann Manifold in (6). This is the same case for the PLRRncut method. We studied the effectiveness of λ for the final clustering result in several experiments. It can be observed that the noise level will change the rank of low-rank representation Z. A larger noise level will increase the rank of the represented coefficient matrix. So a proper way to tune λ is dependent on noise level. Generally. we can make λ small when the noise of data is lower and use a larger λ value if the noise level is higher. In our experiments, we set $\lambda = 0.1$ for the our LRR model and set $\lambda = 0.5$ for PLRR method.

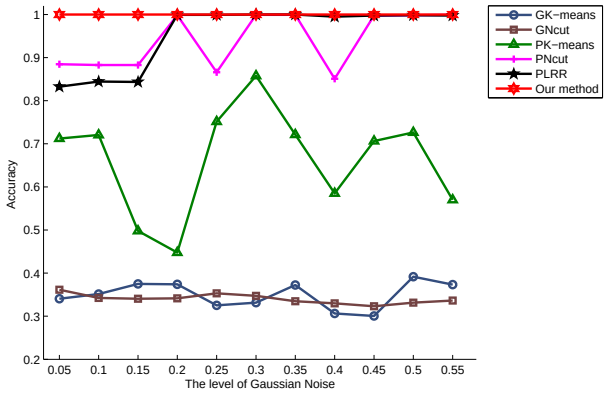


Fig. 4. Clustering on MNIST database with different level of noise.

4.3 Dynamic Texture Clustering

DynTex++ database [35] is derived from a total of 345 video sequences in different scenarios which contains river water, fish swimming, smoke, cloud and so on. These videos are labeled as 36 classes and each class has 100 subsequences (totally 3600 subsequences) with a fixed size of $50 \times 50 \times 50$ (50 gray frames). Some samples of DynTex++ are shown in Fig 5. This is a challenge database for clustering because most of texture from different class is fairly similar.

Our method uses a set of samples to construct points on Grassmann Manifold, it is suitable to process video sequence data as a clip of continuous frames can be naturally collected as an image set. Moreover, as video sequence contain useful space and time context information for clustering, in order to utilize

Fig. 5. DynTex++ samples. Each row includes frames from same video sequence and each four frames is a continuous clip.

these infomation, instead of using SVD method, we use Local Binary Patterns from Three Orthogonal Plans (LBP-TOP) model [37] to construct points on Grassmann Manifold. The LBP-TOP method extract LBP features from three orthogonal planes and concatenate a number of neighbour points' features to form a co-occurrence feature. After extracting LBP-TOP features for the 3600 subsequences, we get 3600 matrices of size 177×14 as the points on Grassmann Manifold. Then these features are represented by LRR model on Grassmann Manifold and its representation Z is used for Ncut clustering. As the data volume of all the 3600 subsequence is huge, we randomly pick K classes from 36 chesses and use these classes data for clustering experiments. The experiments is repeated several times for each K. In the experiments, λ is also set to 0.1. For the PCA based method, the prototype $50 \times 50 \times 50$ subsequence is reduced to 2478 and λ for the PLRRNcut is set to 0.5.

The clustering results for the DynTex++ database are shown in Fig. 6. For different number of classes, the accuracy of the proposed LRR model on Grassmann Manifold are superior to the other methods more than 10 percents, due to information extraction capability in Grassmann Manifold mapping over the LBP-TOP features. We also observed that all of accuracies decreases as the number of classes increases. This may be caused by the clustering challenge when more similar texture images are added into the data set.

5 Conclusion and Future Work

In this paper, we propose a novel LRR model on Grassmann manifold, in which we exploit the property of Grassmann manifold. To resolve the high-dimension issue in the resulting LRR model, we further explore the structured embedding mapping and derive an equivalent optimization problem which is easily solvable. The proposed model and algorithm have been assessed against a number of existing clustering algorithms. The experimental results show the efficiency and robustness of the proposed model. In most of the experimental cases, the proposed method outperforms other benchmark methods. As a future work, we will consider incorporating ℓ_1/ℓ_2 errors in the LRR model on Grassmann manifold and further explore kernelized LRR models on Grassmann manifold.

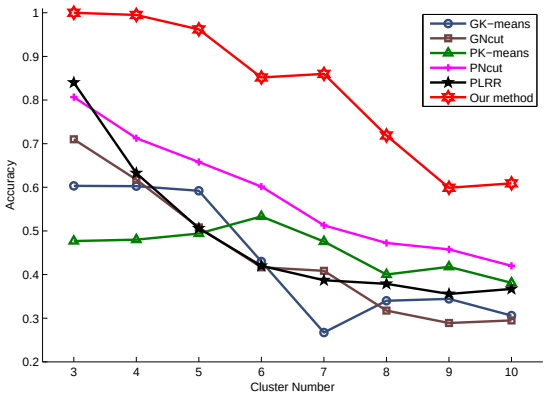


Fig. 6. Clustering results on DynTex++ database with the number of classes ranging from 3 to 10.

6 Acknowledgements

The research project is supported by the Australian Research Council (ARC) through the grant DP130100364 and also partially supported by National Natural Science Foundation of China under Grant No.61390510, 61133003, 61370119, 61171169, 61227004 and Beijing Natural Science Foundation No.4132013.

References

1. Aharon, M., Elad, M., Bruckstein, A.: K-SVD: An algorithm for designing overcomplete dictionaries for sparse representation. IEEE Transactions on Signal Processing **54** (2006) 4311–4322
2. Wright, J., Yang, A., Ganesh, A., Sastry, S., Ma, Y.: Robust face recognition via sparse representation. IEEE Transactions on Pattern Analysis and Machine Intelligence **31** (2009) 210–227
3. Elad, M., Aharon, M.: Image denoising via sparse and redundant representations over learned dictionaries. IEEE Transactions on Image Processing **15** (2006) 3736–3745
4. Mairal, J., Elad, M., Sapiro, G.: Sparse representation for color image restoration. IEEE Transactions on Image Processing **17** (2008) 53–69
5. Yang, J., Wang, Z., Lin, Z., Huang, T.: Coupled dictionary training for image super-resolution. IEEE Transactions on Image Processing **21** (2012) 3467–3478
6. Liu, G., Lin, Z., Yu, Y.: Robust subspace segmentation by low-rank representation. In: International Conference on Machine Learning. (2010) 663–670
7. Liu, G., Lin, Z., Sun, J., Yu, Y., Ma, Y.: Robust recovery of subspace structures by low-rank representation. IEEE Transactions on Pattern Analysis and Machine Intelligence **35** (2013) 171–184

8. Cheng, B., Liu, G., Wang, J., Huang, Z., Yan, S.: Multi-task low-rank affinity pursuit for image segmentation. In: International Conference on Computer Vision. (2011) 2439–2446
9. Lang, C., Liu, G., Yu, J., Yan, S.: Saliency detection by multitask sparsity pursuit. IEEE Transactions on Image Processing **21** (2012) 1327–1338
10. Wright, J., Ganesh, A., Rao, S., Peng, Y., Ma, Y.: Multi-task low-rank affinity pursuit for image segmentation. In: Advances in Neural Information Processing Systems. Volume 22. (2009)
11. Candés, E.J., Li, X., Ma, Y., Wright, J.: Robust principal component analysis? Journal of the ACM **58** (2011) article 11
12. Elhamifar, E., Vidal, R.: Subspace clustering. IEEE Signal Processing Magazine **28** (2011) 52–68
13. Wang, R., Shan, S., Chen, X., Gao, W.: Manifold-manifold distance with application to face recognition based on image set. In: Computer Vision and Pattern Recognition. (2008) 1–8
14. Roweis, S., Saul, L.: Nonlinear dimensionality reduction by locally linear embedding. Science **290** (2000) 2323–2326
15. Tenenbaum, J., Silva, V., Langford, J.: A global geometric framework for nonlinear dimensionality reduction. Optimization Methods and Software **290** (2000) 2319–2323
16. He, X., Niyogi, P.: Locality preserving projections. In: Advances in Neural Information Processing Systems. Volume 16. (2003)
17. Belkin, M., Niyogi, P.: Laplacian eigenmaps and spectral techniques for embedding and clustering. In: Advances in Neural Information Processing Systems. Volume 14. (2001)
18. Absil, P.A., Mahony, R., Sepulchre, R.: Riemannian geometry of Grassmann manifolds with a view on algorithmic computation. Acta Applicadae Mathematicae **80** (2004) 199–220
19. Srivastava, A., Klassen, E.: Bayesian and geometric subspace tracking. Advances in Applied Probability **36** (2004) 43–56
20. Cetingul, H., Vidal, R.: Intrinsic mean shift for clustering on Stiefel and Grassmann manifolds. In: IEEE Conference on Computer Vision and Pattern Recognition. (2009) 1896–1902
21. Hamm, J., Lee, D.: Grassmann discriminant analysis: a unifying view on subspace-based learning. In: International Conference on Machine Learning. (2008) 376–383
22. Harandi, M., Shirazi, S., Sanderson, C., Lovell, B.: Graph embedding discriminant analysis on Grassmannian manifolds for improved image set matching. In: Computer Vision and Pattern Recognition. (2011) 2705–2712
23. Harandi, M., Sanderson, C., Shen, C., Lovell, B.: Dictionary learning and sparse coding on grassmann manifolds: An extrinsic solution. In: International Conference on Computer Vision. Volume 14. (2013) 3120–3127
24. Fazel, M.: Matrix Rank Minimization with Applications. Phd thesis, Stanford University (2002)
25. Elhamifar, E., Vidal, R.: Sparse subspace clustering: Algorithm, Theory, and Applications. IEEE Transactions on Pattern Analysis and Machine Intelligence **35** (2013) 2765–2781
26. Xu, C., Wang, T., Gao, J., Cao, S., Tao, W., Liu, F.: An ordered-patch-based image classification approach on the image Grassmannian manifold. IEEE Transactions on Neural Networks and Learning Systems **25** (2014) 728–737

27. Absil, P., Mahony, R., Sepulchre, R.: Optimization Algorithms on Matrix Manifolds. Princeton University Press (2008)
28. Helmke, J.T., Hüper, K.: Newtons's method on Grassmann manifolds. Technical report, Preprint: [arXiv:0709.2205] (2007)
29. Kolda, G., Bader, B.: Tensor decomposition and applications. SIAM Review **51(3)** (2009) 455–500
30. Gentle, J.E.: Numerical Linear Algebra for Applications in Statistics. Springer-Verlag Press (1998)
31. Shen, Y., Wen, Z., Zhang, Y.: Augmented Lagrangian alternating direction method for matrix separation based on low-rank factorization. Optimization Methods and Software **29** (2014) 239–263
32. Cai, J.F., Candès, E.J., Shen, Z.: A singular value thresholding algorithm for matrix completion. SIAM Journal on Optimization **20** (2008) 1956–1982 http://www-stat.stanford.edu/ candes/papers/SVT.pdf.
33. Shi, J., Malik, J.: Normalized cuts and image segmentation. IEEE Transactions on Pattern Analysis and Machine Intelligence **22** (2000) 888–905
34. Lecun, Y., Bottou, L., Bengio, Y., Haffner, P.: Gradient-based learning applied to document recognition. Proceedings of the IEEE **86** (1998) 2278–2324
35. Ghanem, B., Ahuja, N.: Maximum margin distance learning for dynamic texture recognition. In: European Conference on Computer Vision. (2010) 223–236
36. Harandi, M., Sanderson, C., Hartley, R., Lovell, B.: Sparse coding and dictionary learning for symmetric positive definite matrices: A kernel approach. In: European Conference on Computer Vision. (2012) 216–229
37. Zhao, G., Pietikäinen, M.: Dynamic texture recognition using local binary patterns with an application to facial expressions. IEEE Transactions on Pattern Analysis and Machine Intelligence **29** (2007) 915–928