Reweighted Non-convex Non-smooth Rank Minimization based Spectral Clustering on Grassmann Manifold

Xinglin Piao[1,2,4], Yongli Hu[3,⋆], Junbin Gao[5], Yanfeng Sun[3], Xin Yang[4], and Baocai Yin[3,1]

[1] Peng Cheng Laboratory, Shenzhen, 518055, China;
[2] Peking University Shenzhen Graduate School, Shenzhen, 518055, China;
[3] Beijing Key Laboratory of Multimedia and Intelligent Software Technology, Beijing Artificial Intelligence Institute, Faculty of Information Technology, Beijing University of Technology, Beijing, 100124, China;
[4] Faculty of Electronic Information and Electrical Engineering, Dalian University of Technology, Dalian. 116024, China;
[5] Business Analytics Discipline, The University of Sydney Business School, Camperdown, NSW 2006, Australia.
piaoxl@pcl.ac.cn; huyongli@bjut.edu.cn; junbin.gao@sydney.edu.au; yfsun@bjut.edu.cn; xinyang@dlut.edu.cn; ybc@bjut.edu.cn

Abstract. Low Rank Representation (LRR) based unsupervised clustering methods have achieved great success since these methods could explore low-dimensional subspace structure embedded in original data effectively. The conventional LRR methods generally treat the data as the points in Euclidean space. However, it is no longer suitable for high-dimension data (such as video or imageset). That is because high-dimension data are always considered as non-linear manifold data such as Grassmann manifold. Besides, the typical LRR methods always adopt the traditional single nuclear norm based low rank constraint which can not fully reveal the low rank property of the data representation and often leads to suboptimal solution. In this paper, a new LRR based clustering model is constructed on Grassmann manifold for high-dimension data. In the proposed method, each high-dimension data is formed as a sample on Grassmann manifold with non-linear metric. Meanwhile, a non-convex low rank representation is adopt to reveal the intrinsic property of these high-dimension data and reweighted rank minimization constraint is introduced. The experimental results on several public datasets show that the proposed method outperforms the state-of-the-art clustering methods.

1 Introduction

Unsupervised clustering is a fundamental topic in machine learning, artificial intelligence and data mining areas [1, 2], which attempts to group data into different clusters according to their own intrinsic pattern. In past decades, a large

⋆ Corresponding author

number of clustering methods have been proposed and achieved great success in many applications[3, 4]. The representative ones are the statistical methods [5], the conventional iterative methods [6], the factorization-based algebraic methods [7], and the spectral clustering methods [8]. Among them, the spectral clustering methods are considered having promising performance. In this kind of methods, an affinity matrix is usually learned, and then Normalized Cuts (N-Cut)[9] or other standard clustering algorithms are then used to obtain the final clustering results. Inspired by Sparse and Low Rank representation method, a series of classical methods have been proposed. Elhamifar and Vidal adopted ℓ_1 norm to explore the sparse relationship within data and proposed Sparse Subspace Clustering (SSC) [10] for data clustering. Liu *et al.* used nuclear norm to construct a low-rank representation matrix for data and proposed Low-Rank Representation (LRR) clustering method [11]. Later, Liu *et al.* proposed Latent Low-Rank Representation (LatLRR) method [12] for clustering. Zhang *et al.* proposed Robust Latent Low Rank Representation (RobustLatLRR) clustering method [13]. To get better representation matrix, some researchers adopt the kernel trick and proposed some kernel based clustering methods, such as the kernel SSC clustering method [14] and the kernel LRR clustering method [15].

In the aforementioned methods, vector feature and Euclidean distance are combined for building the affinity matrix by self-expression approach for sample data. However, this conventional linear approach would be no longer suitable for complex or high-dimension data, such as imagesets or video clips data. That is because high-dimension data are always treated as a sample point on non-linear manifold space with non-linear metric. For example, an imageset or a video clip is can be modeled as a data sample on Grassmann manifold [16, 17]. Therefore, to address the high-dimension data clustering task, researchers try to extend the traditional methods and proposed a series of effective clustering approaches for these complex data. Turaga *et al.* proposed a statistical computations based manifold representation method (SCGSM) [18]. Shirazi *et al.* proposed a kernel embedding clustering method on Grassmann manifold (K-GM) [19]. Inspired by low rank and sparse theory, Wang *et al.* first proposed low rank based clustering method on Grassmann manifold (G-LRR) [20]. Liu *et al.* adopted kernel method and proposed kernel sparse representation based clustering method [21]. Wang *et al.* proposed Cascaded Low Rank and Sparse Representation on Grassmann manifold method (G-CLRSR) [22]. Later, Wang *et al.* proposed Partial Sum Minimization of Singular Values Representation on Grassmann manifold method (G-PSSVR) [23], in which partial sum minimization of singular values norm was adopted for better low rank representaion. Piao *et al.* proposed Double Nuclear norm based Low Rank Representation clustering method on Grassmann manifold (G-DNLR) [24]. Further, combined with the Laplacian regularizer, researchers proposed Laplacian Low-Rank Representation on Grassmann manifold method (G-LLRR) [25], Laplacian Partial Sum Minimization of Singular Values Representation on Grassmann manifold method (G-LPSSVR) [23] and Laplacian Double Nuclear norm based Low Rank Representation clustering method

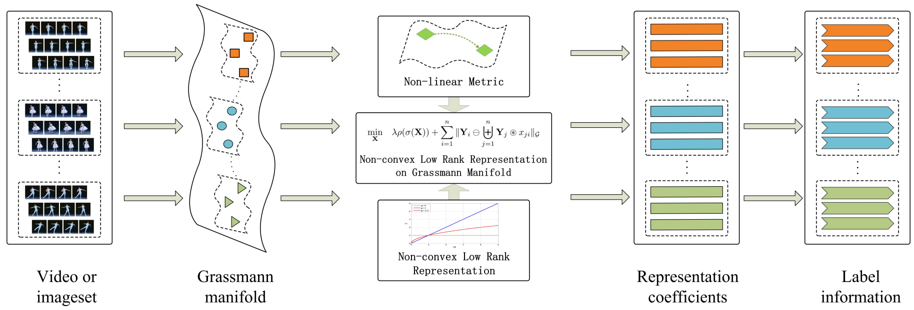


Fig. 1. The pipeline of the paper.

on Grassmann manifold [24], in which authors constructed the affinity matrix by original Grassmann manifold data samples.

Although these current low rank representation and their extension based methods on Grassmann manifold show good performance in clustering task, they generally adopt traditional nuclear norm based low rank constraint for the data representation. This traditional norm treats all singular values equally and prefers to punish the larger singular values than the small ones, which would deviate the optimal solution and lead to suboptimal solution [26, 27]. Recently, to overcome the limitation of convex and smooth nuclear norm, non-convex or non-smooth low rank approximation (such as logarithmic function and Schatten-p norm for $0 < p < 1$) [28] are adopted to replace the traditional nuclear norm for low rank based problems, which could recover a more accurate low rank matrix than the traditional nuclear norm [29]. Especially, these non-convex and non-smooth low rank based methods could increase the punishment on smaller values and decrease the punishment on larger values simultaneously [30]. Inspired from these methods, we propose a novel low rank based clustering model on Grassmann manifold. In the proposed model, the high-dimension data samples are firstly represented as Grassmann points, then a Reweighted Non-convex and Non-smooth Rank Minimization based model on Grassmann manifold (G-RNNRM) is built, where the data representation matrix is constrained by non-convex and non-smooth low rank constraint. The NCut method [9] is used to obtain the final clustering results. Figure 1 shows the pipeline of our paper and the contributions of this paper are following:

- Proposing a novel non-convex and non-smooth low rank representation model on Grassmann for high-dimension data clustering;
- Reweighted approach is introduced in the proposed non-convex and non-smooth low rank approximation norm to reveal low-rank property more exactly. To our best knowledge, this the first reweighted low rank based method on Grassmann manifold;
- An effective algorithm is proposed to solve the complicated optimization problem of the proposed model.

The paper is organized as follows. We introduce the notation and definition of Grassmann manifold in Section 2. Section 3 reviews the related works. We will introduce the formulation and optimization of the proposed G-RNNRM model in Section 4. Section 5 assesses the proposed method on several datasets. Finally, conclusions are discussed in Section 6.

2 Notation and Definition of Grasssmann Manifold

2.1 Notation

We use bold lowercase letters for vectors, e.g. $\mathbf{x}, \mathbf{y}, \mathbf{a}$, bold uppercase for matrices, e.g. $\mathbf{X}, \mathbf{Y}, \mathbf{A}$, calligraphy letters for tensors e.g. $\mathcal{X}, \mathcal{Y}, \mathcal{A}$, lowercase letters for scalars such as dimension and class numbers, e.g. m, n, c. $\mathbf{x}_i$ represents the i-th column of matrix $\mathbf{X}$. x_{ij} represents the i-th element in j-th column from matrix $\mathbf{X}$. $\mathbb{R}$ represents the space of real numbers.

2.2 Definition of Grassmann Manifold

According to [31], a Grassmann manifold is always denoted as $\mathcal{G}(p, m)$, which consists of all linear p-dimension subspaces in m-dimension Euclidean space $\mathbb{R}^m (0 \leq p \leq m)$. It also could be represented by the quotient space of all the $m \times p$ matrices with p orthogonal columns under the p-order orthogonal group. Thus, we could construct a Grassmann manifold as below:

$$\mathcal{G}(p, m) = \{\mathbf{Y} \in \mathbb{R}^{m \times p} : \mathbf{Y}^T \mathbf{Y} = \mathbf{I}_p\} / \mathcal{O}(p), \tag{1}$$

where $\mathcal{O}(p)$ represents the p-order orthogonal group. For two Grassmann manifold data samples $\mathbf{Y}_1$ and $\mathbf{Y}_2$, there are two metric approaches: one is to define a consistent metrics in tangent spaces for Grassmann manifold data [32], the other one is to embed Grassmann manifold data into the symmetric matrix space [33]. The later one is easier and the Euclidean distance could be applied, which could be defined as below :

$$\mathrm{dist}_g(\mathbf{Y}_1, \mathbf{Y}_2) = \frac{1}{2} \|\Pi(\mathbf{Y}_1) - \Pi(\mathbf{Y}_2)\|_F, \tag{2}$$

where $\|\mathbf{X}\|_F = \sqrt{\sum_{i=1,j=1}^{n} x_{ij}^2}$ represents the Frobenius norm, $\Pi(\cdot)$ is a mapping function defined as below:

$$\Pi : \mathcal{G}(p, m) \longrightarrow Sym(m), \Pi(\mathbf{Y}) = \mathbf{Y}\mathbf{Y}^T, \tag{3}$$

where $Sym(m)$ represents the m-dimension symmetric matrix space. With the function $\Pi(\cdot)$, Grassmann manifold could be embedded into the symmetric matrices. Each sample data on Grassmann manifold could be regarded as an equivalent class of all the $m \times p$ orthogonal matrices, any one of which can be converted to the other by a $p \times p$ orthogonal matrix. Thus, Grassmann manifolds is naturally regarded as a good representation for video clips/image sets, thus can be used to tackle the problem of videos matching.

3 Related Works

We first introduce related clustering methods on the Euclidean Space. Given a set of sample vectors $\mathbf{Y} = [\mathbf{y}_1, \mathbf{y}_2, ..., \mathbf{y}_n] \in \mathbb{R}^{m\times n}$ drawn from a union of c subspaces $\{\mathcal{S}_i\}_{i=1}^c$, where m denotes the dimension of each sample $\mathbf{y}_i$ and n represents the number of samples $\mathbf{Y}$. Let $\mathbf{Y}_i \subset \mathbf{Y}$ come from the subspace $\mathcal{S}_i$. The task of subspace clustering is to segment the sample set $\mathbf{Y}$ according to the underlying subspaces. Researchers have proposed a large number of methods to solve this problems. As we mentioned in Section 1, the spectral clustering methods are considered as the state-of-the-art ones, in which the data could be self-represented by introducing a representation matrix $\mathbf{X} \in \mathbb{R}^{n\times n}$ with linear combination as $\mathbf{Y} = \mathbf{Y}\mathbf{X}$. To avoid the trivial solution, some matrix constraints are adopted on $\mathbf{X}$ such as Frobenius norm [34]:

$$\min_{\mathbf{X}} \quad \lambda\|\mathbf{X}\|_F^2 + \|\mathbf{Y} - \mathbf{Y}\mathbf{X}\|_F^2. \tag{4}$$

In the past decade, sparse and low rank theories have been applied to subspace clustering successfully. Elhamifar and Vidal [10] proposed Sparse Subspace Clustering (SSC) method, which aimed to find the sparsest representation matrix $\mathbf{X}$ by using ℓ_1 norm $\|\cdot\|_1$. The SSC model is formulated as follows,

$$\min_{\mathbf{X}} \quad \lambda\|\mathbf{X}\|_1 + \|\mathbf{Y} - \mathbf{Y}\mathbf{X}\|_F^2, \tag{5}$$

where λ is balance parameter and $\|\mathbf{X}\|_1 = \sum_{i=1,j=1}^n |x_{ij}|$. Instead of adopting the sparse constraint, Liu *et al.* [11] proposed Low Rank Representation (LRR) method for clustering by using nuclear norm $\|\cdot\|_*$ for the representation matrix $\mathbf{X}$, which is formulated as follows,

$$\min_{\mathbf{X}} \quad \lambda\|\mathbf{X}\|_* + \|\mathbf{Y} - \mathbf{Y}\mathbf{X}\|_F^2, \tag{6}$$

where $\|\mathbf{X}\|_* = \sum_i^r \sigma_i(\mathbf{X})$ and $\sigma_i(\mathbf{X})$ represents the i-th singular value of $\mathbf{X}$, $\sigma_1 > \sigma_2 > ... > \sigma_r$, r represents the rank of $\mathbf{X}$.

All above related works all construct the representation matrix of data samples by employing Euclidean distance. However, the high-dimension datum are always assumed as Grassmann manifold samples and the Euclidean distance is no longer suitable. Therefore, researchers proposed a series of clustering methods for Grassmann manifold based on the non-distance defined in (2). For a set of Grassmann samples $\mathcal{Y} = \{\mathbf{Y}_1, \mathbf{Y}_2, ..., \mathbf{Y}_n\}$ where $\mathbf{Y}_i \in \mathcal{G}(p, m)$, Wang *et al.* [20] proposed a Low Rank model with non-linear metric for Grassmann (G-LRR) by generating the (6) on Grassmann,:

$$\min_{\mathbf{X}} \quad \lambda\|\mathbf{X}\|_* + \sum_{i=1}^n \|\mathbf{Y}_i \ominus \biguplus_{j=1}^n \mathbf{Y}_j \circledast x_{ji}\|_{\mathcal{G}}, \tag{7}$$

where $\|\mathbf{Y}_i \ominus \biguplus_{j=1}^n \mathbf{Y}_j \circledast x_{ji}\|_{\mathcal{G}}$ represents the reconstruction error of the sample $\mathbf{Y}_i$ on Grassmann manifold, $\biguplus_{j=1}^n \mathbf{Y}_j \circledast x_{ji}$ denotes the "combination" of $\{\mathbf{Y}_j\}_{j=1}^n$

with the coefficients $\{x_{ji}\}_{i=1,j=1}^{n}$, the symbol $\ominus, \biguplus, \circledast$ are abstract symbols which are used to simulated the "linear" operations on Grassmann manifold. They also proposed a cascaded Low Rank and Sparse model on Grassmann manifold (G-CLRSR) model:

$$\begin{aligned} \min_{\mathbf{X},\mathbf{Z}} \quad & \lambda\|\mathbf{X}\|_* + \alpha\|\mathbf{Z}\|_1 + \beta\|\mathbf{X} - \mathbf{X}\mathbf{Z}\|_F^2 \\ & + \sum_{i=1}^{n} \|\mathbf{Y}_i \ominus \biguplus_{j=1}^{n} \mathbf{Y}_j \circledast x_{ji}\|_{\mathcal{G}}. \end{aligned} \tag{8}$$

Further, to achieve a better low rank representation matrix for clustering, Wang *et al.* [23] adopted Partial Sum Minimization of Singular Values (PSSV) norm to instead the nuclear norm for formulating PSSV Low Rank model on Grassmann manifold (G-PSSVLR) model:

$$\min_{\mathbf{X}} \quad \lambda\|\mathbf{X}\|_{>r} + \sum_{i=1}^{n} \|\mathbf{Y}_i \ominus \biguplus_{j=1}^{n} \mathbf{Y}_j \circledast x_{ji}\|_{\mathcal{G}}, \tag{9}$$

where r represents the expected rank of $\mathbf{X}$ and $\|\cdot\|_{>r}$ represents the PSSV norm defined as below [35]:

$$\|\mathbf{X}\|_{>r} = \sum_{i=r+1}^{n} \sigma_i(\mathbf{X}), \tag{10}$$

Although the above methods achieve great performance in Grassmann manifold clustering problem, the traditional convex nuclear based norm is adopted, which would reduce the ability to represent the correlation among data.

4 Reweighted Non-convex and Non-smooth Rank Minimization model on Grassmann manifold

In this section, we will introduce the formulation and optimization of the proposed G-RNNRM model in detail.

4.1 Model Formulation

For a set of Grassmann samples $\mathcal{Y} = \{\mathbf{Y}_1, \mathbf{Y}_2, ..., \mathbf{Y}_n\}$ where $\mathbf{Y}_i \in \mathcal{G}(p, m), i = 1, 2, ..., n$, we could formulate a self-expression clustering model as below:

$$\min_{\mathbf{X}} \quad \lambda f(\mathbf{X}) + g(\mathbf{X}, \mathbf{Y}), \tag{11}$$

where $f(\mathbf{X})$ represents a function satisfying some specific conditions for matrix $\mathbf{X}$, $g(\mathbf{X}, \mathbf{Y})$ represents a function of self-representation of samples $\{\mathbf{Y}_i\}_{i=1}^{n}$ with representation matrix $\mathbf{X}$. According to the preliminary knowledge in Section 2,

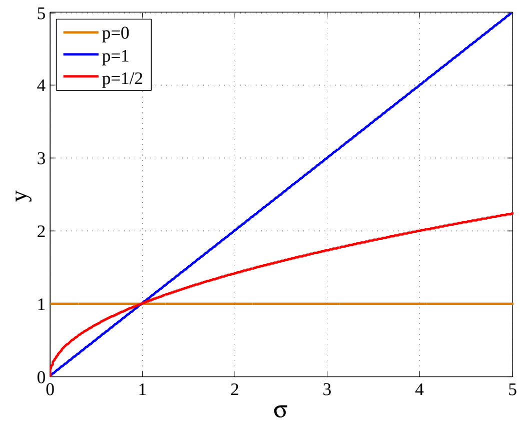


Fig. 2. The illustration of three terms on one dimensional data.

$\{\mathbf{Y}_i\}_{i=1}^n$ could be self-represented by $\mathbf{X}$ with non-linear metric. Therefore, (11) could be rewritten as below:

$$\min_{\mathbf{X}} \quad \lambda f(\mathbf{X}) + \sum_{i=1}^{n} \|\mathbf{Y}_i \ominus \biguplus_{j=1}^{n} \mathbf{Y}_j \circledast x_{ji}\|_{\mathcal{G}}. \tag{12}$$

As we discussed in Section 1, to obtain a better low-rank representation matrix, we intend to replace the traditional nuclear norm by Schatten-p norm $(0 < p < 1)$ which is a non-convex low rank norm. Then the (12) could be rewritten as below:

$$\min_{\mathbf{X}} \quad \lambda \sum_{i=1}^{r} \rho(\sigma_i(\mathbf{X})) + \sum_{i=1}^{n} \|\mathbf{Y}_i \ominus \biguplus_{j=1}^{n} \mathbf{Y}_j \circledast x_{ji}\|_{\mathcal{G}}, \tag{13}$$

where r represents the rank of matrix $\mathbf{X}$, $\rho(\cdot) : \mathbb{R}^+ \to \mathbb{R}^+$ is a proper and lower semicontinuous on $[0, +\infty)$. In this paper, we choose the Schatten-p norm $(0 < p < 1)$ as the low rank constraint. Figure 2 is an illustration of three terms on one dimensional data. The function of Schatten-p norm $(p = \frac{1}{2})$ is a neutralization between rank norm $(p = 0)$ and nuclear norm $(p = 1)$, which could increase the punishment on smaller values and decrease the punishment on larger ones simultaneously. Then, to further reduce the influence of smaller singular values to the matrix $\mathbf{X}$, we introduce the reweighted approach inspired by [36] and (13) could be rewritten as below:

$$\min_{\mathbf{X},\mathbf{w}} \quad \lambda \sum_{i=1}^{r} w_i \rho(\sigma_i(\mathbf{X})) + \sum_{i=1}^{n} \|\mathbf{Y}_i \ominus \biguplus_{j=1}^{n} \mathbf{Y}_j \circledast x_{ji}\|_{\mathcal{G}}, \tag{14}$$

where $\mathbf{w} = (w_1, w_2, ..., w_r)$ represents the weighting vector with $w_1 < w_2 < ... < w_r$, which could be regarded as the adaptive weights for the singular values of $\mathbf{X}$.

Therefore, the first term in (14) could be regarded as a reweighted Non-convex and Non-smooth Rank Minimization based low rank norm, and we call this the reweighted Non-convex and Non-smooth Rank Minimization based clustering model on Grassmann manifold (G-RNNRM).

The objective function in (14) is hard to solve directly owing to the non-linear metric on Grassmann manifold. According to the property and definition in Section 2, we could use the embedding distance defined in (2) to replace the construction error in (14) as below:

$$\begin{aligned} \|\mathbf{Y}_i \ominus \biguplus_{j=1}^{n} \mathbf{Y}_j \circledast x_{ji}\|_{\mathcal{G}} &= \mathrm{dist}_g^2(\mathbf{Y}_i, \biguplus_{j=1}^{n} \mathbf{Y}_j \circledast x_{ji}) \\ &= \|\mathbf{Y}_i\mathbf{Y}_i^T - \sum_{j=1}^{n} x_{ji}\mathbf{Y}_j\mathbf{Y}_j^T\|_F^2. \end{aligned} \tag{15}$$

With this measurement, the function in (14) could be rewritten as below:

$$\min_{\mathbf{X},\mathbf{w}} \quad \lambda \sum_{i=1}^{r} w_i \rho(\sigma_i(\mathbf{X})) + \sum_{i=1}^{n} \|\mathbf{Y}_i\mathbf{Y}_i^T - \sum_{j=1}^{n} x_{ji}\mathbf{Y}_j\mathbf{Y}_j^T\|_F^2. \tag{16}$$

Denoting $g_{ij} = \mathrm{tr}((\mathbf{Y}_j^T\mathbf{Y}_i)(\mathbf{Y}_i^T\mathbf{Y}_j))$ according to [20], we could rewrite (16) as below:

$$\min_{\mathbf{X},\mathbf{w}} \quad \lambda \sum_{i=1}^{r} w_i \rho(\sigma_i(\mathbf{X})) + \mathrm{tr}(\mathbf{X}^T\mathbf{G}\mathbf{X}) - 2\mathrm{tr}(\mathbf{G}\mathbf{X}), \tag{17}$$

where matrix $\mathbf{G} = \{g_{ij}\}_{n\times n} \in \mathbb{R}^{n\times n}$ is a symmetric matrix, tr represents the inner product of matices. With these transformation, the original non-linear self-representation function in (14) could be converted into a linear one, which could be solved by standard optimization.

4.2 Optimization of G-RNNRM

The proposed G-RNNRM model is a complicated optimization problem which is difficult to solve directly. According to [36], we have the following proposition:

Proposition 1: Let $\rho(\cdot) : \mathbb{R}^+ \to \mathbb{R}^+$ be a function such that the proximal operator denoted by $\mathrm{Prox}_\rho()$ is monotone. For $\lambda > 0$, let $\mathbf{Z} = \mathbf{U_Z S_Z V_Z^T}$, where $\mathbf{S_Z} = diag(\sigma_1(\mathbf{Z}), \sigma_2(\mathbf{Z}), ..., \sigma_r(\mathbf{Z}))$ and all weighting values satisfy $0 \le w_1 \le w_2 \le ... \le w_r$. Then, the optimal solution to $\mathbf{X}$ could be written as below:

$$\begin{aligned} \sigma_i(\mathbf{X}) &\in \mathrm{Prox}_\rho(\sigma_i(\mathbf{Z})) \\ &= \arg \min_{\sigma_i(\mathbf{X})\ge 0} \lambda w_i \rho(\sigma_i(\mathbf{X})) + \frac{1}{2}(\sigma_i(\mathbf{X}) - \sigma_i(\mathbf{Z}))^2. \end{aligned} \tag{18}$$

According to Proposition 1, we could obtain the optimal solution to (17). First, let $h(\mathbf{X}) = \mathrm{tr}(\mathbf{X}^T\mathbf{G}\mathbf{X}) - 2\mathrm{tr}(\mathbf{G}\mathbf{X})$. Then we linearize $h(\mathbf{X})$ at $\mathbf{X}^{(t)}$ and add a proximal term as below:

$$h(\mathbf{X}) \approx h(\mathbf{X}^{(t)}) + \langle \nabla h(\mathbf{X}^{(t)}), \mathbf{X} - \mathbf{X}^{(t)} \rangle + \frac{\mu}{2}\|\mathbf{X} - \mathbf{X}^{(t)}\|_F^2, \tag{19}$$

where μ is larger than the Lipschitz constant L_h, $\nabla h(\mathbf{X}^{(t)})$ represents the first derivative of $h(\mathbf{X})$ at $\mathbf{X}^{(t)}$. In our paper, $\mu = 2\|\mathbf{G}\|_2$ and $\nabla h(\mathbf{X}^{(t)}) = 2\mathbf{G}\mathbf{X}^{(t)} - 2\mathbf{G}$. Therefore, we could obtain the update function of $\mathbf{X}$ in (17) as below:

$$\begin{aligned}\mathbf{X}^{(t+1)} = \arg\min_{\mathbf{X}} \quad & \lambda\sum_{i=1}^{r} w_i\rho(\sigma_i(\mathbf{X})) + h(\mathbf{X}^{(t)}) \\ & + \langle \nabla h(\mathbf{X}^{(t)}), \mathbf{X} - \mathbf{X}^{(t)} \rangle \\ & + \frac{\mu}{2}\|\mathbf{X} - \mathbf{X}^{(t)}\|_F^2. \end{aligned} \tag{20}$$

(20) could be rewritten as below:

$$\mathbf{X}^{(t+1)} = \arg\min_{\mathbf{X}} \quad \lambda\sum_{i=1}^{r} w_i^{(t)}\rho(\sigma_i(\mathbf{X})) + \frac{\mu}{2}\|\mathbf{X} - \mathbf{Z}^{(t)}\|_F^2, \tag{21}$$

where $\mathbf{Z}^{(t)} = \mathbf{X}^{(t)} - \frac{\nabla h(\mathbf{X}^{(t)})}{\mu}$. According to (18), we could obtain the solution to (21) by solving the optimal singular values and we select $p = \frac{1}{2}$ for examples. The closed-form of singular values is as below [27, 37]:

- if $\sigma_i(\mathbf{Z}^{(t)}) > \varphi(\lambda w_i^{(t)})$:

$$\sigma_i(\mathbf{X}^{(t+1)}) = \frac{2}{3}\sigma_i(\mathbf{Z}^{(t)})(1 + \cos(\frac{2}{3}(\pi - \phi(\sigma_i(\mathbf{Z}^{(t)}))))), \tag{22}$$

- otherwise:

$$\sigma_i(\mathbf{X}^{(t+1)}) = 0, \tag{23}$$

where $\phi(\sigma_i(\mathbf{Z}^{(t)})) = \arg\cos(\frac{\lambda w_i^{(t)}}{4}(\frac{\sigma_i(\mathbf{Z}^{(t)})}{3})^{-\frac{3}{2}})$, and $\varphi(\lambda w_i^{(t)}) = \frac{3\sqrt[3]{2}}{4}(2\lambda w_i^{(t)})^{\frac{2}{3}}$. After updating $\mathbf{X}^{(t+1)}$, we could update the weighting vector $\mathbf{w}$ as below:

$$w_i^{(t+1)} \in \partial\rho(\rho(\sigma_i(\mathbf{X}^{(t+1)}))), i = 1, 2, ..., r. \tag{24}$$

In our algorithm, the stopping criterion is measured by the following condition:

$$\|\mathbf{X}^{(t+1)} - \mathbf{X}^{(t)}\| \leq \varepsilon. \tag{25}$$

We summarized all update steps in Algorithm 1 for the proposed G-RNNRM.

4.3 Converge and Complexity Analysis

For the proposed G-RNNRM, we first transform the original complex objective function (14) into the standard function (21). Therefore, the algorithm convergence analysis in [36] could be applied to Algorithm 1. Besides, these algorithms always converge in our experiments.

Further, we discuss the complexity of the proposed model. In each iteration step, the complexity of updating $\mathbf{X}$ is $\mathcal{O}(nr^2)$. The complexity of updating $\mathbf{w}$

Algorithm 1 The solution to G-RNNRM

Require: The Grassmann sample set $\mathcal{Y} = \{\mathbf{Y}_1, \mathbf{Y}_2, ..., \mathbf{Y}_n\}$, the parameters λ.
Initialize : $\mathbf{X}^{(0)} \in \mathbb{R}^{n\times n}$ and $\mathbf{w}^{(0)} \in \mathbb{R}^r$, $\varepsilon = 10^{-4}$, the number of maximum iteration $MaxIter = 1000$.
Calculate matrix $\mathbf{G}$ by $g_{ij} = \mathrm{tr}((\mathbf{Y}_j^T\mathbf{Y}_i)(\mathbf{Y}_i^T\mathbf{Y}_j))$;
$t = 0$;
while not converged and $t \leq MaxIter$ **do**
  Update $\mathbf{X}$ by (18) to (23);
  Update $\mathbf{w}$ by (24);
  Check the convergence condition defined as (25);
  $t = t + 1$.
end while
Ensure:
The matrices $\mathbf{X}$ and weighting vector $\mathbf{w}$.

is $\mathcal{O}(r)$. Therefore, the total complexity of G-RNNRM is $\mathcal{O}(nr^2 + r)$. We also list the complexities of other methods in Table 1. Meanwhile, we test all the methods on Extended Yale B dataset as an example. The running time is also shown in Table 1. It demonstrates that the proposed G-RNNRM have acceptable executive time. All methods are coded in Matlab R2014a and implemented on an Intel(R) Xeon(R) Gold 5115 CPU @ 3.60GHz CPU machine with 8G RAM.

Table 1. The complexity and running time (second) on Extended Yale B dataset of various methods.

Method	Complextiy	Running Time
SSC	$\mathcal{O}(tn^2(1+n))$	2.78
LRR	$\mathcal{O}(2tn^3)$	27.67
LS3C	$\mathcal{O}(tn^2(sn^2+1))$	39.02
SCGSM	$\mathcal{O}(m^3p^3d^2t(n+d))$	2198.84
G-KM	$\mathcal{O}(3n^3)$	1.43
G-CLRSR	$\mathcal{O}(tn^2(3n+1))$	108.32
G-LRR	$\mathcal{O}(2tn^3)$	25.42
G-PSSVLR	$\mathcal{O}(2tn^3)$	29.94
G-RNNRM	$\mathcal{O}(t(nr^2+r))$	32.74

5 Experimental results

We test the proposed method for facial imageset and action video clip clustering tasks, and the whole experiments are evaluated on four datasets, including the Extended Yale B face dataset [38], CMU-PIE face dataset [39], Ballet action dataset [40] and SKIG gesture dataset [41]. Facial images are usually corrupted by rich expression, different pose, various illustration intensities and directions, which affect the performance of facial clustering task. Action clustering task is also difficult since human actions are always captured by large range of scenes and viewpoints. Besides, small movement, the illumination and background would change great. These problems would make the clustering task challenging.

The performance of the proposed method is compared with some state-of-the-art clustering algorithms, such as SSC [10], LRR [11], LS3C [42], SCGSM [18], G-KM [19], G-LRR [20], G-PSSVLR [23] and G-CLRSR [22]. In our method, after learning the representation $\mathbf{X}$, we use the NCut method [9] to obtain the final clustering results. The clustering results are measured by the clustering Accuracy (ACC), Normalized Mutual Information (NMI), Rand Index(RI) and Purity (PUR). The details of data setting and results analysis are given below.

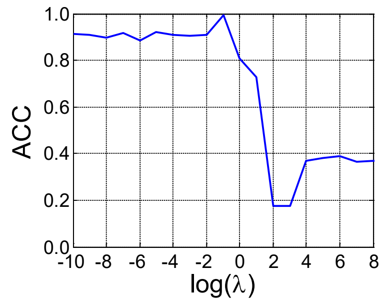
5.1 Data setting

In our experiments, we first transform each image into a m-dimension vector. For vector based LRR, SSC and LS3C methods, we stack all image vectors from the same imageset as a long vector and adopt PCA to reduce the dimension which equals to the dimension of PCA components retaining 95% of its variance energy. For other Grassmann manifold based methods, we form all image vectors from the same imageset as a matrix. Then SVD is applied on the matrix and we pick up the first p columns of the left singular matrix as a sample data on Grassmann manifold $\mathcal{G}(p, m)$. The detailed data setting for Grassmann manifold based methods is as below:

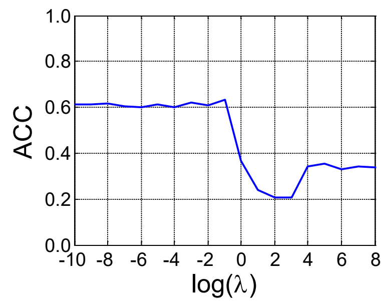
Extended Yale B face dataset Extended Yale B face dataset contains 2,414 frontal face images of $c = 38$ subjects under different light directions and illumination conditions, and each subject has about 64 images. In our experiments, we resize images into 20×20. To construct Grassmannian data, we randomly choose 8 images from the same subject to construct an image-set for clustering. We set the subspace dimension of Grassmann manifolds as $p = 4$.

CMU-PIE face dataset The CMU-PIE face dataset is composed of 68 subjects with 1632 front face images. Each subject has 42 images under different lighting conditions. In our experiments, each grey image is down-sampled as a fixed size of 32×32. Every 4 images from the same subject are selected to form an imageset sample. The image-sets are also represented as Grassmannian points $\mathcal{G}(4, 1024)$.

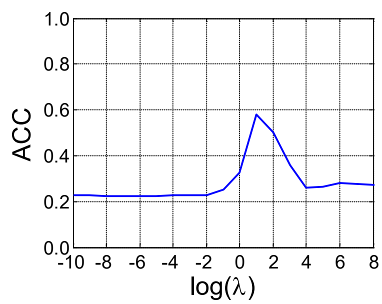
Ballet action dataset This dataset comprises of 8 basic ballet actions performed by 3 persons. The 8 ballet actions are *left-toright hand opening, right-to-left hand opening, standing hand opening, leg swinging, jumping, turning, hopping* and *standing still*. In our experiments, we resize each image into 30×30 and divide each video clip into sections of 12 images to form the imagesets. We set the dimension of Grassmann manifolds as $p = 6$, then we construct a Grassmann manifolds $\mathcal{G}(6, 900)$ for clustering.



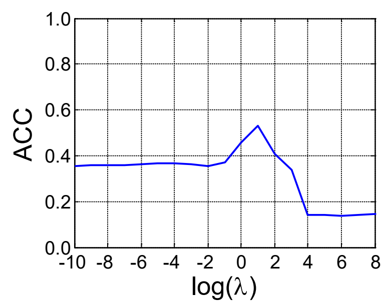
(a)



(b)



(c)



(d)

Fig. 3. The clustering accuracy of the proposed method on four datasets with different λ: (a) Extended Yale B dataset; (b) CMU-PIE face dataset; (c) Ballet action dataset; (d) SKIG gesture dataset.

SKIG gesture dataset The SKIG dataset consists of 1080 RGB-D videos collected from 6 subjects captured by a Kinect sensor. In this dataset, there are 10 gesture types: *circle, triangle, up-down, rightleft, wave, Z, cross, comehere, turn-around,* and *pat*. All the gestures are performed by fist, finger, and elbow, respectively, under three backgrounds (wooden board, white plain paper, and paper with characters) and two illuminations (strong light and poor light). In our experiments, each image is resized as 24×32 and each clip is regarded as an imageset.

5.2 Parameters setting

To obtain the suitable parameter for the proposed model, the influence of λ on the clustering accuracy is learned by some pre-experiments. λ is tuned within $[10^{-10}, 10^{8}]$ and Figure 3 shows the influence of λ on Extended Yale B dataset, CMU-PIE face dataset, Ballet action dataset and SKIG gesture dataset respectively. The parameter for each dataset is set as: $\lambda = 0.1$ for Extended Yale B dataset; $\lambda = 0.15$ for CMU-PIE dataset; $\lambda = 20$ for Ballet dataset; $\lambda = 2.5$ for SKIG dataset.

Table 2. The clustering results of various methods on four datasets: (a) Accuracy (ACC), (b) Normalized Mutual Information (NMI), (c) Rand Index(RI) and (d) Purity (PUR).

(a)

Method	SSC	LRR	LS3C	SCGSM	G-KM	G-CLRSR	G-LRR	G-PSSVLR	G-RNNRM
Extended Yale B	0.4032	0.4659	0.2461	0.7946	0.8365	0.8194	0.8135	0.9035	**0.9872**
CMU-PIE	0.5231	0.4034	0.2761	0.5732	0.6025	0.6289	0.6153	0.6213	**0.6418**
Ballet	0.2962	0.2923	0.4262	0.5613	0.5699	0.5931	0.5912	0.6013	**0.6143**
SKIG	0.3892	0.2537	0.2941	0.3716	0.5308	0.5083	0.5022	0.5502	**0.5949**
avg.	0.4029	0.3538	0.3106	0.5733	0.6349	0.6374	0.6306	0.6691	**0.7127**

(b)

Method	SSC	LRR	LS3C	SCGSM	G-KM	G-CLRSR	G-LRR	G-PSSVLR	G-RNNRM
Extended Yale B	0.6231	0.6813	0.4992	0.9326	0.9341	0.9103	0.8903	0.9262	**0.9921**
CMU-PIE	0.7865	0.7321	0.6313	0.5736	0.5976	0.8132	0.8103	0.7926	**0.8341**
Ballet	0.2813	0.2910	0.4370	0.5646	0.5779	0.5862	0.5762	0.5837	**0.6405**
SKIG	0.4762	0.3343	0.3421	0.5367	0.5671	0.5679	0.5450	0.5692	**0.6342**
avg.	0.5418	0.5097	0.4774	0.6519	0.6692	0.7194	0.7055	0.7179	**0.7752**

(c)

Method	SSC	LRR	LS3C	SCGSM	G-KM	G-CLRSR	G-LRR	G-PSSVLR	G-RNNRM
Extended Yale B	0.9503	0.9581	0.9525	0.9537	0.9647	0.9772	0.9793	0.9812	**0.9868**
CMU-PIE	0.9727	0.9752	0.9737	0.9235	0.9482	0.9721	0.9811	0.9727	**0.9878**
Ballet	0.8135	0.8273	0.8202	0.8301	0.8319	0.8321	0.8377	0.8382	**0.8496**
SKIG	0.8595	0.7223	0.8160	0.8135	0.8392	0.8577	0.8782	0.8763	**0.8910**
avg.	0.8990	0.8707	0.8906	0.8802	0.8960	0.9098	0.9191	0.9171	**0.9288**

(d)

Method	SSC	LRR	LS3C	SCGSM	G-KM	G-CLRSR	G-LRR	G-PSSVLR	G-RNNRM
Extended Yale B	0.4347	0.4932	0.2375	0.8104	0.8582	0.8375	0.8275	0.9017	**0.9937**
CMU-PIE	0.5371	0.4415	0.2695	0.5637	0.5976	0.6559	0.6429	0.6711	**0.6941**
Ballet	0.4175	0.4302	0.4581	0.5854	0.5867	0.6281	0.6298	0.6376	**0.6578**
SKIG	0.4352	0.3577	0.3102	0.3502	0.6097	0.5819	0.5322	0.6268	**0.6630**
avg.	0.4311	0.4307	0.3188	0.5774	0.6631	0.6759	0.6581	0.7093	**0.7522**

5.3 Results analysis

We show clustering results in Tables 2. Each clustering experiment is repeated 20 times and the average results are reported. The best results are bold, the second ones are underlined. From the results, Grassmann manifold representation based methods always have better performances than the vectors based ones (SSC, LRR, LS3C), which explains that the manifold representation have the advantage of revealing the complicated relationship within the imageset or video data effectively. In all the methods, the low rank representation based ones always obtain the top results, which shows the benefit of low rank representation. From the results, our proposed G-RNNRM always obtains the best results. Especially, G-RNNRM outperform the second ones with about 4 to 5 percentage points gap in terms of ACC, NMI and PUR on avenges respectively. The significant improvement of our method is analyzed and own to the superiority that the proposed method not only adopts the Non-convex Non-smooth Rank Minimization but also constructs reweighted approach.

6 Conclusion

In this paper, we propose a new low rank model on Grassmann manifold for high-dimension data clustering task. Instead of the traditional convex nuclear norm, we adopt non-convex and non-smooth rank minimization approach to formulate a novel clustering model on Grassmann manifold with non-linear metric. Further, reweighting approach has been introduced to obtain a better low-rank representation matrix. In addition, an effective alternative algorithm is proposed as solution. The proposed model has been evaluated on four public datasets. The experimental results show that our proposed model outperforms state-of-the-art ones.

Acknowledgements

The research project is supported by National Natural Science Foundation of China under Grant No. 61902053, 61672071, U1811463, U19B2039, 61632006, 61972067, 91748104.

References

1. Biswas, A., Jacobs, D.: Active image clustering: Seeking constraints from humans to complement algorithms. In: Procceedings of IEEE Conference on Computer Vision and Pattern Recognition. (2012) 2152–2159
2. Krinidis, S., Chatzis, V.: A robust fuzzy local information c-means clustering algorithm. IEEE Transactions on Image Processing **19** (2010) 1328–1337
3. You, C., Robinson, D., Vidal, R.: Scalable sparse subspace clustering by orthogonal matching pursuit. In: Procceedings of IEEE Conference on Computer Vision and Pattern Recognition. (2016) 3918–3927
4. Liu, H., Shao, M., Li, S., Fu, Y.: Infinite ensemble for image clustering. In: Proceedings of ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. (2016) 1745–1754
5. Gruber, A., Weiss, Y.: Multibody factorization with uncertainty and missing data using the em algorithm. In: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition. (2004) 707–714
6. Ho, J., Yang, M.H., Lim, J., Lee, K.C., Kriegman, D.: Clustering appearances of objects under varying illumination conditions. In: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition. (2003) 11–18
7. Ma, Y., Yang, A.Y., Derksen, H., Fossum, R.: Estimation of subspace arrangements with applications in modeling and segmenting mixed data. SIAM Review **50** (2008) 413–458
8. Elhamifar, E., Vidal, R.: Sparse subspace clustering. In: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition. (2009) 1–8
9. Shi, J., Malik, J.: Normalized cuts and image segmentation. IEEE Transactions on Pattern Analysis and Machine Intelligence **22** (2010) 888–905
10. Elhamifar, E., Vidal, R.: Sparse subspace clustering: Algorithm, theory, and applications. IEEE Transactions on Pattern Analysis and Machine Intelligence **35** (2013) 2765–2781

11. Liu, G., Lin, Z., Yu, Y.: Robust subspace segmentation by low-rank representation. In: Proceedings of International Conference on Machine Learning. (2010) 663–670
12. Liu, G., Yan, S.: Latent low-rank representation for subspace segmentation and feature extraction. In: Procceedings of International Conference on Computer Vision. (2011) 1615–1622
13. Zhang, H., Lin, Z., Zhang, C., Gao, J.: Robust latent low rank representation for subspace clustering. Neurocomputing **145** (2014) 2765–2781
14. Patel, V.M., Vidal, R.: Kernel sparse subspace clustering. In: Proceedings of IEEE International Conference on Image Processing. (2014) 2849–2853
15. Xiao, S., Tan, M., Xu, D., Dong, Z.Y.: Robust kernel low-rank representation. IEEE Transactions on Neural Networks and Learning Systems **27** (2015) 2268–2281
16. Wu, T., Bajwa, W.U.: Learning the nonlinear geometry of high-dimensional data: Models and algorithms. IEEE Transactions on Signal Processing **63** (2015) 6229–6244
17. Wang, B., Hu, Y., Gao, J., Sun, Y., Yin, B.: Laplacian lrr on product grassmann manifolds for human activity clustering in multicamera video surveillance. IEEE Transactions on Circuits and Systems for Video Technology **27** (2017) 554–566
18. Turaga, P., Veeraraghavan, A., Srivastava, A., Chellappa, R.: Statistical computations on grassmann and stiefel manifolds for image and video-based recognition. IEEE Transactions on Pattern Analysis and Machine Intelligence **33** (2011) 2273–2286
19. Shirazi, S., Harandi, M.T., Sanderson, C., Alavi, A., Lovell, B.C.: Clustering on grassmann manifolds via kernel embedding with application to action analysis. In: Proceedings of IEEE International Conference on Image Processing. (2012) 781–784
20. Wang, B., Hu, Y., Gao, J., Sun, Y., Yin, B.: Low rank representation on Grassmann manifolds. In: Proceedings of Asian Conference on Computer Vision. (2014) 81–96
21. Liu, T., Shi, Z., Liu, Y.: Visual clustering based on kernel sparse representation on grassmann manifolds. In: Proceedings of IEEE 7th Annual International Conference on CYBER Technology in Automation, Control, and Intelligent Systems. (2017) 920–925
22. Wang, B., Hu, Y., Gao, J., Sun, Y., Yin, B.: Cascaded low rank and sparse representation on grassmann manifolds. In: Proceedings of International Joint Conference on Artificial Intellignce. (2018) 2755–2761
23. Wang, B., Hu, Y., Gao, J., Sun, Y., Yin, B.: Partial sum minimization of singular values representation on grassmann manifolds. ACM Transactions on Knowledge Discovery from Data **12** (2018) Article 13
24. Piao, X., Hu, Y., Gao, J., Sun, Y., Yin, B.: Double nuclear norm based low rank representation on grassmann manifolds for clustering. In: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition. (2019) 12075–12084
25. Wang, B., Hu, Y., Gao, J., Sun, Y., Yin, B.: Localized lrr on grassmann manifold: An extrinsic view. IEEE Transactions on Circuits and Systems for Video Technology **28** (2018) 2524–2536
26. Candès, E.J., Wakin, M.B., Boyd, S.P.: Enhancing sparsity by reweighted ℓ_1 minimization. Journal of Fourier analysis and applications **14** (2008) 877–905
27. Zhang, H., Yang, J., Shang, F., Gong, C., Zhang, Z.: LRR for subspace segmentation via tractable schatten-p norm minimization and factorization. IEEE Transactions on Cybernetics **49** (2019) 172–1734
28. Nie, F., Hu, Z., Li, X.: Matrix completion based on non-convex low rank approximation. IEEE Transactions on Image Processing **28** (2019) 2378–2388

29. Malek-Mohammadi, M., Babaie-Zadeh, M., Skoglund, M.: Performance guarantees for schatten-p quasi-norm minimization in recovery of low-rank matrices. Signal Processing **114** (2015) 225–230
30. Lu, C., Tang, J., Yan, S., Lin, Z.: Nonconvex nonsmooth low rank minimization via iteratively reweighted nuclear norm. IEEE Transactions on Image Processing **25** (2016) 829–839
31. Absil, P.A., Mahony, R.E., Sepulchre, R.: Optimization Algorithms on Matrix Manifolds. Princeton University Press (2009)
32. Goh, A., Vidal, R.: Clustering and dimensionality reduction on riemannian manifolds. In: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition. (2008) 1–7
33. Harandi, M., Sanderson, C., Shen, C., Lovell, B.C.: Dictionary learning and sparse coding on grassmann manifolds: An extrinsic solution. In: Procceedings of International Conference on Computer Vision. (2013) 3120–3127
34. Lu, C.Y., Min, H., Zhao, Z.Q., Zh, L., Huang, D.S., Yan, S.: Robust and efficient subspace segmentation via least squares regression. In: Proceedings of European Conference on Computer Vision. (2012) 347–360
35. Oh, T.H., Tai, Y.W., Bazin, J.C., Kim, H., Kweon, I.S.: Partial sum minimization of singular values in robust pca: Algorithm and applications. IEEE Transactions on Pattern Analysis and Machine Intelligence **38** (2015) 744–758
36. Zhang, H., Gong, C., Qian, J., Zhang, B., Xu, C., Yang, J.: Efficient recovery of low-rank matrix via double nonconvex nonsmooth rank minimization. IEEE Transactions on Neural Networks and Learning Systems **30** (2019) 2916–2925
37. Cao, W., Sun, J., Xu, Z.: Fast image deconvolution using closed-form thresholding formulas of $l_q(q = 1/2, 2/3)$ regularization. Journal of Visual Communication and Image Representation **24** (2013) 31–41
38. Lee, K.C., Ho, J., Kriegman, D.J.: Acquiring linear subspaces for face recognition under variable lighting. IEEE Transactions on Pattern Analysis and Machine Intelligence **27** (2005) 684–698
39. Sim, T., Baker, S., Bsat, M.: The cmu pose, illumination, and expression database. IEEE Transactions on Pattern Analysis and Machine Intelligence **25** (2003) 1615–1618
40. Fathi, A., Mori, G.: Action recognition by learning mid-level motion features. In: Procceedings of IEEE Conference on Computer Vision and Pattern Recognition. (2008) 1–8
41. Liu, L., Shao, L.: Learning discriminative representations from RGB-D video data. In: Procceedings of International Joint Conference on Artificial Intelligence. (2013) 1493–1500
42. Patel, V.M., Nguyen, H.V., Vidal, R.: Latent space sparse subspace clustering. In: Procceedings of IEEE International Conference on Computer Vision. (2013) 225–232