

Affinity Matrix with Large Eigenvalue Gap for graph-based Subspace Clustering and Semi-Supervised Classification

Xiaofang Liu, Jun Wang, Dansong Cheng, Feng Tian and Yongqiang Zhang

Abstract—In the graph-based learning method, the data graph or similarity matrix reveals the relationship between data, and reflects similar attributes within a class and differences between classes. Inspired by Davis-Kahan Theorem that the stability of matrix eigenvector space depends on its spectral distance (i.e. its eigenvalue gap), in this paper, we propose a global local affinity matrix model with low rank subspace sparse representation (GLAM-LRSR) based on global information of eigenvalue gap and local distance between samples. This method approximate the similarity matrix with ideally diagonal block structure from the perspective of maximizing the eigenvalue gap, and the local distance between data is utilized as a regular term to prevent the eigenvalue gap from being too large to ensure the efficacy of similarity matrix. We have shown that the combination of subspace (LRSR) partitioning method such as Sparse Subspace Clustering(SSC) and the similarity matrix constructed by GLAM can improve the accuracy of subspace clustering, and that the similarity matrix constructed by GLAM-LRSR can be successfully applied to graph-based semi-supervised classification task. Our experiments on synthetic data as well as the real-world datasets for face clustering, face recovery and motion segmentation have clearly demonstrate the significant advantages of GLAM-LRSR and its effectiveness.

Index Terms—affinity matrix, subspace clustering, semi-supervised Classification, low rank representation, sparse representation

I. INTRODUCTION

HIGH-volume and high-dimensional datasets such as videos and millions of images are ubiquitous due to advances in sensing and storage technology and dramatic growth in applications [1]. Subspace segmentation refers to the problem of segmenting high-dimensional data according to their underlying subspaces. Recent subspace clustering algorithms are mainly categorized into algebraic methods, statistical methods and graph-based methods. Generalized Principal Component Analysis (GPCA) [2] is an algebraic method by fitting the data with polynomials. The drawback is that it is hard to estimate polynomials when data contains large noises. Statistical approaches usually use independent samples drawn from a mixture of probabilistic distributions to model data generation processes. The processes can be

tackled by Expectation-Maximization (EM) algorithm with alternating between data clustering and subspaces estimation or estimating the mixture structure by iteratively finding a min-max estimation [3]. However, the optimization of statistical methods is difficult and the use of the EM also leads to a local minimum. Graph based clustering methods have attracted increasing attention over the last decade and have succeeded in many applications [4], [5], [6], [7].

One representative graph based clustering approach is the graph based subspace clustering, which may further be categorized as local spectral clustering and global spectral clustering. The local spectral clustering based approaches such as Local Subspace Affinity (LSA)[8] and Locally Linear Manifold Clustering (LLMC) [9] do not perform well on tackling with points near the intersection of subspaces. In addition, the number of neighborhood also affects the performance of these methods. As a local spectral clustering methods, Sparse Subspace Clustering (SSC) [7], [10] exploit the discriminative nature of sparse representation and discovers the sparsest representations for the data set. According to the theoretical work [7], [11], the subspace-sparse recovery holds when a certain value of the data distribution in each subspace is smaller than the smallest principal angle between each subspace and any other subspace. However, the pursuit of sparsity leads to sparse within class homogeneity so as to separate points in the same cluster into several small singleton which leads to bad results of poor recall rates. Global approaches such as Low-Rank Representation (LRR) [6], [12] are actually a general version of Robust principal component analysis (RPCA) [13], [14], which is based on minimizing the rank of the representation matrix. While RPCA is assumed that data lie in a single low-rank subspace, LRR considers data from mixed subspaces by seeking the lowest-rank representation among all the candidates that can represent the data vectors as linear combinations of the basis in a given dictionary. Although LRR recovers the affinity matrix of dense within-class affinities, the between-class are also dense especially when the subspace are not independent. However, the spectrum properties are not considered during the construction of the affinity matrix for the follow-up spectral clustering step.

To tackles the limitations above, Feng *et al.* [15] constrains the rank of Laplacian with a *Block-Diagonal* prior by optimizing the coefficient representation matrix and the Laplacian matrix alternatively. Liu *et al.* [16] assumes the affinity matrix is of low rank by regularizing a nuclear norm term of the affinity matrix. However, the proposed optimization model

The authors D Cheng, Y Zhang and J Wang are with the School of Computer Science and Technology, Harbin Institute of Technology, 92 West Dazhi Street, Nan Gang District, Harbin, 150001, P.R.China.

The authors X Liu is with The School of Electrical Engineering and Automation, Harbin Institute of Technology, P.R.China.

The author F Tian is with the Faculty of Science & Technology, Bournemouth University, UK.

Dansong Cheng, e-mail: cdsinhit@hit.edu.cn.

directly computes the pairwise distance of the original data as local constraint. Moreover, minimizing the nuclear norm of the affinity matrix tends to minimize its spectrum variance by simultaneously minimizing all the singular values, which may reduce the relative distances among k clusters. Actually, each cluster can be represented by its corresponding eigenvector and only the top k eigenvectors of the affinity matrix, span the subspaces denoting the clusters.

According to Davis-Kahan Theorem [17], the stability of matrix eigenvector space depends on its spectral distance, i.e. the size of its eigenvalue gap. Therefore, we propose to construct similarity matrix model for clustering or semi-supervised classification by using global information of spectral spacing (maximizing the affinity matrix's *eigenvalue gap*) and local distance between samples. First, a new optimization model called Global Local Affinity model (GLAM) is proposed to maximize the sum of the top k eigenvalues and minimize the sum of the rest $n-k$ ones for getting k tight clusters [18]. Second, we combine the similarity matrix constructed by GLAM with a subspace segmentation method based on low-rank subspace sparse clustering [19] seeking low-rank 'Dictionary', called as GLAM-LRSR, this method can not only recover the lowest-rank subspace but also get a relatively sparse segmentation with regard to the intersection of subspace.

II. GLOBAL AND LOCAL AFFINITY MATRIX MODEL WITH LOW RANK SUBSPACE SPARSE REPRESENTATION (GLAM-LRSR)

In this section, We propose a comprehensive model GLAM-LRSR by combining GLAM with low rank subspace sparse representation (LRSR) [19] to learn low-rank coefficient and the graph affinity matrix with large eigenvalue gap simultaneously.

A. Global and Local Affinity Matrix model with Low Rank subspace Sparse Representation (GLAM-LRSR)

The objective function of our model (GLAM-LRSR) is as follows:

$$\begin{aligned} \min_{D,E} \|Z\|_* + \lambda \|E\|_{2,1} + \min_{Z,E} \|J\|_* + \lambda \|K\|_1 \\ + \gamma (-\text{tr}(ZAZ^T) + \beta \cdot \xi_k(A)) \\ \text{s.t. } A1_n = 1_n, A \in S_n, A \geq 0, X = XZ + E \end{aligned} \quad (1)$$

The nuclear norm $\|\cdot\|_*$ and $\|\cdot\|_{2,1}$ is the surrogate of low-rankness rank (\cdot) and column sparsity $\|\cdot\|_{2,0}$ respectively. Given the parameter is self-adaptive, the combined model concludes two sub-problems, i.e. multiple subspace recovery (low-rank coding) and graph construction. So we transfer Eq. (1) into three sub-models as follows:

$$\begin{aligned} \min_{Z,E} \|Z\|_* + \lambda_2 \|E\|_{2,1} \\ \text{s.t. } X = XZ + E \end{aligned} \quad (2)$$

$$\begin{aligned} \min_{Z,E} \|J\|_* + \lambda_1 \|K\|_1 \\ \text{s.t. } A = AZ + E, Z = J, Z = K \end{aligned} \quad (3)$$

$$\begin{aligned} \min_A -\text{tr}(ZAZ^T) + \beta \cdot \xi(A) \\ \text{s.t. } A1_n = 1_n, A \in S_n, A \geq 0. \end{aligned} \quad (4)$$

In terms of Eq. (1), we set $\gamma = 0$ when dealing with Z and E thus it refers to LRR. GLAM-LRSR model consists of two consecutive stages, the low rank representation methods to get the coefficient matrix and GLAM. Similarly, we also combine GLAM with Robust Shape Interaction Matrix (RSIM) [20] as GLAM-RSIM and combine GLAM with Low Rank Representation (LRR) [6] as GLAM-LRR. The effectiveness of these approaches is evaluated in Section VI.

With the auxiliary variables J and K , the problem (Eq. (4)) is equivalent to

$$\begin{aligned} \min_A -\text{tr}(ZAZ^T) + \beta \cdot \xi_k(J) \\ \text{s.t. } A1_n = 1_n, A = J, A = K, \\ J \in S_n, K \geq 0. \end{aligned} \quad (5)$$

We introduce three Lagrangian multipliers M_1, M_2 and M_3 to remove the equality constraints in Eq. (5) and formulate the augmented Lagrangian function of Eq. (5) as follows:

$$\begin{aligned} J_{gl}(A, J, K, \{M\}_{i=1}^3) = -\text{tr}(ZAZ^T) + \beta \cdot \xi_k(J) \\ + \langle M_1, A1_n - 1_n \rangle + \langle M_2, A - J \rangle + \langle M_3, A - K \rangle \\ + \frac{\mu_1}{2} \|A1_n - 1_n\|_F^2 + \frac{\mu_2}{2} \|A - J\|_F^2 + \frac{\mu_3}{2} \|A - K\|_F^2 \end{aligned} \quad (6)$$

where μ_1, μ_2 and μ_3 are positive penalty parameters. Then ADM updates $A, J, K, \{M\}_{i=1}^3$ by minimizing the energy function J_{gl} with respect to one variable while fixing others, respectively. Specifically, the updating schemes can be formed as follows:

$$\begin{aligned} A^+ &= (Z^T Z + \mu_1 1_n 1_n^T + \mu_2 J + \mu_3 K - M_1 1_n^T - M_2 - M_3) Q, \\ J^+ &= \arg \min_{J \in S_n} \frac{1}{2} \left\| J - \left(A^+ + \frac{M_2}{\mu_2} \right) \right\|_F^2 + \beta \cdot \xi_k(J), \\ K^+ &= \chi(A^+ + M_3 / \beta_3), \\ M_1^+ &= M_1 + \beta_1 (A^+ 1_n - 1_n), \\ M_2^+ &= M_2 + \beta_2 (A^+ - J^+), \\ M_3^+ &= M_3 + \beta_3 (A^+ - K^+), \end{aligned} \quad (7)$$

where the superscript "+" means that the value is updated, $Q = (\mu_1 1_n 1_n^T + (\mu_2 + \mu_3) I)^{-1}$ and $\chi(O)$ is a matrix defined as follows:

$$\chi([O])_{ij} = \begin{cases} [O]_{ij} & \text{if } [O]_{ij} \geq 0; \\ 0 & \text{if } [O]_{ij} < 0. \end{cases} \quad (8)$$

The overall process of solving GLAM-Lagrange via ADM is shown in Algorithm 1.

B. Randomized algorithm of Eigen Decomposition

In Algorithm 1, when updating of variable J , we need to compute the eigenvalue decomposition of $n * n$ matrix. Even if deterministic methods, such as the fastest Housholder transformation method, are directly used, so the time complexity

Algorithm 1 Solving GLAM-Lagrange via ADM

```

1: Input: Representation matrix  $Z$ , number  $k$  of clusters, the parameter  $\beta > 0, \rho > 1, \varepsilon, \beta_{max}$ 
2: Output: Affinity Matrix  $A^*$ 
3: Initialization:  $J = K = M_1 = M_2 = M_3 = 0$ 
4: while not converged do
5:   Follow the updating scheme Eq. (7) (Note that the eigenvalue decomposition to solve  $J$  is done with Algorithm 2),
6:   If  $\max(\|A^+ \mathbf{1}_n - \mathbf{1}_n\|_F^2, \|A^+ - J^+\|_F^2, \|A^+ - K^+\|_F^2) < \varepsilon$  then
7:      $\beta = \min(\beta_{max}, \rho\beta)$ 
8:   else
9:     Return  $A^* = A^+$ .
10:  end
11: end while

```

is $O(\frac{4n^3}{3} + n^2)$. Our method use a random algorithm for matrix feature decomposition, which can effectively reduce the time complexity of matrix feature decomposition operations. In Algorithm 2, Phase 1 is the orthogonal basis of the subspace in which the data J is to be constructed, where the number of orthogonal bases is the rank of the matrix. The time-consuming operation in step 3 is matrix multiplication, and the time complexity is $O(n^2l)$. In fact, the subsampling randomized Fourier Transform (SRFT) is used to reduce the time complexity to $O(n^2 \log(l))$ [21]. Step 3 solves the problem of slow singular value decay of matrix J using a power iteration method (where $q=2$ or $q=4$). After the orthogonal matrix Q is obtained, a small-scale matrix $A \in R^{l \times l}$ is constructed in the second phase, which reduces the scale of subsequent matrix feature decomposition.

Algorithm 2 Randomized algorithm of Eigen Decomposition

```

1: Input: Decomposed matrix  $J$ , estimated rank  $k$ 
2: Output: Matrix composed of eigenvectors  $U$ , Diagonal Matrix Constituted by Eigenvalues  $\Lambda_J (J = U \Lambda_J U^T)$ 
3: //Phase 1: Estimate the base vector  $Q$  of matrix  $J$  by power iteration;
4: sampling a random matrix from a Gaussian distribution  $\Omega \in R^{n \times l}$ , where  $l > k$ ;
5: construct a matrix  $Y = (JJ^T)^q J \Omega$  of  $n \times l$ ;
6:  $QR$  decomposition of  $Y, Y = QR$ , and then  $\Omega \in R^{n \times l}$ , the column vector  $Y$  is the orthogonal basis of the subspace in which matrix  $Y$  is located;
7: // Phase 2: Direct feature decomposition
8: construct a small-scale matrix  $Q^T J Q$ 
9: Calculate the eigendecomposition for matrix  $A$ :  $A = V \Lambda_J V^T$ 
10: Construct an orthogonal matrix:  $U = QV$ 

```

C. Time complexity analysis

In Algorithm 1, the main time consuming computation lies in the update of variables A and J . The update of variable A only involves matrix multiplication, and its time complexity is $O(n^3 + mn^2)$. But the matrix multiplication is simple and facilitates large-scale parallelization processing. The update of variable J contains the eigenvalue decomposition of the $n \times n$ matrix, which is accelerated by algorithm 2. The time

complexity of the first phase is $O(n^2 \log(l) + nl^2)$, the time complexity of the second phase is $O(n^2 \log(l) + nl^2 + l^3)$, and the main time-consuming operation is the sixth step. Matrix multiplication $O(n^2l)$, while matrix multiplication is simple, and the actual time-consuming overhead is low. In summary, the time complexity of the overall algorithm is $O(T(n^3 + n^2 \log(l) + mn^2 + l^3))$, where T is the number of iterations. Obviously, the computation time complexity of the GLAM approach (Algorithm 1) are not more than those of the traditional method approach, such as K-nearest neighbour(KNN)[22].

III. APPLICATIONS BASED ON GRAPH LEARNING VIA GLAM-LRSR

Graph-based subspace learning methods are usually divided into three stages (as shown in Fig. 1). 1) For high-dimensional noisy data, subspace partitioning is used to obtain data representation in low-rank subspace; 2) graph construction is equivalent to the construction of similarity matrix; 3) related applications based on graph model, such as clustering, transductive semi-supervised classification and inductive semi-supervised classification. In this section, we will introduce three representative graph based learning applications, including clustering, transductive semi-supervised classification and inductive semi-supervised classification.

A. Graph based clustering

In the Algorithm 3, we show the process of the spectral clustering method of Normalized Cut [23] combined with the similarity matrix obtained by GLAM-LRSR model for subspace clustering.

B. Transductive semi-supervised classification

The given data X contains the labeled samples X_l and unlabeled samples X_u . $X = \{x_1, x_2, \dots, x_l, x_{l+1}, \dots, x_{l+u}\} = [X_l, X_u]$, the sample number $n = l + u$. Our purpose is to deduce labels from labeled sample. The similarity matrix of the graph can be easily combined with the label propagation method, such as Gaussian Harmonic Function [24]. The label probability distribution of samples is expressed by $Y = [Y_l, Y_u]$, where Y_l be the Label probability distribution of labelled samples and Y_u be the label probability distribution of

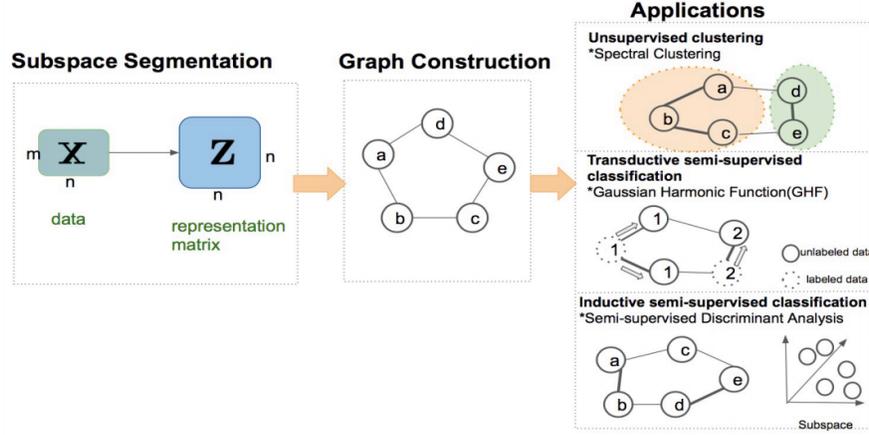


Fig. 1. The flowchart of the framework about graph based learning methods

Algorithm 3 The Subspace Clustering Framework of Normalized Cut based on GLAM-LRSR

- 1: **Input:** Data matrix $X \in R^{m \times n}$, number k of clusters
- 2: **Output:** Cluster labels for each sample in X .
- 3:
- 4: Obtain the low rank representation matrix Z using LRSR (Eq. (1)) from the given data X ,
- 5: Construct the affinity matrix A using Z via Algorithm 1,
- 6: Normalize the affinity matrix $A = D^{-1/2}AD^{-1/2}$, where $D = \text{diag}(d_1, \dots, d_N)$ is the degree matrix,
- 7: Perform eigenvalue decomposition (EVD) of A , select top k eigenvectors to $V_k \in R^{n \times k}$,
- 8: Perform KNN on V_k .

unlabelled samples, the probability value vectors of unlabelled samples can be obtained from the following formula:

$$Y_u = -Y_l L_{uu} L_{uu}^{-1} \quad (9)$$

where L is the Laplace matrix obtained by the similarity matrix W . The details of the procedure are shown in Algorithm 4. In addition to Gaussian Harmonic Function, the constructed graph can also be combined with other label propagation methods, such as local and global consistency [25] and Linear neighbor propagation [26].

C. Inductive semi-supervised classification

Semi-supervised Discriminant Analysis (SDA) [27] extends the smoothness constraint of graphs to the objective function of Linear Discriminant Analysis (LDA) [28].

$$\arg \max_a \frac{a^T S_B a}{a^T (S_T + \alpha X L X^T) a} \quad (10)$$

where X is data matrix, a is transformation matrix, the matrix S_B and S_W is the inter class and intra class matrix obtained by LDA, $S_T = S_B + S_W$ is the sum of covariance matrices with inter class and intra class, $L = I - D^{-1/2} W D^{-1/2}$ is Laplacian matrix. W is a graph similarity matrix constructed in

Algorithm 4 The framework of transductive Semi-supervised classification based on GLAM-LRSR

- 1: **Input:** Data matrix $X = [X_l, X_u] \in R^{m \times n}$, number k , labeled samples $\{X_l, Y_l\}$
- 2: **Output:** Label probability value matrix of unlabeled samples Y_u .
- 3:
- 4: Obtain the low rank representation matrix Z using LRSR (Eq. (1)) from the given data X ,
- 5: Construct the affinity matrix A via Algorithm 1,
- 6: Normalize the affinity matrix $W = D^{-1/2}AD^{-1/2}$, where $D = \text{diag}(d_1, \dots, d_N)$ is the degree matrix,
- 7: Finding Laplace matrix: $L = I - W$,
- 8: Perform the label probability value on Y_u by Eq.(9).

Algorithm 5 The framework of inductive Semi-supervised classification based on GLAM-LRSR

- 1: **Input:** Data matrix $X = [X_l, X_u] \in R^{m \times n}$, number k , labeled samples $\{X_l, Y_l\}$
- 2: **Output:** Label probability value matrix of unlabeled samples Y_u .
- 3:
- 4: Obtain the low rank representation matrix Z using LRSR (Eq. (1)) from the given data X ,
- 5: Construct the affinity matrix A via Algorithm 1,
- 6: Normalize the affinity matrix $W = D^{-1/2}AD^{-1/2}$, where $D = \text{diag}(d_1, \dots, d_N)$ is the degree matrix,
- 7: Find Laplace matrix: $L = I - W$,
- 8: Calculated the inter class covariance matrix S_B and inner class matrix S_W of labeled sample X_l , $S_T = S_B + S_W$
- 9: Eigenvalue decomposition for Eq.(11), the subspace $A_k = [a_1, \dots, a_k]$ composed of the first k eigenvectors is obtained, projecting samples to subspaces $C_k = A_k^T C$.
- 10: Perform the label probability value on Y_u by KNN.

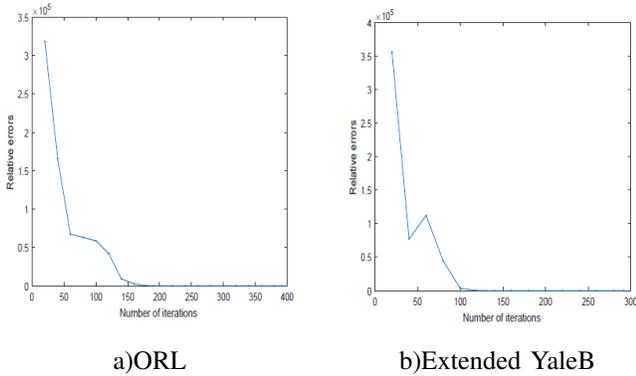


Fig. 2. Convergence curve of GLAM(GLAM-LRSR) on ORL and ExtendedYaleB (15 faces) database.

this paper. The maximization problem in Eq. (10) is converted to the maximum eigenvalue problem for solving the following:

$$S_B a = \lambda(S_T + \alpha X L X^T) a \quad (11)$$

Finally, we map all the samples to the learned subspace A_k and classify the test samples with the k-nearest neighbor method. The details of the procedure are shown in Algorithm 5.

IV. EXPERIMENTS

To verify the effectiveness of the proposed model, in this section we conduct three experiments on face clustering, face recovery, motion segmentation and semi-supervised classification.

A. Face Clustering

In this section, we evaluate the clustering performance of GLAM-LRSR, GLAM-LRR and GLAM-RSIM compared with BD-LRR [15], BD-SSC [15], LRR [6], SSC [7], RSIM [20] and Model-II [29] on the Extended Yale B dataset. The dataset consists of 192×168 pixel cropped face images of $n = 38$ individuals, where there are $N_i = 64$ frontal face images for each subject acquired under various lighting conditions. We downsample the images to 48×42 pixels and treat each 2016-dimensional vectorized image as a data point. The face images are corrupted by errors such as shades or highlight.

In this experiment, we test on five groups of n subjects, $n \in \{5, 8, 10, 15, 20\}$. And we set the parameters $\alpha = 18$ for SSC, $\lambda = 0.12$ for LRR, the parameters $\beta = [0.32, 0.25, 0.25]$ for GLAM-RSIM, GLAM-LRR and GLAM-LRSR. At the same time, this paper also compares with the method Model-II [29], which uses the characteristics of the equilibrium graph to construct the similarity matrix reflecting clustering information. Fig. 2 shows the convergence of the GLAM method on the ORL and Extended YaleB data sets, where the relative error is minimum value of $\|A1_n - 1_n\|_F, \|A - J\|_F, \|A - K\|_F$. As can be seen from the figure, the GLAM method can quickly converge within a limited number of steps.

We use both *F-measure* and NMI as the evaluation metric. Table I and Table II reveal the clustering results of various methods on five sets of Extended YaleB datasets, Table III

and Table IV show the clustering results of different methods performed on the ORL, PIE and USPS database. It can be seen from the tables that the results obtained by the traditional method KNN are the worst, possibly because the similarity criterion based on Gaussian kernel construction is not robust to noise. The coefficient representation matrix obtained by LRR algorithm is inferior to LRR-COS because the post-processing technique filters out non-zero elements on non-diagonal blocks to some extent. Using the GLAM model to construct the similarity matrix of the graph can greatly improve the accuracy of the spectral clustering algorithm. As can be seen from Table I and Table II, GLAM-LRR, GLAM-LRSR and GLAM-RSIM have a significant improvement compared to LRR, LRSR and RSIM. It is more obvious that the GLAM-LRR method has an average increase of 20% in LRR and about 10% in comparison with LRR-COS. Since BD-LRR and BD-SSC also take into account the spectral information, there is no essential difference between the GLAM model and the BD-LRR and BD-SSC models. The main difference is that the former maximizes the spectral spacing while the latter introduces constraints on the Laplacian matrix on LRR and SSC. The rank of the matrix is $n - k$. BD-LRR averages 12% higher than the *F-measure* value of the LRR method. BD-SSC also improves the clustering accuracy of SSC. However, as can be seen from Table I ~ Table IV, the GLAM based methods have higher average accuracy than BD-LRR and BD-SSC for various image data sets. For example, in the five sample sets of Extended YaleB, the average *F-measure* of GLAM-LRSR is 4.9% higher than that of BD-LRR, 6.8% higher than that of BD-SSC and 7.8% higher than that of Model-II. On PIE data sets, the *F-measure* of GLAM-LRSR is 3% higher than that of BD-LRR and 4% higher than that of Model-II method; on USPS data sets, the F-value accuracy of GLAM-RSIM method is 4.3% higher than that of BD-LRR, 6.8% higher than that of BD-SSC, 4.6% higher than that of BD-SSC and 3.8% higher than that of Model-II method; on ORL data sets, GLAM-LRSR method performs best (*F-measure* is 0.835, NMI is 0.857), followed by Model-II method (*F-measure* is 0.832, NMI is 0.851).

To more intuitively demonstrate the superiority of the GLAM approach, in Fig.3, the similarity matrix results plots for various methods on the Extended YaleB dataset (15 face objects) are shown. The graph of similarity matrix results on the extended YaleB dataset (15 face objects) can be seen in the figure. It can be seen from the figure that although the structure of the diagonal block obtained by the LRR is obvious, there are still many non-zero elements on the non-diagonal block, which may cause serious clustering errors. The similarity matrix obtained by SSC is relatively sparse, but the diagonal block structure is not obvious, and there are more non-zero elements on the non-diagonal block. The same problem occurs in the result graph of the BD-LRR, and the horizontal lines in the result graph can be clearly seen. In the result graph of BD-SSC, the diagonal block structure is very loose, which may divide the original larger cluster into multiple smaller clusters, which affects the propagation of labels in semi-supervised classification. The results of Model-II show that there are fewer non-zero elements on non-diagonal blocks, but

TABLE I
F-MEASURES OF DIFFERENT METHODS PERFORMED ON THE EXTENDED YALEB DATABASE

	5 Subjects	8 Subjects	10 Subjects	15 Subjects	20 Subjects
KNN [22]	0.76±0.32	0.75±0.40	0.70±0.35	0.73±0.40	0.72±0.11
SSC [7]	0.82±0.08	0.94±0.05	0.87±0.12	0.84±0.11	0.85±0.13
LRR [6]	0.78±0.08	0.75±0.15	0.69±0.17	0.72±0.16	0.70±0.16
LRR-COS [6]	0.85±0.18	0.95±0.05	0.84±0.12	0.82±0.17	0.82±0.16
GLAM-LRR	0.92±0.02	0.96±0.03	0.91±0.01	0.92±0.11	0.90±0.12
LRSR [19]	0.878±0.068	0.941±0.052	0.909±0.089	0.867±0.111	0.877±0.182
GLAM-LRSR	0.943±0.02	0.961±0.032	0.932±0.072	0.927±0.12	0.931±0.035
RSIM [20]	0.90±0.08	0.95±0.13	0.91±0.19	0.91±0.18	0.87±0.26
GLAM-RSIM	0.94±0.07	0.98±0.03	0.95±0.04	0.93±0.12	0.95±0.05
BD-LRR[15]	0.88±0.11	0.95±0.06	0.87±0.10	0.86±0.14	0.87±0.13
BD-SSC [15]	0.86±0.31	0.93±0.12	0.86±0.27	0.82±0.37	0.88±0.21
Model-II [29]	0.872±0.012	0.8803±0.017	0.8694±0.007	0.8576±0.0165	0.822±0.0079

TABLE II
NMI OF DIFFERENT METHODS PERFORMED ON THE EXTENDED YALEB DATABASE

	5 Subjects	8 Subjects	10 Subjects	15 Subjects	20 Subjects
KNN [22]	0.78±0.31	0.73±0.33	0.72±0.37	0.75±0.29	0.73±0.27
SSC [7]	0.79±0.21	0.93±0.17	0.82±0.27	0.842±0.17	0.832±0.21
LRR [6]	0.82±0.09	0.77±0.20	0.71±0.31	0.73±0.19	0.74±0.21
LRR-COS [6]	0.86±0.19	0.942±0.09	0.94±0.13	0.84±0.21	0.831±0.22
GLAM-LRR	0.937±0.03	0.941±0.039	0.921±0.11	0.932±0.057	0.913±0.19
LRSR [19]	0.867±0.043	0.937±0.032	0.887±0.047	0.871±0.071	0.868±0.082
GLAM-LRSR	0.940±0.05	0.953±0.056	0.941±0.063	0.932±0.08	0.919±0.020
RSIM [20]	0.91±0.17	0.941±0.21	0.921±0.18	0.93±0.21	0.939±0.21
GLAM-RSIM	0.932±0.14	0.964±0.15	0.947±0.10	0.932±0.25	0.941±0.12
BD-LRR[15]	0.891±0.19	0.931±0.09	0.853±0.09	0.862±0.19	0.83±0.04
BD-SSC [15]	0.86±0.31	0.93±0.19	0.849±0.31	0.853±0.24	0.861±0.17
Model-II [29]	0.883±0.023	0.874±0.023	0.8794±0.027	0.867±0.0166	0.846±0.0149

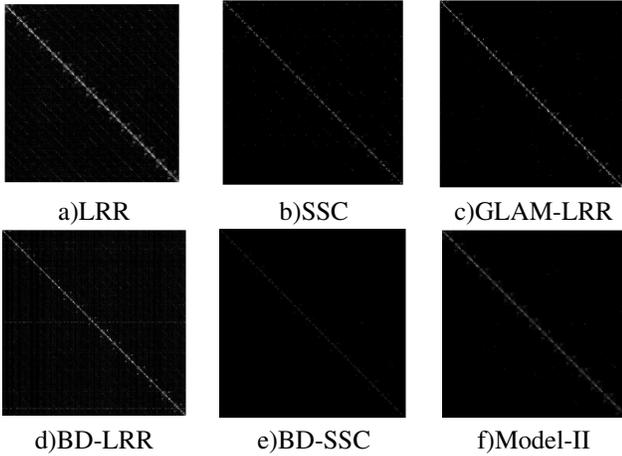


Fig. 3. Visualization of different graphs on Extended YaleB database.

the diagonal block structure is clear but the numerical value is small. In the similarity matrix obtained by GLAM-LRR, the number of non-zero elements on non-diagonal blocks is relatively small, and the structure of diagonal blocks is clear and the value is large.

B. Face Recovery

This section introduces two groups of face recovery experiments on two face datasets, Extended YaleB and ORL. For the Extended YaleB dataset, we randomly selects 640 images of 10 categories to form the matrix X by column expansion, and then randomly selects 300 samples to add random noise.

TABLE III
F-MEASURES OF DIFFERENT METHODS ON THE ORL, PIE AND USPS DATABASE.

	ORL [30]	PIE [31]	USPS [32]
KNN [22]	0.75±0.12	0.33±0.01	0.68±0.12
SSC [7]	0.80±0.03	0.49±0.04	0.69±0.11
LRR [6]	0.78±0.08	0.75±0.15	0.69±0.17
LRR-COS	0.82±0.07	0.60±0.01	0.70±0.03
GLAM-LRR	0.87±0.08	0.76±0.12	0.80±0.04
LRSR [19]	0.802±0.03	0.76±0.14	0.70±0.05
GLAM-LRSR	0.88±0.04	0.76±0.17	0.82±0.07
RSIM [20]	0.83±0.02	0.65±0.13	0.77±0.11
GLAM-RSIM	0.89±0.03	0.72±0.15	0.81±0.02
BD-LRR[15]	0.82±0.04	0.67±0.14	0.75±0.02
BD-SSC[15]	0.84±0.07	0.69±0.21	0.73±0.17
Model-II [29]	0.832±0.012	0.681±0.018	0.715±0.007

TABLE IV
NMI OF DIFFERENT METHODS ON THE ORL, PIE AND USPS DATABASE.

	ORL [30]	PIE [31]	USPS [32]
KNN [22]	0.761±0.13	0.421±0.19	0.651±0.13
SSC [7]	0.78±0.142	0.532±0.23	0.725±0.26
LRR [6]	0.737±0.02	0.627±0.12	0.721±0.102
LRR-COS	0.802±0.06	0.60±0.01	0.70±0.03
GLAM-LRR	0.865±0.102	0.71±0.13	0.77±0.13
LRSR [19]	0.81±0.08	0.71±0.05	0.747±0.12
GLAM-LRSR	0.875±0.08	0.79±0.1	0.817±0.051
RSIM [20]	0.825±0.07	0.69±0.24	0.772±0.15
GLAM-RSIM	0.876±0.15	0.76±0.06	0.825±0.17
BD-LRR[15]	0.84±0.04	0.70±0.12	0.75±0.08
BD-SSC[15]	0.83±0.13	0.653±0.14	0.735±0.12
Model-II [29]	0.851±0.012	0.674±0.0108	0.7154±0.0102

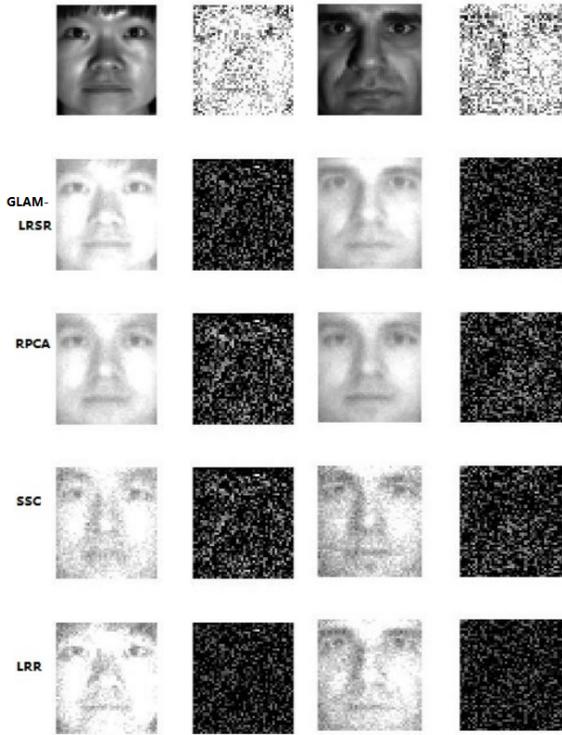


Fig. 4. Face recovery on Extended YaleB datasets (1).

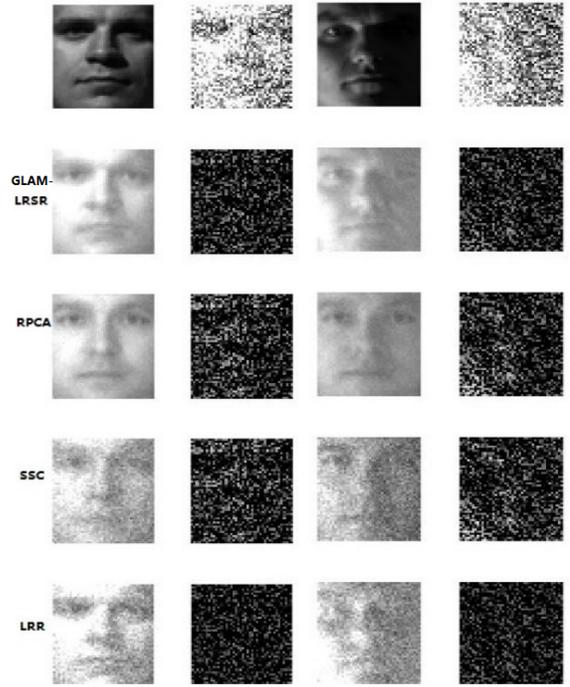


Fig. 5. Face recovery on Extended YaleB datasets (2).

The proportion of noise and data is $\frac{\|N\|_F^2}{\|X\|_F^2} \approx 1$. Note that only 300 columns selected in the noise matrix are non-zero values, and the rest are all zero.

Fig.4 and Fig.5 show the restoration results of four face images in the contaminated image set for Extended YaleB. In the first row, the first and third columns are the original face images, and the second and fourth columns are the noise added face images. The first and third columns from the second row to the fifth row are the recovered face of method GLAM-LRSR, RPCA [13], LRR [6] and SSC [7] respectively, and the second and fourth columns are the corresponding noise image. Through cross validation, the parameters of LRSR method in this experiment are set to $\lambda_1 = 0.7$; $\lambda_2 = 0.6$. The other methods all choose the parameter setting that give the best result, in which the RPCA parameter setting is $\lambda = 0.008$, the SSC parameter setting is $\alpha = 9$, and the LRR parameter setting is $\lambda = 0.015$.

RPCA method is a very classical space recovery method. It can be seen from Fig.4 and Fig.5 that although RPCA method can recover clear and high-quality face images, each damaged face image has got a similar average face, which does not restore the characteristics of different faces. The images restored by SSC and LRR still show the influence of noise. Moreover, the face restored by SSC is not the same as the original image. Because of the noise, different face images are similar to each other. The L_1 norm constraint in SSC lacks the ability to describe the data globally, so the same face may be represented by a few other face combinations. In contrast, LRSR method can recover clear and accurate face image.

For ORL dataset, we downsample images so their dimension becomes 56×46 . The sample size of ORL is a small dataset

consisting of 40 categories of faces in total, and in each category has only 10 images with different expressions. So, it's challenging to recover the damaged face. We randomly select one third of face images and add noise to them, and the ratio of noise to data is $\frac{\|N\|_F^2}{\|X\|_F^2} \approx 0.25$. Through cross validation, the parameters of GLAM-LRSR are set to $\lambda_1 = 0.8$; $\lambda_2 = 0.5$. For the other methods, the parameter setting under the best result is selected. The RPCA parameter setting is $\lambda = 0.006$, the SSC parameter setting is $\alpha = 9$, and the LRR parameter setting is $\lambda = 0.01$.

It can be seen in Fig.6 and Fig.7 that the results of LRR and RPCA methods are very fuzzy in facial contour and details. The results of SSC and GLAM-LRSR are relatively clear, but the result of SSC does not completely eliminate the noise, while GLAM-LRSR achieves relatively smooth results.

C. Motion Segmentation

The motion segmentation refers to the problem of separating a video sequence into multiple spatio-temporal regions corresponding to different rigid-body motions in the scene. The motion segmentation problem can be preceded by first extracting a set of feature points $\{\mathbf{x}_{fi} \in R^2\}_{i=1}^N$ from the video sequences $f = 1, \dots, F$ using standard tracking methods. Each data point \mathbf{y}_i , which is also called a feature trajectory, corresponds to a $2F$ -dimensional vector obtained by stacking the feature points \mathbf{x}_{fi} in the video as

$$\mathbf{y}_i \triangleq [\mathbf{x}_{1i}^T \quad \mathbf{x}_{2i}^T \quad \dots \quad \mathbf{x}_{Fi}^T]^T \in R^{2F}.$$

Then the problem is reduced to clustering these points trajectories according to different rigid-body motions. Under



Fig. 6. Face recovery on ORL dataset (1).

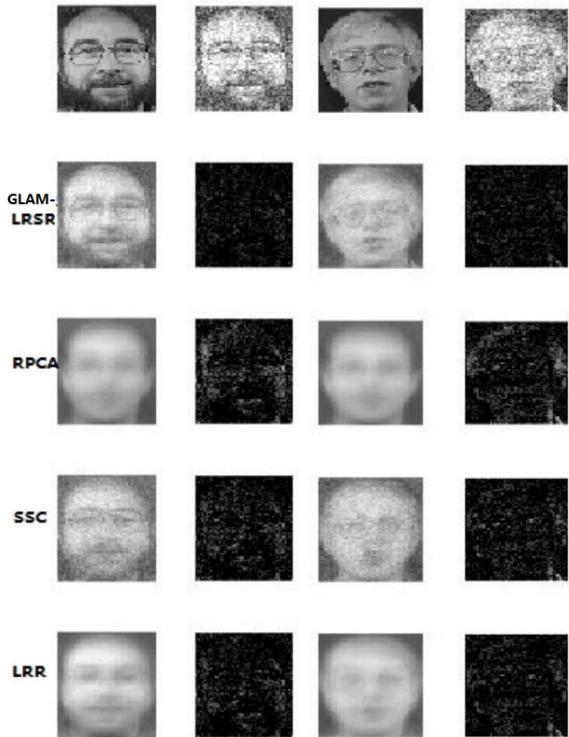


Fig. 7. Face recovery on ORL dataset (2).

the affine projection model, all feature trajectories associated with a single rigid motion lie in a linear subspace of dimension at most 4 in R^{2F} [33]. Hence, the motion segmentation becomes clustering of data points in a union of subspaces.

Here, we apply GLAM-LRR and GLAM-RSIM to motion segmentation problem and evaluate them on the Hopkins155 motion database. The database consists of 155 sequences of two and three motions which can be divided into three main categories: *Checkerboard*, *Articulated* and *Traffic* sequences. The trajectories are extracted automatically by a tracker and outliers are manually removed. Therefore, the trajectories are only corrupted by noises without any missing entries or outliers. We set the parameters $\alpha = 15$ for SSC, $\lambda = 0.6$ for LRR, $\beta = 0.15, 0.0.9, 0.09$ for GLAM-RSIM, GLAM-LRR and GLAM-LRSR, respectively. The proposed models achieve better performance than LRR, LRSR and RSIM, as shown in Table V and Table VI.

Table V and Table VI show the F -measure value and NMI average error rate for the Hopkins155 dataset clustering, where GLAM-LRSR parameters are set to $\lambda_1 = 0.2$; $\lambda_2 = 0.01$. GLAM-LRSR performs worse on the Traffic video than SSC, mainly because the moving vehicles in the video are sometimes close together, and SSC can better avoid dividing the calibration points on different vehicles into the same moving vehicle. In general, the average error rate (F -measure value) of GLAM-LRSR is 3.17%, which is not much improvement, mainly because the data set itself contains less noise, the previous method has achieved high accuracy on Hopkins 155. Secondly, when the dimension of the affine subspace is not more than 4, the correlation between the data is too strong, resulting in the accuracy of GLAM-LRSR and SSC algorithm is similar [7].

D. Semi-supervised classification

In this experiment, our model and other methods (SSC [7], LRR[6], RSIM[20], BD-LRR[15], BD-SSC [15]) on transductive semi-supervised classification and the inductive semi-supervised classification are repeated 5 times. Each time we randomly select some labeled samples for each class of images. For Extended YaleB, PIE, ORL and USPS data sets, the proportion of labeled samples is varied from 5% to 30% intervals. Fig.8 and Fig.9 show the combination of the graph (similarity matrix) model obtained by various methods on transductive semi-supervised classification and inductive semi-supervised classification, respectively. The classification accuracy is obtained by the label propagation strategy of the Gaussian harmony function [24]. The horizontal axis represents percentages of labeled samples.

In most cases, the graph model constructed by the sparse representation is better than the graph from KNN because the similarity criterion of Gauss kernel structure is not robust to noise. LRR, SSC, RSIM, BD-LRR and BD-SSC displays the noise that is implied in the data, and constructs the graph model under the representation of low rank subspace. GLAM-LRR and GLAM-RSIM are all under the low rank subspace representation, if the global and local properties of the data class space structure are taken into account. The

TABLE V
F-MEASURES OF DIFFERENT METHODS ON THE HOPKINS155 WITH 2 MOTIONS AND 3 MOTIONS RESPECTIVELY, WHERE $\beta = 0.15, 0.15, 0.09$ FOR GLAM-LRSR, GLAM-LRR AND GLAM-RSIM.

	<i>Checkerboard</i>	<i>Articulated</i>	<i>Traffic</i>	<i>Checkerboard</i>	<i>Articulated</i>	<i>Traffic</i>
	Two Motions			Three Motions		
SSC [7]	0.73±0.02	0.95±0.03	0.96±0.10	0.70±0.14	0.93±0.17	0.94±0.09
LRSR [19]	0.86±0.19	0.90±0.13	0.98±0.31	0.85±0.25	0.90±0.20	0.96±0.20
GLAM-LRSR	0.87±0.19	0.975±0.12	0.97±0.21	0.87±0.20	0.97±0.15	0.97±0.01
LRR [6]	0.82±0.21	0.90±0.13	0.98±0.31	0.81±0.25	0.88±0.20	0.96±0.22
GLAM-LRR	0.85±0.18	0.97±0.12	0.97±0.21	0.84±0.19	0.96±0.15	0.97±0.01
RSIM [20]	0.83±0.02	0.92±0.12	0.97±0.21	0.83±0.21	0.95±0.16	0.96±0.12
GLAM-RSIM	0.89±0.12	0.96±0.11	0.98±0.21	0.87±0.02	0.96±0.30	0.97±0.11
BD-LRR [15]	0.86±0.11	0.93±0.21	0.95±0.14	0.85± 0.31	0.91±0.18	0.97±0.36
BD-SSC [15]	0.87±0.21	0.90±0.33	0.91±0.21	0.83± 0.45	0.88±0.10	0.92±0.28
Model-II [29]	0.846±0.11	0.929±0.21	0.949±0.14	0.851± 0.31	0.911±0.18	0.966±0.36

TABLE VI
 NMI OF DIFFERENT METHODS ON THE HOPKINS155 WITH 2 MOTIONS AND 3 MOTIONS RESPECTIVELY, WHERE $\beta = 0.15, 0.15, 0.09$ FOR GLAM-LRSR, GLAM-LRR AND GLAM-RSIM.

	<i>Checkerboard</i>	<i>Articulated</i>	<i>Traffic</i>	<i>Checkerboard</i>	<i>Articulated</i>	<i>Traffic</i>
	Two Motions			Three Motions		
SSC [7]	0.70±0.02	0.92±0.03	0.94±0.10	0.68±0.14	0.91±0.17	0.92±0.09
LRSR [19]	0.83±0.19	0.89±0.13	0.96±0.31	0.84±0.25	0.89±0.20	0.95±0.20
GLAM-LRSR	0.86±0.19	0.962±0.12	0.96±0.21	0.86±0.20	0.96±0.15	0.96±0.01
LRR [6]	0.81±0.21	0.89±0.13	0.97±0.31	0.79±0.25	0.87±0.20	0.95±0.22
GLAM-LRR	0.852±0.18	0.96±0.12	0.965±0.21	0.84±0.19	0.956±0.15	0.967±0.01
RSIM [20]	0.83±0.02	0.912±0.12	0.957±0.21	0.823±0.21	0.945±0.16	0.956±0.12
GLAM-RSIM	0.889±0.12	0.956±0.11	0.978±0.21	0.88±0.02	0.956±0.30	0.968±0.11
BD-LRR [15]	0.856±0.11	0.932±0.21	0.935±0.14	0.845± 0.31	0.901±0.18	0.97±0.36
BD-SSC [15]	0.864±0.21	0.89±0.33	0.90±0.21	0.823± 0.45	0.878±0.10	0.912±0.28
Model-II [29]	0.846±0.11	0.923±0.21	0.945±0.14	0.851± 0.31	0.901±0.18	0.967±0.36

prominent feature of GLAM is the spectral spacing describing the properties of similarity matrix to construct a graph model of balanced class structure.

The GLAM model based on LRR and RSIM outperforms other methods in the datasets with obvious subspace structure. It performs well on Extended YaleB, PIE and ORL datasets. Because of its constraints on the global structure of the analog data class space, the GLAM model is also superior to the USPS handwritten digital set which is not obvious in the spatial structure, though the performance margin is not large.

On three face image sets (Extended YaleB, ORL, PIE), GLAM-LRR, GLAM-LRSR and GLAM-RSIM improve the accuracy based on the original algorithms (corresponding to LRR, LRSR, RSIM respectively), and the classification effect is better than other algorithms. There is little difference between the various methods when the proportion of labeled samples becomes larger. However, when there are fewer labeled samples, the GLAM method is much larger than the original method alone. In general, the accuracy of GLAM-LRR and GLAM-LRSR is about 7% higher than that of LRR and LRSR, and GLAM-RSIM is about 5% higher than RSIM. kNN performs the worst because it is sensitive to the noise contained in the data. However, on the USPS dataset, it may be because some of the numbers are similar, resulting in a similar subspace. The highest average accuracy is the kNN algorithm though GLAM-RSIM perform similarly. As can be seen from Fig.8 and Fig.9, even with very limited labeled data, the GLAM method can achieve higher accuracy.

V. CONCLUSION AND FUTURE WORK

In this paper, a method based on global information of spectral spacing and similarity matrix of local distance between samples and low rank subspace sparse representation(LRSR) model is proposed, called as GLAM-LRSR. This paper notices that graph similarity matrix plays a key role in subsequent clustering or semi-supervised classification tasks. Therefore, this paper proposes a global-local Affinity Matrix Model (GLAM) from the global and local perspectives of data. The innovations of the proposed algorithm are as follows: 1) In this paper, spectral spacing reflecting class information is used as a global constraint, and larger spectrum is obtained. Spacing resists noise to some extent and maintains the stability of the feature space of similarity matrix. Therefore, the similarity matrix of ideal diagonal block structure is approximated from the angle of maximizing spectral spacing. 2) Regarding the local distance between data as a regularization term not only prevents the failure of similarity matrix caused by too large spectral distance, but also maintains the local smoothness of data embedding manifold. The combination of similarity matrix constructed by GLAM and subspace partitioning (LRSR) improves the accuracy of subspace clustering significantly. At the same time, the similarity matrix constructed by GLAM-LRSR is successfully applied to graph-based semi-supervised classification tasks. The experimental results on real public datasets demonstrate the effectiveness of GLAM-LRSR algorithm. We observed that the proposed methods outperform significantly some other state-of-the-art methods. Our model

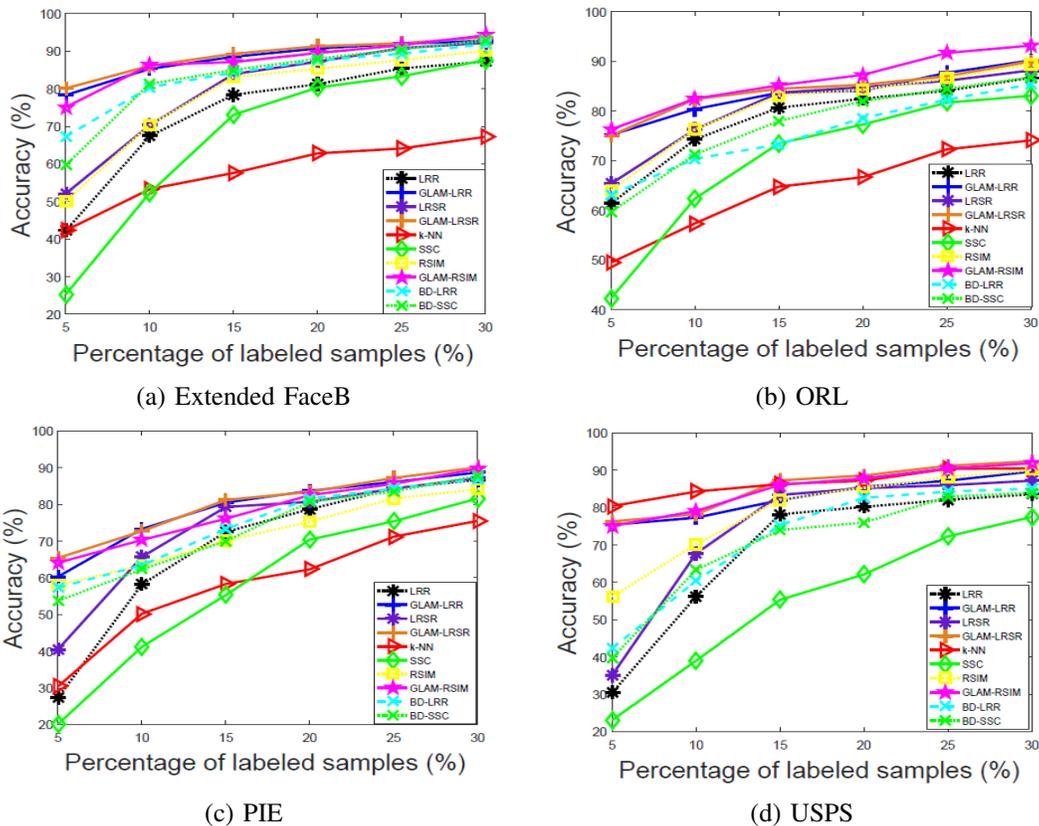


Fig. 8. Accuracy comparison on the transductive semi-supervised classification versus different percentages of labeled samples on four datasets

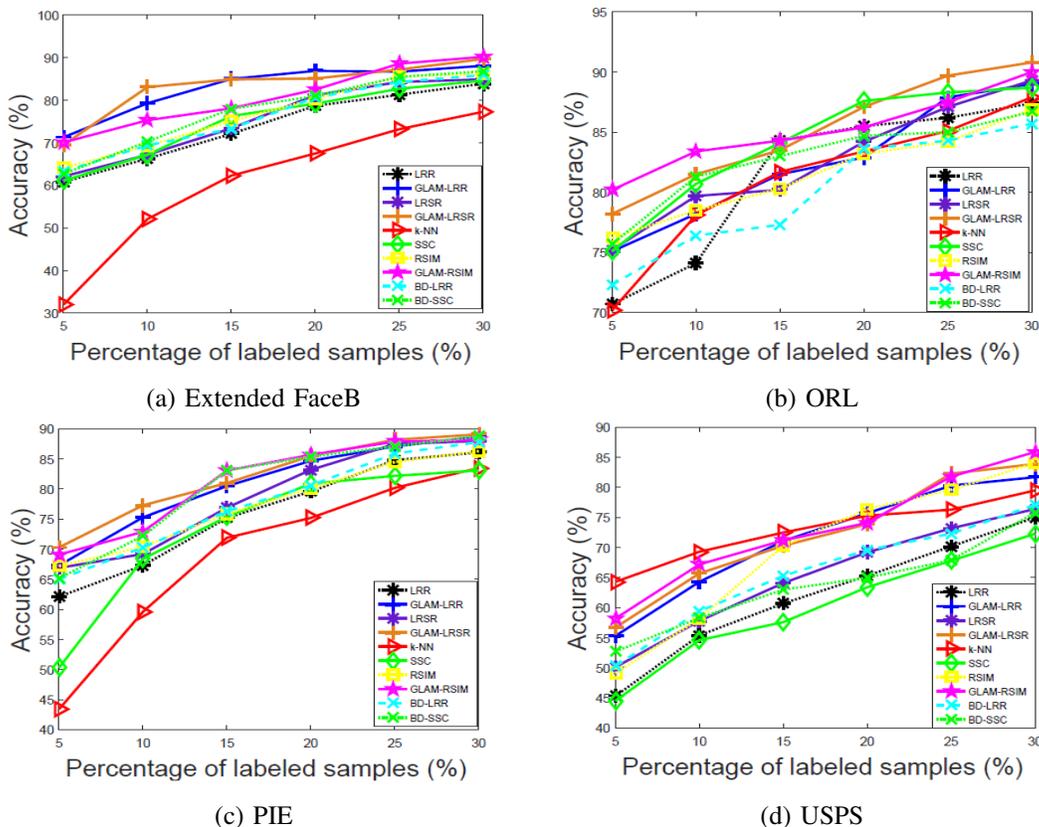


Fig. 9. Accuracy comparison on the inductive semi-supervised classification versus different percentages of labeled samples on four datasets

will be applied to supervised models in future work.

ACKNOWLEDGMENT

This research was supported by the National Natural Science Foundation of China (Grant No. 51677042)

All Author declares that he/she has no conflict of interest.

REFERENCES

- [1] K.-C. Lee, J. Ho, and D. J. Kriegman, "Acquiring linear subspaces for face recognition under variable lighting," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 27, no. 5, pp. 684–698, 2005.
- [2] R. Vidal, Y. Ma, and S. Sastry, "Generalized principal component analysis (GPCA)," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 27, no. 12, pp. 1945–1959, 2005.
- [3] D. Cheng, M. N. Nguyen, J. Gao, and D. Shi, "On the construction of the relevance vector machine based on bayesian ying-yang harmony learning," *Neural Networks the Official Journal of the International Neural Network Society*, vol. 48, no. 6, pp. 173–179, 2013.
- [4] M. Grundmann, V. Kwatra, M. Han, and I. Essa, "Efficient hierarchical graph-based video segmentation," in *Computer Vision and Pattern Recognition (CVPR), IEEE Conference on*. IEEE, 2010, pp. 2141–2148.
- [5] Y. Zhou, H. Cheng, and J. X. Yu, "Graph clustering based on structural/attribute similarities," *Proceedings of the Vldb Endowment*, vol. 2, no. 1, pp. 718–729, 2009.
- [6] G. Liu, Z. Lin, S. Yan, J. Sun, Y. Yu, and Y. Ma, "Robust recovery of subspace structures by low-rank representation," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 35, no. 1, pp. 171–184, 2013.
- [7] E. Elhamifar and R. Vidal, "Sparse subspace clustering: Algorithm, theory, and applications," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 35, no. 11, pp. 2765–2781, 2013.
- [8] L. Zappella, X. Llad, E. Provenzi, and J. Salvi, "Enhanced local subspace affinity for feature-based motion segmentation," *Pattern Recognition*, vol. 44, no. 2, pp. 454–470, 2011.
- [9] A. Goh and R. Vidal, "Segmenting motions of different types by unsupervised manifold clustering," in *IEEE Conference on Computer Vision and Pattern Recognition*, 2007, pp. 1–6.
- [10] C. R. Dong, D. S. Yeung, and X. Z. Wang, "An improved differential evolution and its application to determining feature weights in similarity based clustering," in *International Conference on Machine Learning and Cybernetics*, 2014, pp. 95–103.
- [11] E. L. Dyer, A. C. Sankaranarayanan, and R. G. Baraniuk, "Greedy feature selection for subspace clustering," *Journal of Machine Learning Research*, vol. 14, no. 1, pp. 2487–2517, 2013.
- [12] H. Zhang, Z. Lin, C. Zhang, and J. Gao, "Robust latent low rank representation for subspace clustering," *Neurocomputing*, vol. 145, no. 18, pp. 369–373, 2014.
- [13] X. Li, Y. Ma, and J. Wright, "Robust principal component analysis?" *Journal of the Acm*, vol. 58, no. 3, p. 11, 2011.
- [14] D. Cheng, J. Yang, J. Wang, D. Shi, and X. Liu, "Double-noise-dual-problem approach to the augmented lagrange multiplier method for robust principal component analysis," *Soft Computing*, vol. 21, pp. 1–10, 2015.
- [15] J. Feng, Z. Lin, H. Xu, and S. Yan, "Robust subspace segmentation with block-diagonal prior," in *Computer Vision and Pattern Recognition (CVPR), IEEE Conference on*. IEEE, 2014, pp. 3818–3825.
- [16] R. Liu, Z. Lin, and Z. Su, "Learning markov random walks for robust subspace clustering and estimation," *Neural Networks*, vol. 59, pp. 1–15, 2014.
- [17] C. Davis, "The rotation of eigenvectors by a perturbation," *Journal of Mathematical Analysis and Applications*, vol. 6, no. 2, pp. 159–173, 1963.
- [18] D. Shi, J. Wang, D. Cheng, and J. Gao, "A global-local affinity matrix model via eigengap for graph-based subspace clustering," *Pattern Recognition Letters*, vol. 89, pp. 67–72, 2017.
- [19] J. Wang, D. Shi, D. Cheng, Y. Zhang, and J. Gao, "Lrsr: Low-rank-sparse representation for subspace clustering," *Neurocomputing*, vol. 214, pp. 1026–1037, 2016.
- [20] P. Ji, M. Salzmann, and H. Li, "Shape interaction matrix revisited and robustified: Efficient subspace clustering with corrupted and incomplete data," in *International Conference on Computer Vision (ICCV), IEEE Conference on*, 2015, pp. 2141–2148.
- [21] F. Woolfe, E. Liberty, V. Rokhlin, and M. Tygert, "A fast randomized algorithm for the approximation of matrices," *Applied Computational Harmonic Analysis*, vol. 25, no. 3, pp. 335–366, 2008.
- [22] L. E. Peterson, "K-nearest neighbor," *Scholarpedia*, vol. 4, no. 2, p. 1883, 2009.
- [23] J. Shi and J. Malik, "Normalized cuts and image segmentation," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 22, no. 8, pp. 888–905, 2000.
- [24] X. Zhu, "Semi-supervised learning using gaussian fields and harmonic functions," *Proc Icml*, pp. 912–919, 2003.
- [25] D. Zhou, O. Bousquet, T. N. Lal, and J. Weston, "Learning with local and global consistency," in *International Conference on Neural Information Processing Systems*, 2003, pp. 321–328.
- [26] F. Wang and C. Zhang, "Label propagation through linear neighborhoods," *IEEE Transactions on Knowledge Data Engineering*, vol. 20, no. 1, pp. 55–67, 2007.
- [27] D. Cai, X. He, K. Zhou, J. Han, and H. Bao, "Locality sensitive discriminant analysis," vol. 2007, pp. 708–713, 2007.
- [28] P. N. Belhumeur, J. P. Hespanha, and D. J. Kriegman, "Eigenfaces vs. fisherfaces: Recognition using class specific linear projection," *IEEE Transactions on Pattern Analysis Machine Intelligence*, vol. 19, no. 7, pp. 711–720, 2002.
- [29] S. Li and Y. Fu, "Learning balanced and unbalanced graphs via low-rank coding," *IEEE Transactions on Knowledge Data Engineering*, vol. 27, no. 5, pp. 1274–1287, 2015.
- [30] F. S. Samaria and A. C. Harter, "Parameterisation of a stochastic model for human face identification," in *Applications of Computer Vision, 1994., Proceedings of the Second IEEE Workshop on*, 1994, pp. 138–142.
- [31] T. Sim, S. Baker, and M. Bsat, "The cmu pose, illumination, and expression (pie) database," in *IEEE International Conference on Automatic Face and Gesture Recognition, 2002. Proceedings*, 2002, pp. 46–51.
- [32] J. J. Hull, "A database for handwritten text recognition research," in *IEEE Trans. Pattern Anal. Mach. Intell.*, 1994, pp. 550–554.
- [33] C. Tomasi and T. Kanade, "Shape and motion from image streams under orthography: a factorization method," *International Journal of Computer Vision*, vol. 9, no. 2, pp. 137–154, 1992.