Multiview Similarity Learning for Robust Visual Clustering

Ao Li[0000−0003−0735−2917], Jiajia Chen[0000−0003−2705−1185], Deyun Chen[0000−0002−5176−7725], and Guanglu Sun[0000−0003−2589−1164]

School of Computer Science and Technology, Harbin University of Science and Technology, Harbin China.
dargonboy@126.com,544953065@qq.com,chendeyun@hrbust.edu.cn,sunguanglu@hrbust.edu.cn

Abstract. Multiview similarity learning aims to measure the neighbor relationship between each pair of samples, which has been widely used in data mining and presents encouraging performance on lots of applications. Nevertheless, the recent existing multiview similarity learning methods have two main drawbacks. On one hand, the comprehensive consensus similarity is learned based on previous fixed graphs learned from all views separately, which ignores the latent cues hidden in graphs from different views. On the other hand, when the data are contaminated with noise or outlier, the performance of existing methods will decline greatly because the original true data distribution is destroyed. To address the two problems, a Robust Multiview Similarity Learning(RMvSL) method is proposed in this paper. The contributions of RMvSL includes three aspects. Firstly, the recent low-rank representation shows some advantage in removing noise and outliers, which motivates us to introduce the data representation via low-rank constraint in order to generate clean reconstructed data for robust graph learning in each view. Secondly, a multiview scheme is established to learn the consensus similarity by dynamically learned graphs from all views. Meanwhile, the consensus similarity can be used to propagate the latent relationship information from other views to learn each view graph in turn. Finally, the above two processes are put into a unified objective function to optimize the data reconstruction, view graphs learning and consensus similarity graph learning alternatingly, which can help to obtain overall optimal solutions. Experimental results on several visual data clustering demonstrates that RMvSL outperforms the most existing methods on similarity learning and presents great robustness on noisy data.

1 Introduction

Similarity learning is a key and fundamental issue in data mining and machine learning, which not only is used to measure the data neighbor relationship, but also plays an important role in constructing the graph from data. Graph is a significant structure for presenting the relationships among large amount of objects, which is consist of nodes and their edges. Each node of graph corresponds to a object, and the edges represent the linkage among objects. By the

graph, the global and local data structure can be obtained simultaneously, which is widely used in many data analysis based applications. Thanks for the distinguished ability in relationship measurement, a good partition will be produced by the learned graph, which shows encouraging performance on clustering task. Therefore, graph-based clustering approaches attract increased attention and are widely studied in recent years. It is worth noting that similarity structure is at the core of graph-based clustering, so the quality of graph learning is crucial to final performance. In a direct way, the gaussian kernel can be used to construct a graph by computing exponential Euclidean distance between instances, such as k nearest neighbor graph. Nevertheless, the direct way would be failed because the data within similar distribution may suffer large gap in Euclidean space due to external outlier. To considered structural similarity on distribution, a l_1-graph learning method is proposed in [1] with the help of data representation model, where each weight edge of graph is learned by the l_1-norm based reconstruction from instances. Motivated by l_1-graph, Fang et al [2] seek the non-negative low-rank representation from data for subspace clustering, which can be used to explore the more meaningful similarity relationship by exploiting subspace structure. From the viewpoint of probability, Nie et al [3] propose a probabilistic graph learning approach based on adaptive neighbors. In their method, all the instances are connected to each other as a neighbor with a probability s_{ij} , which should be large if the distance between instance x_i and x_j is small while verse vice. So, the obtained s_{ij} can be utilized as a well similarity measurement. To earn extra robustness, Kang et al [4] introduces the robust principle component analysis(RPCA) to probabilistic graph, where the similarity probability is assigned based on latent data component.

Nevertheless, the above graph learning methods are designed in a single view. More recently, multiview learning breeds a new paradigm in machine learning, which aims to establish the learning model from different views. In essence, multiview data often presents in our life. For example, an image can be characterized by different types of features, a piece of news can be described by different languages, et al. Hence, thanks for the ability in utilizing the redundant and complementary information across views, multiview learning has been widely studied in many literatures [5–11]. Specially, many multiview graph learning approaches are proposed and applied to clustering problem. Multiview graph learning aims to seek a more accurate fusion similarity graph from the cues in all views. And then, the corresponding clustering algorithm can be implemented on the learned fusion graph to obtain low-dimensional data representation for final cluster results. In [12], a multiple graph learning method is presented to share a same cluster indicator matrix. Also, a auto-weighted scheme is adopt to assign the optimal parameter-free weight for each view graph. Nie et al [13] supposes that graphs from all views are closed to each other, and they would like to seek the centroid of graphs by exploring the Laplacian rank constraint as the fusion graph. Meanwhile, the self-weighted scenario with hyperparameter is deisgned to compute the confidences for different graphs. Inspired by it, to further make full use of data correlation among views, Zhan et al [14] proposes a concept

Factorization-based multivew clustering. In this approach, affinity weights from different views are correlated to jointly learn the unified graph matrix. Assuming that the underlying cluster structure is shared across multiple views, a common graph is learned by minimizing the disagreement between each pair of views in [15]. Moreover, the rank constraints are also imposed on each Laplacian matrix for further improving the graph consensus. Similarly, a multi-graph fusion scheme is proposed in [16] for multiview clustering, which enforces the fusion graph to be approximated to original graph from each view but with an explicit cluster structure. Although these multiview graph learning methods have shown promising results and proved to be effective in clustering, they are still limited in two aspects. Firstly, the noise and outlier influence are ignored when individual graph is learned from each view, by which the true similarity structure in graph might be damaged. Secondly, the graphs from each view are isolated pre-learned and fixed when multiview learning is implemented, which leads that the complementary information cross views is not considered sufficiently. Finally, the robust modeling, individual view graph learning and common graph learning are not be well unified to improve each other. Hence, the correlations among these three aspects are not utilized efficiently.

Towards the mentioned problems, in this paper, we proposes a robust multiview similarity learning method for visual data clustering, which can be used in many practical applications, such as image annotation, visual pattern analysis and so on. Thanks for the help of robust representation and multiview learning, RMvSL can learn an effective graph with reliable similarity relationship from robust multiple graphs. The contributions of RMvSL are summaried as follows.

1) Data representation model is introduced to seek a robust space via low-rank constraint, by which the corrupted data can be recovered with linear combination of instances. That is, what we want is to learn the view graphs from the robust compensated counterpart of corrupted data.

2) A consensus similarity graph learning scheme will be designed in a multivew way to cover multiple cues from all views, in which the latent true similarity can be well explored in the learned consensus graph. Meanwhile, the complementary information can be propagated among views by common graph when the individual view graph is constructed.

3) We put the data compensation model, view graphs and consensus graph learning into a unified framework where variables can be jointly optimized to benefit each other and obtain the overall optimal solutions with a developed numerical algorithm.

The comprehensive idea of our proposed approach is shown in Fig.1. The joint learning framework is presented, in which the iterations are conducted among the data reconstruction, view graph and consensus graph learning. And then, the spectral clustering will be implemented on the learned consensus graph to get the final cluster structures.

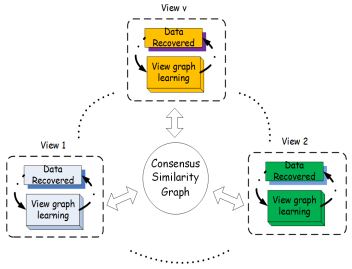


Fig. 1. The joint learning framework of RMvSL.

2 Related works

2.1 Graph learning revisited

As illustrated in [3] , learning similarity with adaptive neighbors is a kind of easy and elegant graph constructed method. In this branch, it emphasizes that the connectivity of neighbor is a useful strategy for similarity learning. Inspired by it, the neighbors of instance can be connected to each other with a probability. Assuming s_{ij} denotes the probabilistic similarity between two instance x_i and x_j respectively, the similarity with adaptive neighbors can be formulated as follows.

$$\min_{s_i^T \mathbf{1}=1, 0 \le s_{ij} \le 1} \sum_{j=1}^{n} \|x_i - x_j\|_2^2 s_{ij} \tag{1}$$

where $\|\cdot\|_2$ denotes the l_2-norm. The Eq.(1) is established with an intuitive assumptions that samples closed to each other should have a large connected probability s_{ij} . In other words, the s_{ij} has a negative correlation to the distance of each pair of instances. In addition, to avoid the trivial solution, an additional l_2-norm based constraint is imposed on the similarity matrix in Eq.(1) as follows.

$$\begin{gathered} \min_{s_i \in \mathbb{R}^{n \times 1}} \sum_{i,j}^{n} Tr(XLX^T) + \lambda \|S\|_F^2 \\ s.t. \forall i, s_i^T \mathbf{1} = 1, 0 \le s_{ij} \le 1 \end{gathered} \tag{2}$$

where $L = D - S$ denotes the graph Laplacian matrix, and D is a diagonal matrix with $D_{ii} = 0.5\left(\sum S_{i*} + \sum S_{*i}\right)$

By Eq.(2), the larger probability will be assigned to the closed instances pair, and the obtained S can be used as the similarity matrix to present the neighbor relationship among data. Next, the spectral clustering can be implemented on S to obtain the final cluster result. Moreover, many other applications based on the above graph learning model are further studied in literatures[17–19].

Nevertheless, from Eq.(2), we can see that the similarity probability is learned from raw data directly. It means that the true similarity relationship may be damaged when the data are contaminated with noise or outlier, which will destroy the latent distribution property belongs to clean data. Meanwhile, in practice, it is easy to collect more noisy data while few clean data in the open

visual environment. So, a more robust learning method is urgently expected to obtain high quality graph with noisy data recently.

2.2 Self-representation via low-rank constraint

Low-rank representation(LRR) is a typical representation model to learn subspace structure hidden among data, which attracts much attention and has shown promising performance in data recovery problem [20]. In general, the LRR can be formulated as

$$\begin{aligned} &\min_{Z} \|Z\|_* + \tau\|E\|_l \\ &s.t. X = DZ + E \end{aligned} \tag{3}$$

where $X \in R^{d\times n}$ denotes the data matrix including n instances with d dimensions, D is the dictionary used to span the space for X, and Z is the representation coefficient matrix. E denotes the residual matrix for modeling kinds of typical noise with different norm, such as Frobenius norm for Gaussian noise, l_1 norm for random corruptions, et al. $\|\cdot\|_*$ and $\|\cdot\|_l$ indicate the nuclear norm and l norm respectively.

Specially, if the dictionary is replaced by data matrix X itself, the above formulation are changed into a self-representation based LRR problem as follows.

$$\begin{aligned} &\min_{Z} \|Z\|_* + \tau\|E\|_l \\ &s.t. X = XZ + E \end{aligned} \tag{4}$$

With the Eq.(4), the intrinsic subspace structure of data is uncovered, and the robust recovered data may be reconstructed from two aspects. On one hand, thank for Z, a nearly clean counterpart of raw data are recovered by the linear combination of all the data as XZ, and much useful detail can be reconstructed by the instances mainly drawn from the same subspace due to the block diagonal character of Z. Meanwhile, the damaged latent distribution hidden in data would be complementary with the help of combination of other instances. On the other hand, the error matrix E can also remove extra noise influence existed in the noisy raw data. So, if the LRR is introduced into the neighbor graph learning model to form a unified framework that learns the recovered data and similarity probability alternately, a more robustness similarity graph will be constructed to improve the subsequent clustering results.

3 The proposed framework of RMvSL

Multiview graph learning aims to learn a fusion graph from the different views, which can cover all the useful cues from all views and obtain further performance improvement. For the most existing multiview graph learning approaches, they generally focus on the elegant graphs construction model and their corresponding multiview learning mechanism but ignore the data themselves, especially for uncertain noisy data. In the open environment, the data are easily contaminated with kinds of noise, which will damage the original data

distribution greatly and lead to awful similarity measurement obviously. However, the data quality restoration are not sufficiently and especially considered in the most existing multiview graph learning approaches.

Motivated by the above considerations, the details of our proposed RMvSL will be discussed in this section. Firstly, we will talk about how to learn the similarity probability with the cleaning recovered data to earn a reliable graph matrix from the noisy raw data in each view. And then, a multiview scheme is designed to not only learn the consensus fusion graph but also propagate the cues from other views for each single view graph learning.

For each view v, the robust graph learning method is formulated as

$$\begin{aligned} &\min_{Z^v,E^v,S^v} \|Z^v\|_* + \alpha\|E^v\|_l + \beta\sum_{i,j}\|X^vZ_i^v - X^vZ_j^v\|_2^2 S_{ij}^v + \gamma\|S^v\|_F^2 \\ &s.t. X^v = X^vZ^v + E^v, S^v\mathbf{1} = \mathbf{1}, 0 \le S_{ij}^v \le 1 \end{aligned} \tag{5}$$

where $X^v \in R^{d\times n}$ denotes the data in v-th view, $Z^v \in R^{n\times n}$ and Z_i^v present the low-rank representation matrix and its i-th column. α, β and γ are the positive parameters to balance the constrained terms. By our proposed framework in Eq.(5), the view graph $S^v \in R^{n\times n}$ is learned from the compensated data $X^vZ_i^v$ with LRR model instead of the noisy raw data X_i^v, which can further preserve true data distribution when the similarity probability S_{ij}^v is assigned. In turn, the learned similarity graph S_{ij}^v can also be used to guide representation matrix(Z^v) learning, which helps to obtain more rational recovered data. The two mutual parts will benefit each other during the iterations. Furthermore, different from the direct two-stage operation that separates the data recovery and graph learning, we put them into a unified framework. It should be noted that graph presents the neighbor relationship between pair of instances, which is believed to be capable of providing useful correlation information when the subspace representation is learned. Hence, the unified framework will improve the two variables each other effectively. That is, in this framework, the more accurate subspace structure will be learned for better data recovering, and it helps to obtain a higher quality view graph with nearly clean compensated data in turn.

As mentioned above, to fuse cues from other views, a consensus graph learning scheme is designed in a multivew manner, which makes the view graphs and their consensus fusion version to be mutually optimized by jointly learning. Assuming G is the consensus fusion graph, which can be solved with the following weighted learning model as

$$\begin{aligned} &\min_{G} \sum_{v=1}^{m} w_v\|G - S^v\|_F^2 \\ &s.t. \forall_i, g_{ij} \ge 0, \mathbf{1}^T\mathbf{g}_i = 1 \end{aligned} \tag{6}$$

where m denotes the total number of views, and g_i presents the i-th row of G. Similarly, to keep the probability property, the nonnegative and normalization constraints are enforced on G in Eq.(6). w_v is the weighted constraint coefficient for the v-th view, which is computed automatically in our designed multiview

graph learning scheme. Naturally, a compact form is set to be the reciprocal of distance between pair of graphs as $w_v = \frac{1}{2\sqrt{\|G-S^v\|_F^2}}$. In practice, the weighted coefficient for each view is determined during the iteration dynamically to present stronger adaption than the fixed situation.

Incorporating Eq.(6) into Eq.(5), our proposed RMvSL model is turned into

$$\begin{aligned}
&\min_{Z,E,S,G} \sum_{v=1}^{m} \Bigg(\|Z^v\|_* + \alpha\|E^v\|_l + \beta \sum_{i,j} \|X^v Z_i^v - X^v Z_j^v\|_2^2 S_{ij}^v \\
&+ \gamma\|S^v\|_F^2 + w_v\|G - S^v\|_F^2 \Bigg) \\
&s.t. X^v = X^v Z^v + E^v, S^v\mathbf{1} = \mathbf{1}, 0 \le S_{ij}^v \le 1, v = 1, ..., m, \\
&\forall i, g_{ij} \ge 0, \mathbf{1}^T\mathbf{g}_i = 1
\end{aligned} \tag{7}$$

Based on the comprehensive formulation in RMvSL, the view subspace structure Z^v, view graph S^v and consensus graph G are jointly learned in a unified framework, resulting in the alternating optimizing among the recovered data and graph learning to improve each variable. That is to say, by Eq.(7), we establish a multiview graph learning framework with dynamic data compensation mechanism, which could mutually learn the recovered data, view graphs and fusion graph and show some advantage on similarity learning for uncertain noisy data. Furthermore, with the proposed unified framework, each variable is not optimized in isolation during iteration. For example, when the subspace structure Z^v is updated, the view graph S^v is involved to provide relationships between data, which is benefit to intrinsic structure exploration and data reconstruction. Similarly, for each view graph leaning, the cues from all other views are propagated by the fusion graph G. So, the overall optimal solutions can be obtained by our proposed framework, which is believed to learn a more robust and high quality graph finally. In the next section, a numerical algorithm is developed to solve the objective function in RMvSL efficiently.

3.1 Numerical algorithm

Solving all the variables at once is a challenging problem because they are coupled in objective function. So, we adopt an alternating scheme to optimize variables iteratively. That is, we updates only one variable once while fixing others.

Eq.(7) can be solved by using the inexact augmented Lagrange multiplier method (IALM). Introducing the auxiliary variable J , it can be changed into:

$$\begin{aligned}
&\min_{Z,E,S,G} \sum_{v=1}^{m} \Bigg(\|Z^v\|_* + \alpha\|E^v\|_l + \beta Tr(X^v Z^v L^v (X^v Z^v)^T) \\
&+ \gamma\|S^v\|_F^2 + w_v\|G - S^v\|_F^2 \Bigg) \\
&s.t. X^v = X^v Z^v + E^v, S^v\mathbf{1} = \mathbf{1}, 0 \le S_{ij}^v \le 1, v = 1, ..., m, \\
&\forall i, g_{ij} \ge 0, \mathbf{1}^T\mathbf{g}_i = 1, Z^v = J^v.
\end{aligned} \tag{8}$$

Removing the equality constraints on $X^v$and Z^v, we can get the following Lagrangian functions:

$$\begin{aligned}\min_{Z,E,S,J,G} \sum_{v=1}^{m} \Big(&\|J^v\|_* + \alpha\|E^v\|_l + \beta Tr(X^v Z^v L^v (X^v Z^v)^T) + \gamma\|S^v\|_F^2 \\ &+ w_v\|G - S^v\|_F^2 + \|X^v - X^v Z^v - E^v - \frac{R_1^v}{\mu}\|_F^2 + \|Z^v - J^v - \frac{R_2^v}{\mu}\|_F^2 \Big)\end{aligned} \tag{9}$$

where μ is penalty parameter. R_1 and R_2 are Lagrange multipliers. Next,the variables in (9) will be solved one by one.

Updating E^v for each view:

$$\min_{E^v} \alpha\|E^v\|_l + \|X^v - X^v Z^v - E^v - \frac{R_1^v}{\mu}\|_F^2 \tag{10}$$

For different types of noise, the error matrix can be solved with the corresponding l-norm based optimization. For example, if l_1 norm is enforced, then $E_{ij} = (|t_{ij}| - \alpha/\mu)^+ \cdot sign(t_{ij})$, where $T = X^v - X^v Z^v - \frac{R_1^v}{\mu}$.

Updating Z^v for each view:

$$\min_{Z} \|Z^v - J^v - \frac{R_2}{\mu}\|_F^2 + \beta Tr(X^v Z^v L^v (X^v Z^v)^T) + \|X^v - X^v Z^v - E^v - \frac{R_1^v}{\mu}\|_F^2 \tag{11}$$

Setting its derivative to be zero, we have

$$2\beta Z^v L^v + ((X^v)^T X^v)^{-1}(2I + 2(X^v)^T X^v) Z^v = ((X^v)^T X^v)^{-1} D \tag{12}$$

where $D = 2J^v + 2\frac{R_1}{\mu} + 2(X^v)^T X^v - 2(X^v)^T E^v - 2(X^v)^T \frac{R_2}{\mu}$, and I denotes the identity matrix. Eq.(12) is a standard Sylvester equation, which can be solved by existing method[21].

Updating J^v for each view:

$$\min_{J} \|J^v\|_* + \|Z^v - J^v - \frac{R_2^v}{\mu}\|_F^2 \tag{13}$$

Eq.(13) is a nuclear norm based minimization problem, which can be easily solved by using the singular value shrinkage operator proposed in [22].

Updating graph S^v for each view: To make it more clear, the S^v is solved in a row-wise way by

$$\min_{s_i} \sum_{j=1}^{n} \frac{\beta}{2}\|o_i - o_j\|^2 s_{ij} + \gamma s_{ij}^2 + (g_{ij} - s_{ij})^2 \tag{14}$$

where $O = X^v Z^v$ is the recovered data by linear combination of raw data, and o_i denotes its i-th column. Assuming $f_{ij} = \|o_i - o_j\|^2 - \frac{4}{\beta} g_{ij}$, Eq.(14) can be converted to

$$\min_{s_i^T \mathbf{1} = 1, 0 \le s_{ij} \le 1} \|s_i + \frac{\beta}{4\gamma} f_i\|^2 \tag{15}$$

Eq.(15) have a natural sparse solution due to the neighborhood connections. Therefore, we only need to update its first k neighborhoods. In other words, s_i

has k positive entries. The Lagrangian function of Eq.(15) can be expressed as follows:

$$\Gamma(s_i, \eta, \zeta) = \|s_i + \frac{\beta}{4\gamma_i} f_i\|^2 - \eta(s_i^T \mathbf{1} - 1) - \zeta_i^T s_i \tag{16}$$

where η and $\zeta \in R^{n\times 1}$ are Lagrangian multipliers. By the Karush-Kuhn-Tucker condition, it produces $s_i = ((\eta/2) - (\beta f_i/4\eta))_+$.

Ranking f_i in ascending order, we have

$$\begin{cases} s_{ik} = \frac{\eta}{2} - \frac{\beta f_{ik}}{4\gamma_i} > 0 \\ s_{i,k+1} = \frac{\eta}{2} - \frac{\beta f_{i,k+1}}{4\gamma_i} \in 0 \\ s_i^T \mathbf{1} = \sum_{j=1}^{k} (\frac{\eta}{2} - \frac{\beta f_{ij}}{4\gamma_i}) = 1 \end{cases} \tag{17}$$

After infering, we can get

$$\begin{cases} s_{ij} = \frac{f_{i,k+1} - f_{ij}}{k f_{i,k+1} - \sum_{r=1}^{k} f_{ir}}, j \leq k \\ \gamma_i = \frac{\beta}{4}(k f_{i,k+1} - \sum_{j=1}^{k} f_{ij}) \\ \eta = \frac{2}{k} + \frac{\beta}{2k\gamma_i} \sum_{j=1}^{k} f_{ij} \end{cases} \tag{18}$$

Updating consensus similarity graph G: When the above variables of all views are solved, the consensus fusion graph G can be optimized by

$$\begin{aligned} &\min_{G} \sum_{v=1}^{m} w_v \|G - S^v\|_F^2 \\ &s.t. g_i^T \mathbf{1} = 1, g_{ij} > 0 \end{aligned} \tag{19}$$

The solution can be obtained by solving its each row separately as

$$\min_{g_i^T \mathbf{1}=1, g_{ij}>0} \sum_{v=1}^{m} \|w_v(g_i - s_i^v)\|_F^2 \tag{20}$$

Eq.(20) can be solved by an effective iterative scheme proposed in [23].

In summary, the developed scheme for solving our proposed objective function of **RMvSL** in Eq.(7) is listed in **Algorithm 1** as follows.

Algorithm 1

Input: Multiview Data $\{X^v\}_{v=1}^m$, Parameters α, β, γ, μ;

Initialize: $E^v = 0$, $Z^v = \left(X^{vT} X^v + 10^{-3} I\right)^{-1} X^{vT} X^v$, $R_1^v = R_2^v = 0$; S^v is initialized with normalized Euclidean distance;

while not converged do

for each view do

1. Update J^v using Eq.(14);
2. Update Z^v according to Eq.(13);
3. Update S^v using Eq.(18);
4. Update E^v by Eq.(11);
5. $R_1^v = R_1^v + \mu(X^v Z^v + E^v - X^v)$
6. $R_2^v = R_2^v + \mu(J^v - Z^v)$

end for

7. Update G according to Eq.(20), and $w_v = \frac{1}{2\sqrt{\|G - S^v\|_F^2}}$

until convergence

Output: Similarity graph G

4 Experiments

In this part, we will verify the performance of our RMvSL model via visual clustering experiments on four datasets, including two facial datasets, a digital handwritten dataset and a object dataset. The detailed description and setting for each dataset will be elaborated later. Our proposed RMvSL is compared with six recent simliarity graph learning approaches, including two single-view based approaches and four multiview based approaches: classical k-nearest neighbor Graph Construction with Gaussian distance(GCG), Robust Graph Clustering(RGC) [4], Graph-based Multi-view Clustering(GMC) [24], Multi-Graph Fusion for Multi-view Spectral Clustering(GFSC) [16], parameter-free Auto-weighted Multiple Graph Learning(AMGL) [12], Self-weighted Multiview Clustering with Multiple Graphs(SwMC) [13], Multiview concept factorization(MVCF) [14] and Multiview consensus graph clustering(MCGC)[15]. Among them, RGC and GCG are single-view based methods that are implemented on each view and the average indicator of all views are taken as the final clustering performance. To be balanced, the same spectral clustering are implemented on the learned graphs from comparison methods, and the obtained clustering indicators are used to evaluate their performance. Three classical clustering evaluation indicators(accuracy(ACC), normalized mutual information(NMI) and purity(PUR)) are used in our experiments.

4.1 Datasets

ORL/YALE dataset: The **ORL** contains 400 facial images from 40 different people. **YALE** contains total 165 facial images from 15 people, each of whom has 11 photos under different lighting, posture and expression. We extract three kinds of features for these datasets that are gray intensity, Local Binary Pattern(LBP)[25] and Gabor feature [26] as three views for each subject.

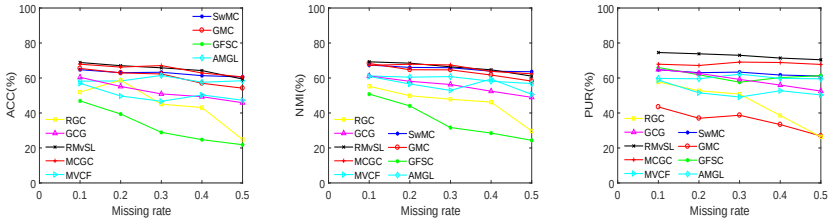
COIL20 dataset: COIL-20 is a collection of gray-scale images, including 20 objects taken from different angles. The image is taken every 5 degrees and size of 32×32, and each object has 72 images, a total of 1440 images. We directly divide them into four different views according to the taken degree: V1$[0°, 85°]$, V2$[90°, 175°]$, V3$[180°, 265°]$, V4$[270°, 360°]$. In essence, by this setting, each object has 18 individuals, meanwhile each individual has 4 views from different taken degree interval.

UCI-handwritten(UCI-H) dataset: UCI-H contains ten kinds of handwritten images, which are 0,1,2,..., 9. Each subject has 200 samples, so the entire dataset has 2000 samples and 10 clusters. We select 500 samples and use three views for each instance. The first view is the profile-correlation feature with 216 dimensions, the second is the Fourier-coefficient with 76 dimensions, and the third is morphological feature with 6 dimensions.

4.2 Experimental Results and Analysis

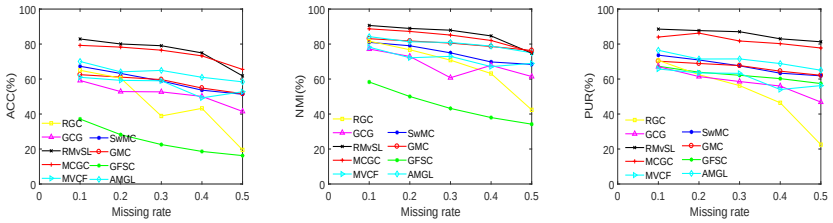
In our experiments, there are two types of representative noise will be introduced to valid the robust performance of comparison methods under noisy

visual data, which are Gaussion noise and feature missing respectively. The test on each dataset are repeated five times and the average indicator performance are reported. We add the Gaussion noise to normalized view feature with different variances with interval 0.01, and take l_2 norm as the constraint on error matrix E. The average clustering results and their variances in parenthesis are shown in Table.1-4. The best results are highlighted in boldface. Specially, only the variance larger than 10^{-3} is presented while others are shown as $\rightarrow$0. For feature missing, we randomly select part of the features of data in each view to be zero. Meanwhile, the l_1 norm is taken to be enforced on E. The average clustering results under different missing rates are plotted in Fig.2-5.



(a) ACC (b) NMI (c) PUR

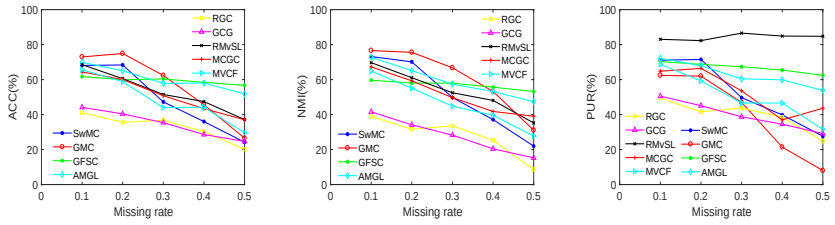
Fig. 2. Clustering results on **YALE** with missing rate.



(a) ACC (b) NMI (c) PUR

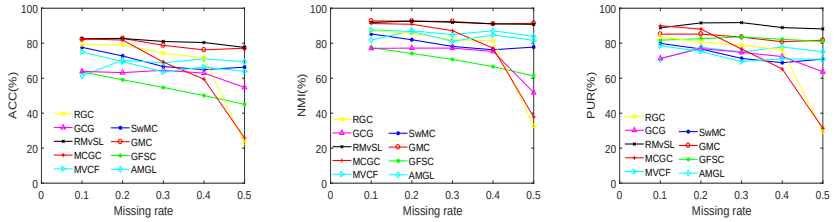
Fig. 3. Clustering results on **ORL** with missing rates.

From the results in tables, we can see that, when the data are contaminated with Gaussion noise, our proposed method outperforms other compared methods on visual clustering task in almost all experiments, which demonstrates the robust performance of RMvSL. In particular, thank for the introduced low-rank representation model, the view graph can be learned from the clean reconstructed data, which will help to obtain a better similarity metric with nearly true data distribution.



(a) ACC (b) NMI (c) PUR

Fig. 4. Clustering results on **UCI-H** with missing rates.



(a) ACC (b) NMI (c) PUR

Fig. 5. Clustering results on **COIL20** with missing rates.

Table 1. Clustering results on **UCI-H**(%).

	Metric	SwMC	GMC	GFSC	AMGL	RGC	GCG	MCGC	MVCF	RMvSL
σ = 0.02	ACC	70.56(→0)	**74.24**(→0)	39.16(→0)	72.88(0.008)	35.66(→0)	40.27(→0)	72.36(0.005)	16.64(→0)	73.12(0.004)
	NMI	75.24(→0)	**76.10**(→0)	42.98(→0)	75.63(0.003)	33.78(→0)	36.31(→0)	70.02(0.003)	3.65(→0)	75.85(0.003)
	PUR	75.08(→0)	60.48(→0)	73.22(→0)	75.64(0.006)	36.33(→0)	51.20(→0)	72.52(0.004)	16.52(→0)	**88.68**(→0)
σ=0.03	ACC	69.96(0.005)	**73.80**(0.002)	38.75(0.0004)	69.56(0.005)	34.49(→0)	44.60(0.0008)	73.12(→0)	15.88(→0)	72.84(0.001)
	NMI	72.83(0.001)	**76.19**(0.001)	40.21(0.0004)	73.27(0.003)	34.40(→0)	38.66(0.0007)	70.48(→0)	3.49(→0)	72.95(0.002)
	PUR	74.16(0.002)	58.94(0.006)	75.16(→0)	73.16(0.003)	40.13(→0)	50.07(→0)	73.6(→0)	16.52(→0)	**86.80**(→0)
σ=0.04	ACC	69.36(0.003)	72.51(→0)	32.80(0.002)	**74.92**(0.002)	36.67(→0)	37.34(0.003)	67.48(0.002)	17.0(→0)	73.24(0.001)
	NMI	70.69(0.001)	**74.35**(→0)	33.32(0.002)	72.09(0.002)	32.85(→0)	30.89(→0)	63.07(0.001)	4.66(→0)	73.92(0.005)
	PUR	71.96(0.001)	57.61(0.002)	81.20(0.001)	74.92(0.004)	51.34(→0)	43.87(→0)	67.76(0.001)	18.2(→0)	**87.28**(→0)
σ=0.05	ACC	67.88(→0)	**69.20**(0.006)	31.66(0.001)	66.28(0.004)	32.66(→0)	39.87(→0)	66.16(0.007)	17.8(→0)	68.16(0.003)
	NMI	**67.96**(→0)	70.35(0.004)	30.72(0.001)	65.76(0.002)	30.10(→0)	30.13(→0)	62.67(0.003)	5.22(→0)	67.27(0.004)
	PUR	71.44(→0)	45.36(0.012)	80.55(0.002)	67.84(0.003)	38.26(→0)	47.40(→0)	67.52(0.007)	19.0(→0)	**88.88**(→0)
σ=0.06	ACC	60.56(0.003)	58.72(0.005)	29.33(0.001)	**61.25**(0.003)	32.34(→0)	33.60(→0)	62.72(0.003)	15.4(→0)	60.94(0.002)
	NMI	**61.16**(0.002)	60.06(0.001)	27.60(0.001)	58.69(0.001)	27.37(→0)	25.17(→0)	56.27(0.002)	3.58(→0)	58.81(0.002)
	PUR	65.08(0.003)	30.42(0.007)	82.51(0.002)	64.92(0.002)	41.00(→0)	40.93(→0)	63.04(0.003)	15.8(→0)	**89.80**(0.001)

For experiments on feature missing, RMvSL achieves almost best performance among comparison methods on all evaluation indicators in Yale, ORL and COIL20. Only for UCI-H dataset, the GMC and AMGL are a little better than our proposed method on ACC and NMI under a few of missing rates. Nevertheless, the competitive results can still show advantage of RMvSL on robustness under feature missing situation.

4.3 Parameters sensitivity and convergence

There are total five parameters α, β, γ, and μ in our proposed method. We set $\alpha = 1/\sqrt{(max\,(d,n))}$ for all the experiments according to [22]. As for γ, it can

Table 2. Clustering results on **YALE**(%).

	Metric	SwMC	GMC	GFSC	AMGL	RGC	GCG	MCGC	MVCF	RMvSL
σ = 0.02	ACC	52.97(0.001)	54.30(0.001)	47.96(→0)	55.03(0.001)	48.28(→0)	41.82(→0)	**60.36**(→0)	41.82(0.006)	58.30(0.003)
	NMI	56.94(0.002)	58.88(→0)	52.10(→0)	55.87(0.001)	50.66(→0)	49.96(→0)	**59.85**(→0)	45.96(0.007)	59.52(0.001)
	PUR	54.67(0.001)	29.21(0.003)	64.62(→0)	56.85(0.002)	51.92(→0)	48.69(→0)	64.48(→0)	44.85(0.005)	**70.55**(→0)
σ=0.03	ACC	43.15(0.002)	32.85(0.006)	43.67(→0)	44.91(→0)	33.73(→0)	29.29(→0)	**46.12**(0.002)	33.94(→0)	44.91(0.001)
	NMI	45.13(0.002)	42.43(0.001)	44.75(→0)	**47.78**(0.003)	33.73(→0)	32.60(→0)	47.26(0.003)	41.25(0.001)	43.47(0.001)
	PUR	44.00(0.002)	7.67(→0)	63.90(→0)	49.21(0.003)	37.57(→0)	31.92(→0)	49.33(0.002)	35.15(→0)	**67.52**(→0)
σ=0.04	ACC	25.58(→0)	25.82(0.001)	27.07(→0)	28.79(→0)	23.23(→0)	24.24(→0)	31.87(0.001)	27.27(0.001)	**28.94**(→0)
	NMI	28.59(0.001)	30.75(0.001)	28.26(→0)	**30.34**(→0)	27.59(0.009)	26.57(0.009)	31.61(0.002)	31.31(0.001)	29.33(0.02)
	PUR	27.88(→0)	3.52(→0)	64.05(→0)	32.97(→0)	26.46(→0)	25.66(→0)	62.90(0.002)	29.70(0.001)	**75.39**(0.002)
σ=0.05	ACC	21.24(→0)	19.06(→0)	16.10(→0)	21.15(→0)	21.41(→0)	18.61(→0)	20.96(0.001)	22.42(→0)	**21.32**(→0)
	NMI	23.28(→0)	23.11(→0)	21.60(→0)	**25.89**(→0)	25.83(→0)	25.59(→0)	20.16(→0)	26.71(→0)	23.46(→0)
	PUR	24.48(→0)	0.72(→0)	64.40(→0)	28.85(→0)	23.03(→0)	25.86(→0)	22.66(0.001)	24.45(→0)	**77.33**(→0)
σ=0.06	ACC	17.82(→0)	18.42(→0)	15.89(→0)	**21.30**(→0)	19.62(→0)	20.61(→0)	17.33(→0)	18.18(→0)	20.84(→0)
	NMI	18.1(→0)	**24.71**(→0)	21.37(→0)	22.57(→0)	21.20(→0)	20.43(→0)	17.45(→0)	18.72(→0)	22.86(→0)
	PUR	18.91(→0)	0.36(→0)	63.76(→0)	23.76(→0)	23.84(→0)	22.42(→0)	18.66(→0)	20.00(→0)	**79.52**(→0)

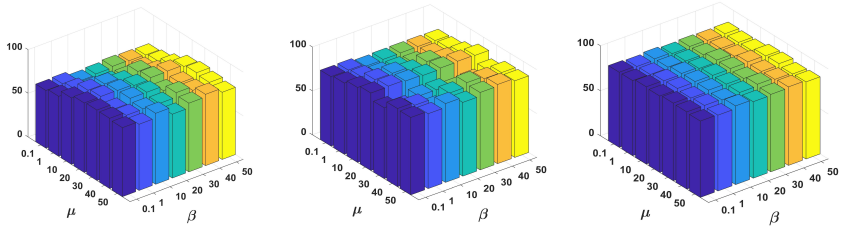
Table 3. Clustering results on **ORL**(%).

	Metric	SwMC	GMC	GFSC	AMGL	RGC	GCG	MCGC	MVCF	RMvSL
σ = 0.02	ACC	51.55(→0)	49.10(→0)	41.17(→0)	59.50(→0)	43.58(→0)	37.50(→0)	70.00(→0)	20.00(0.03)	**72.65**(0.002)
	NMI	65.76(→0)	73.88(→0)	61.58(→0)	76.96(→0)	61.64(→0)	43.33(→0)	80.50(→0)	37.56(0.03)	**83.37**(0.007)
	PUR	60.60(→0)	59.30(→0)	68.47(→0)	65.95(→0)	46.67(→0)	44.08(→0)	75.85(→0)	21.5(0.04)	**82.05**(0.001)
σ=0.03	ACC	33.40(→0)	36.25(→0)	38.54(→0)	49.15(0.002)	31.25(→0)	26.25(→0)	46.70(0.001)	47.00(→0)	**50.80**(→0)
	NMI	41.91(0.001)	60.89(→0)	59.68(→0)	**64.69**(0.001)	51.59(→0)	47.29(→0)	56.78(0.002)	64.67(→0)	62.16(→0)
	PUR	40.15(→0)	45.70(→0)	68.34(→0)	54.85(0.001)	34.08(→0)	30.83(→0)	52.60(0.002)	52.50(→0)	**75.05**(→0)
σ=0.04	ACC	19.05(→0)	18.70(→0)	18.51(→0)	**28.50**(→0)	17.92(→0)	20.00(→0)	22.00(→0)	32.00(0.01)	28.15(→0)
	NMI	23.71(→0)	37.42(→0)	33.16(→0)	**42.92**(→0)	36.47(→0)	37.98(→0)	29.84(→0)	48.54(→0)	40.79(→0)
	PUR	21.25(→0)	23.60(→0)	67.04(→0)	31.20(→0)	23.08(→0)	22.83(→0)	24.55(→0)	95.75(0.001)	**76.40**(→0)
σ=0.05	ACC	13.75(0.008)	14.00(→0)	13.82(→0)	18.40(→0)	16.66(→0)	16.42(→0)	14.80(→0)	**19.5**(→0)	18.45(→0)
	NMI	19.32(→0)	30.29(→0)	38.23(→0)	35.87(→0)	**39.87**(→0)	37.50(→0)	22.90(→0)	37.46(→0)	38.68(→0)
	PUR	15.20(→0)	16.35(→0)	65.59(→0)	20.80(→0)	19.34(→0)	22.00(→0)	16.10(→0)	24.50(→0)	**77.00**(→0)
σ=0.06	ACC	12.50(→0)	13.20(→0)	12.87(→0)	17.05(→0)	16.34(→0)	15.92(→0)	13.25(→0)	**17.00**(→0)	16.94(→0)
	NMI	18.61(0.001)	31.03(→0)	33.42(→0)	34.60(→0)	39.69(→0)	32.02 (→0)	21.38(→0)	32.06(→0)	**36.30**(→0)
	PUR	13.90(→0)	14.80(→0)	64.73(→0)	17.90(→0)	44.84(→0)	19.58(→0)	14.30(→0)	18.50(→0)	**77.45**(→0)

Table 4. Clustering results on **COIL20**(%).

	Metric	SwMC	GMC	GFSC	AMGL	RGC	GCG	MCGC	MVCF	RMvSL
σ = 0.02	ACC	77.15(→0)	82.22(→0)	64.22(→0)	66.56(0.002)	76.80(→0)	65.07(→0)	**84.61**(→0)	59.28(0.001)	83.50(0.001)
	NMI	85.56(→0)	92.71(→0)	77.52(→0)	83.37(→0)	84.51(→0)	77.68(→0)	**93.50**(→0)	76.45(→0)	91.09(→0)
	PUR	80.55(→0)	85.17(→0)	82.22(→0)	74.50(0.001)	80.07(→0)	73.75(→0)	90.11(→0)	66.67(0.001)	**90.72**(→0)
σ=0.03	ACC	66.60(0.002)	77.94(→0)	62.31(→0)	64.00(→0)	77.08(→0)	59.16(→0)	**79.66**(→0)	63.06(→0)	79.06(0.002)
	NMI	78.36(→0)	90.86(→0)	76.13(→0)	83.20(→0)	84.66(→0)	70.81(→0)	90.50(→0)	75.81(→0)	**91.13**(→0)
	PUR	71.55(→0)	81.17(→0)	82.77(→0)	72.06(0.001)	80.83(→0)	63.68(→0)	**90.11**(→0)	66.67(→0)	89.78(→0)
σ=0.04	ACC	62.25(0.003)	72.00(0.002)	57.95(→0)	64.56(0.003)	71.88(→0)	46.88(→0)	72.27(→0)	61.44(→0)	**72.33**(→0)
	NMI	73.20(0.002)	89.07(→0)	73.31(→0)	80.72(→0)	80.37(→0)	58.14(→0)	**89.38**(→0)	70.44(→0)	86.92(→0)
	PUR	66.65(0.002)	78.17(0.001)	83.06(→0)	70.89(→0)	77.03(→0)	53.33(→0)	85.50(→0)	63.72(→0)	**85.83**(→0)
σ=0.05	ACC	57.25(0.001)	56.78(→0)	51.68(→0)	57.89(0.001)	52.56(→0)	35.76(→0)	58.72(0.002)	50.61(→0)	**59.17**(→0)
	NMI	67.92(0.002)	**76.45**(→0)	72.04(→0)	74.73(→0)	64.94(→0)	45.18(→0)	75.21(→0)	61.75(→0)	73.78(0.001)
	PUR	61.65(0.002)	63.11(→0)	70.5(→0)	64.22(0.001)	58.68(→0)	40.91(→0)	73.66(0.001)	54.44(→0)	**76.44**(→0)
σ=0.06	ACC	50.35(→0)	46.33(0.002)	46.68(→0)	**48.61**(0.004)	42.15(→0)	28.68(→0)	48.00(0.001)	42.0(0.005)	47.22(→0)
	NMI	63.73(→0)	67.98(→0)	69.22(→0)	64.98(0.002)	54.98(→0)	35.58(→0)	**70.35**(0.001)	52.56(0.003)	61.77(→0)
	PUR	57.10(→0)	50.39(0.001)	66.61(→0)	56.33(0.002)	46.87(→0)	32.15(→0)	60.22(0.001)	44.78(0.004)	**75.50**(→0)

be computed with Eq.(18) during the iteration. So, the unknown parameters are just β and μ. To evaluate their influence on the clustering performance, we take the ORL as testing dataset. The experimental results under missing rate 0.3 are drawn with varying parameters. As shown in Fig.6, the two unknown parameters are not quite sensitive to three evaluation indicators. Moreover, the experimental results in section 3.2 on each dataset are obtained with the empirical parameters setting in this way. In addition, $max\left\{\left\|X^vZ^v+E^v-X^v\right\|_\infty,\left\|J^v-Z^v\right\|_\infty\right\}$ is taken as the criteria to evaluate the convergence in each iteration. The convergence curves with increasing iterative step on two selected datasets are shown



(a) ACC (b) NMI (c) PUR

Fig. 6. The influence of β and μ on clustering performance of **ORL**

in Fig.7, from which it is confirmed that our proposed method converges fast within a few of iterations.

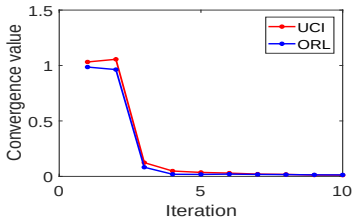


Fig. 7. Convergence curves on two selected datasets.

5 Conclusions

This paper presents a novel multiview similarity graph learning method named RMvSL for visual clustering task. With a unified objective function, RMvSL can optimize the view graphs, consensus fusion graph and constructed data alternately, which can greatly improve each other during the iteration. By introducing the low-rank self-representation model, the recovered data with distribution compensation can be obtained to learn the true similarity between pairs of instances, which shows more advantage than raw data and help to construct meaningful graphs. With the learned fusion graph, spectral clustering can be implemented to obtain the final cluster results. Extensive experiments on four visual datasets demonstrate the superior and robustness of RMvSL compared with various existing excellent similarity graph learning approaches.

6 Acknowledgments

This work was supported in part by the National Natural Science Foundation of China under Grant 62071157, University Nursing Program for Young Scholars with Creative Talents in Heilongjiang Province under Grant UNPYSCT-2018203, Natural Science Foundation of Heilongjiang Province under Grant YQ2019F011, Fundamental Research Foundation for University of Heilongjiang Province under Grant LGYC2018JQ013, and Postdoctoral Foundation of Heilongjiang Province under Grant LBH-Q19112.

References

1. Cheng, B., Yang, J., Yan, S., Fu, Y., Huang, T.S.: Learning with l1-graph for image analysis. IEEE Transactions on Image Processing **19** (2010) 858–66
2. Fang, X., Xu, Y., Li, X., Lai, Z., Wong, W.K.: Robust semi-supervised subspace clustering via non-negative low-rank representation. IEEE Transactions on Systems **46** (2016) 1828–1838
3. Nie, F., Wang, X., Huang, H.: Clustering and projected clustering with adaptive neighbors. In: ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. (2014) 977–986
4. Kang, Z., Pan, H., Hoi, S.C.H., Xu, Z.: Robust graph learning from noisy data. IEEE Transactions on Cybernetics (2019) 1–11
5. Liu, M., Luo, Y., Tao, D., Xu, C., Wen, Y.: Low-rank multi-view learning in matrix completion for multi-label image classification. National Conference on Artificial Intelligence (2015) 2778–2784
6. Wang, Q., Dou, Y., Liu, X., Lv, Q., Li, S.: Multi-view clustering with extreme learning machine. Neurocomputing **214** (2016) 483–494
7. Zhang, C., Hu, Q., Fu, H., Zhu, P., Cao, X.: Latent multi-view subspace clustering. Computer Vision and Pattern Recognition (2017) 4333–4341
8. Li, B., Yuan, C., Xiong, W., Hu, W., Peng, H., Ding, X., Maybank, S.: Multi-view multi-instance learning based on joint sparse representation and multi-view dictionary learning. IEEE Transactions on Pattern Analysis and Machine Intelligence **39** (2017) 2554–2560
9. Jing, X., Wu, F., Dong, X., Shan, S., Chen, S.: Semi-supervised multi-view correlation feature learning with application to webpage classification. Proceedings of the Thirty-First AAAI Conference on Artificial Intelligence (2017) 1374–1381
10. Wu, J., Lin, Z., Zha, H.: Essential tensor learning for multi-view spectral clustering. IEEE Transactions on Image Processing **28** (2019) 5910–5922
11. Xing, J., Niu, Z., Huang, J., Hu, W., Zhou, X., Yan, S.: Towards robust and accurate multi-view and partially-occluded face alignment. IEEE Transactions on Pattern Analysis and Machine Intelligence **40** (2018) 987–1001
12. Nie, F., Li, J., Li, X.: Parameter-free auto-weighted multiple graph learning: a framework for multiview clustering and semi-supervised classification. International Joint Conference on Artificial Intelligence (2016) 1881–1887
13. Nie, F., Li, J., Li, X.: Self-weighted multiview clustering with multiple graphs. International joint conference on artificial intelligence (2017) 2564–2570
14. Zhan, K., Shi, J., Wang, J., Wang, H., Xie, Y.: Adaptive structure concept factorization for multiview clustering. Neural Computation **30** (2018) 1080–1103
15. Zhan, K., Nie, F., Wang, J., Yang, Y.: Multiview consensus graph clustering. IEEE Transactions on Image Processing **28** (2019) 1261–1270
16. Kang, Z., Shi, G., Huang, S., Chen, W., Pu, X., Zhou, J.T., Xu, Z.: Multi-graph fusion for multi-view spectral clustering. Knowledge Based Systems **189** (2020) 102–105
17. Zhang, L., Zhang, Q., Du, B., You, J., Tao, D.: Adaptive manifold regularized matrix factorization for data clustering. International joint conference on artificial intelligence (2017) 3399–3405
18. Du, L., Shen, Y.: Unsupervised feature selection with adaptive structure learning. ACM SIGKDD Int. Conf. Knowl. Disc. Data Mining (2015) 209–218
19. Cai, S., Kang, Z., Yang, M., Xiong, X., Peng, C., Xiao, M.: Image denoising via improved dictionary learning with global structure and local similarity preservations. Symmetry **10** (2018) 167

20. Liu, G., Lin, Z., Yan, S., Sun, J., Yu, Y., Ma, Y.: Robust recovery of subspace structures by low-rank representation. IEEE Transactions on Pattern Analysis and Machine Intelligence **35** (2013) 171–184
21. Bartels, R.H., Stewart, G.W.: Solution of the matrix equation ax + xb = c [f4]. Communications of The ACM **15** (1972) 820–826
22. Candes, E.J., Li, X., Ma, Y., Wright, J.: Robust principal component analysis. Journal of the ACM **58** (2011) 11
23. Duchi, J.C., Shalevshwartz, S., Singer, Y., Chandra, T.D.: Efficient projections onto the l 1 -ball for learning in high dimensions. International Conference on Machine Learning (2008) 272–279
24. Wang, H., Yang, Y., Liu, B.: Gmc: Graph-based multi-view clustering. IEEE Transactions on Knowledge and Data Engineering (2019) 1–1
25. Ojala, T., Pietikainen, M., Maenpaa, T.: Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. IEEE Transactions on Pattern Analysis and Machine Intelligence **24** (2002) 971–987
26. Lades, M., Vorbruggen, J.C., Buhmann, J.M., Lange, J., Der Malsburg, C.V., Wurtz, R.P., Konen, W.: Distortion invariant object recognition in the dynamic link architecture. IEEE Transactions on Computers **42** (1993) 300–311