FROM LOCAL TO GLOBAL SUBSPACE CLUSTERING FOR IMAGE DATA

Maryam Abdolali, Mohammad Rahmati

Amirkabir university of technology
Department of computer engineering and information technology
Tehran, Iran

ABSTRACT

The subspace clustering problem arises in many applications that involve processing high-dimensional data, i.e. images and videos. In many of these applications, high dimensional data is often well approximated by union of low-dimensional subspaces. This motivated the development of various algorithms to cluster high dimensional data based on the underlying intrinsic low-dimensional subspaces. However, the existing approaches are based on global representation of data whereas this representation can be easily affected by errors, occlusions and severe illumination conditions. Here, we propose a multi-scale approach based on extracting local patches from different scales and then merging the shared information using a weighted scheme based on Grassmann manifolds. This approach not only benefits from the discriminative information from global representation of data but also makes the clustering task more robust using the information from local representations. Numerical results show that the proposed approach significantly outperforms existing subspace clustering algorithms.

Index Terms— Subspace clustering, Grassmann manifold, Multi-Scale approximation, Multilayer graph

1. INTRODUCTION

In many applications, including signal and image processing tasks, the collection of high dimensional data from multiple categories can be well represented by union of low-dimensional subspaces. Recovering these multiple low-dimensional structures plays a crucial role in the performance of analyzing and processing high dimensional data algorithms. This motivated the development of several algorithms for clustering data according to their underlying subspaces, a problem that is often referred to as *subspace clustering* [1]. These algorithms include iterative[2], algebraic[3], statistical[4] and spectral-based methods [5, 6].

Among many existing methods, sparse subspace clustering (SSC) [7] which merges the advances of sparse representation literature with the celebrated algorithm of spectral clustering[8], has strong theoretical guarantees[9] and works well in practice. The SSC is based on the self-expressiveness property which assumes that each sample can be represented as linear combination of other samples from the same subspace. In particular, given a data matrix $X = [x_1, x_2, ..., x_N] \in \mathbb{R}^{D\times N}$ from a collection of k linear subspace $\{S_\ell\}_{\ell=1}^{k}$, SSC optimizes the following problem[7]:

$$\min_{C\in\mathbb{R}^{N\times N}, E\in\mathbb{R}^{D\times N}} ||C||_1 + \lambda g(E) \quad (1)$$
$$\text{such that } X = XC + E \text{ and } C_{i,i} = 0 \text{ for all } i,$$

where $C \in \mathbb{R}^{N\times N}$ and $E \in \mathbb{R}^{D\times N}$ are the coefficient and error matrices respectively . The norm $||\cdot||_1 = \sum_{i,j} |C_{i,j}|$ is ℓ_1 norm which enhances the sparsity of the solution. $g(\cdot)$ is the regularization term for error matrix which is usually ℓ_1 or Frobenius norm. The obtained coefficient matrix is symmetrized using $|C| + |C|^T$ which can be considered as a data affinity matrix. By applying spectral clustering to this affinity matrix, the final clusters are obtained. The sparsity constraint and the collaborative representation are the main bases of SSC and the keys to its superior performance in many applications.

However, SSC is a global approach, in which every sample is written as linear combination of other samples entirely. This global view point can degrade the performance significantly in presence of occlusions and gross continuous blocks of error. Moreover, the traditional regularization of error matrix with ℓ_1 or Frobenius norms are insufficient in modeling contiguous noises (this is due to the implicit independent assumption for the elements of the support of error matrix in these norms). On the other hand, the performance of a mere majority voting approach based on a collection of local patches may get affected by non-discriminative patches. In order to handle gross contiguous noises, we propose an efficient multi-scale approach that elegantly integrates advantages of both global and local representations. This is carried out by combining affinity information of local patches from multiple scales using a weighted multi-layer graph based on Grassman manifolds.

To the best of our knowledge, this issue is neglected in majority of existing subspace clustering approaches. However, there are candidate works in sparse representation literature (for classification tasks)[10, 11] that highlighted the importance of local representations in collaborative representa-

tions and few works that magnified the role of local structures in correctly recovering the ground-truth intrinsic global structure for different applications[12, 13]. Nevertheless, these works mainly target supervised tasks or assume a *single* low dimensional structure for the data. Moreover, the proposed approach is based on completely different concepts.

In the remainder of this paper, we present our proposed framework in details in Section 2 and evaluate the framework on real datasets in Section 3. Finally, we conclude the paper in Section 4.

2. MULTI-SCALE PATCH-BASED SUBSPACE CLUSTERING

Subspace clustering approaches, including the state-of-the-art SSC, are based on global representation of data where the global collaborative representation is enforced by the constraint $X = XC$. However, global representation is very sensitive to severely corrupted regions of data. While local representations are more robust to these corruptions, the non-discriminative local patches might affect the performance. Moreover, the smaller the patches are, the more robust they are, but generally the less discriminative information they might contain. In order to integrate the advantages of both representations and to decrease the effect of the size of patches, we consider a multi-scale structure of data. The overview of proposed approach, referred to as *MG-SSC* (Multi-layer Graph based SSC), is presented in Fig 1.

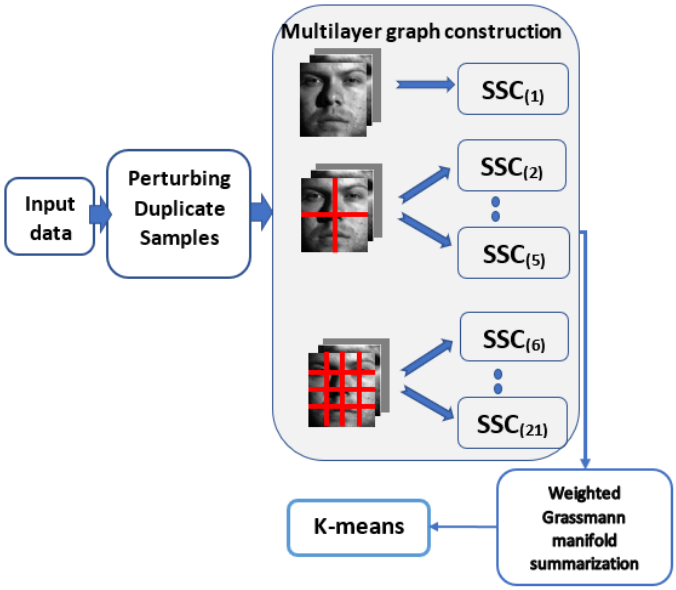

Fig. 1. Overview of proposed framework (MG-SSC).

In particular,the data matrix X first undergoes a simple pre-processing step in which duplicate samples are perturbed by a small Gaussian noise to avoid *possible* meaningless clusters. Next, we divide each sample x_j (here an image) to b non-overlapping patches. The division is conducted in a multi-scale manner, from a fine-scale (no division) to coarser scales (smaller patches) and following this structure, each obtained patch is further divided to b other patches. Usually 2-3 scales are sufficient, because in the very coarse scales, the patches would lose meaningful discriminative information. Let s be the number of scales. $(x_j^{(i)})_k$ indicates the (vectorized) k-th patch in the scale of i from the sample x_j. Each patch is considered as a representative for the corresponding sample. SSC algorithm is applied on the collection of obtained patches in different scales separately. This leads to a collection of coefficient matrices, namely $(C^{(i)})_k$ (for $k = 1, ...,$ number of patches in the scale of i and for $i = 1, ..., s$).

Let M be the total number of coefficient matrices. Each collection of patches in different scales are considered as a representative of data matrix from a different *view* point. In other words, the obtained set of (symmetrized) coefficient matrices correspond to M *multi-view* affinities of the same data matrix. Each coefficient matrix corresponds to a weighted and undirected graph, where each patch represents a vertex and the values of elements of the coefficient matrix represents weights of the edges. Obviously the number of vertices among different graphs are the same and equal to the number of data (N). Therefore, a multi-layer graph structure can be constructed from the set of coefficient matrices. Each graph encodes different information/view about the affinity and similarity of data samples.

Based on the approach in [14], the information shared by different layers can be elegantly combined using subspace analysis on a Grassmann manifold. In this work, authors reformulated the problem as finding a low dimensional representation from multiple low dimensional subspaces on a Grassmann manifold. Let L_r be the normalized laplacian matrix correspond to graph G_r which is calculated by:

$L_r = D_r^{-\frac{1}{2}}(D_r - C_r)D_r^{-\frac{1}{2}}$, for r=1,...,M

where C_r is the corresponding affinity matrix, D_r is the corresponding diagonal degree matrix and M is the total number of coefficient matrices. Let the columns of $U_r \in \mathbb{R}^{N\times k}$ be the first k eigenvectors of L_r (which corresponds to the k smallest eignevalues). The matrix U_r contains information on the connectivity of the graph and provides a k-dimensional subspace representation for the original matrix. Hence, we have a collection of k-dimensional subspaces $\{U_r\}_{r=1}^M$ and by definition, each linear subspace represents a point on a Grassmann manifold $\mathcal{G}(k, N)$. Following this line of thought, the problem of merging the information from different layers of graphs is converted to finding a low-dimensional subspace $U \in \mathbb{R}^{N\times k}$ on the Grassmann manifold which is close to the collection of subspaces. This problem is formulated as[14]:

$$\min_{U\in\mathbb{R}^{N\times k}} \sum_{r=1}^{M} tr(U'L_rU) - \alpha \sum_{r=1}^{M} tr(UU'U_rU_r') \qquad (2)$$
$$\text{such that } U'U = I,$$

where the second term is the sum of squared projection distances between representative subspace U and each individual subspace. The first term ensures that the vertex connectivity in each layer is preserved. The paramter α controls the tradeoff between these two criteria. Based on our experiments, we set the value of α to 0.5.

However, the different graphs do not contain equal amount of information. In general the graphs corresponding to coarser scales (smaller patches) contain less discriminative information and should contribute less in the final representation. Hence, we assign weights to the graphs in this multi-layer structure such that the graphs corresponding to coarser layers have lesser impact compared to the graphs corresponding to finer and more discriminant scales:

$$\min_{U\in\mathbb{R}^{N\times k}} \sum_{r=1}^{M} w_r \; tr(U' L_r U) - \alpha \sum_{r=1}^{M} w_r \; tr(UU' U_r U_r') \quad (3)$$
$$\text{such that } \; U'U = I,$$

where w_r (for $r = 1$, ..., M) are the weights. The solution to this problem can be obtained efficiently in closed form by calculating the first k eigenvectors of the following matrix:

$$L_{final} = \sum_{r=1}^{M} w_r \; L_r - \alpha \sum_{r=1}^{M} w_r \; U_r U_r'. \quad (4)$$

We suggest the following procedure to indicate the values of weights: all patches in the scale i (the fine scale is considered the first with $i = 1$) are assigned the weight $b^{\frac{s-i}{2}}$. For example if we have 3 scales ($s = 3$) and each patch is divided to 4 patches at every scale ($b = 4$), the weights corresponding to scale 1 , 2 and 3 would be 4, 2 and 1. We believe by this procedure, not only the small patches would not get overshadowed but also the entire sample in the finest scale would have the most effect on the final representation.

3. EXPERIMENTS

In this section, we evaluate the proposed approach using two real-world well-known data sets, namely AR and Extended Yale B. The clustering error is calculated by following criterion:

$$\text{error rate} = \frac{\text{\# of wrongly classified samples}}{\text{Total \# of samples}} \times 100$$

3.1. Extended Yale B data set

Extended Yale B[15] consists of face images of 38 human subject under 64 severe illumination conditions. It is generally assumed that face images of a subject under varying illuminations can be approximated by a linear subspace with intrinsic dimension of 9 [16]. Hence, the set of face images of multiple subjects can be modeled by a union of low-dimensional subspaces. For a fair comparison, the same common experimental settings as [7] are used and the performance of proposed MG-SSC is compared to other recent existing works in the literature.

For MG-SSC, 3 scales are considered ($s = 3$) and in each scale, each patch is divided into 4 non overlapping patches ($b = 4$). The average and median error rate for different number of subjects is reported in Table 1. As it can be seen, MG-SSC has a significantly better performance compared to other approaches, especially as the number of subjects increases. This significant drop in error rate highlights the effect of local representations in increasing robustness in presence of illumination variations. Moreover the weighted scheme in MG-SSC helps to emphasize the more discriminant patches (in finer scales) compared to patches in coarser scales and at the same time, benefiting from the information in the coarser scales.

Table 1. Clustering error rate (%) on the Extended Yale B data set with different number of subjects. The best performance is indicated in bold.

Algorithm	SSC	LRR[17]	LatLRR[18]	S^{3}C[19]	LRSC[20]	MG-SSC
2 subjects						
Mean	1.87	6.74	2.54	0.52	3.15	**0.06**
Median	**0**	7.03	0.78	**0**	2.34	**0**
3 subjects						
Mean	3.35	9.30	4.21	0.89	4.71	**0.09**
Median	0.78	9.90	2.60	0.52	4.17	**0**
5 subjects						
Mean	4.32	13.94	6.90	1.51	13.06	**0.15**
Median	2.19	14.38	5.63	1.25	8.44	**0**
8 subjects						
Mean	5.99	25.61	14.34	2.31	26.83	**0.20**
Median	4.49	24.80	10.06	2.25	28.71	**0.09**
10 subjects						
Mean	7.29	29.53	22.92	2.81	35.89	**0.41**
Median	5.47	30.00	23.59	2.50	34.84	**0.46**

For a better intuition, 3 specific subjects (1, 7, 10) are selected and the embedded vectors corresponding to individual graphs, namely $\{U_r\}_{r=1}^{M}$, are plotted in Fig 2. The 16 plots in the first 4 columns (from left) correspond to the coefficient matrices of the most coarse patches in the scale 3. There are 16 patches in this scale. The plots in the next column correspond to coefficient matrices of the patches in the next scale. There are 4 patches in this scale and finally the coefficient matrix corresponding to the patch in the fine scale (the whole images) is plotted in the rightmost column. For comparison, the embedded vectors obtained from coefficient matrix of SSC and our MG-SSC is plotted in Fig 3 (a) and (b) respectively. The samples belonging to the same grand-truth clusters are drawn in same colors. Note that 2 dimensions are sufficient for plotting the embedded vectors in this case. It

is evident that the embedded vectors in SSC are not discriminant enough and they appear to be very close in the middle part. In this situation, K-means cannot be successful in distinguishing the samples from different clusters which leads to the error rate of 10.94%. On the other hand the coefficient matrix from MG-SSC framework leads to clearly separated embedded vectors and hence a successful k-means clustering can be applied in this space (error rate of 0). Besides, the embedded vectors plotted in Fig 2 indicate that almost none of these plots (and hence corresponding coefficient matrices) are informative enough for accurate clustering and in fact, MG-SSC has the ability to merge their information without any prior knowledge on location of discriminative important clean patches. It can be inferred from this figure that naive techniques e.g. majority voting cannot perform well in such cases.

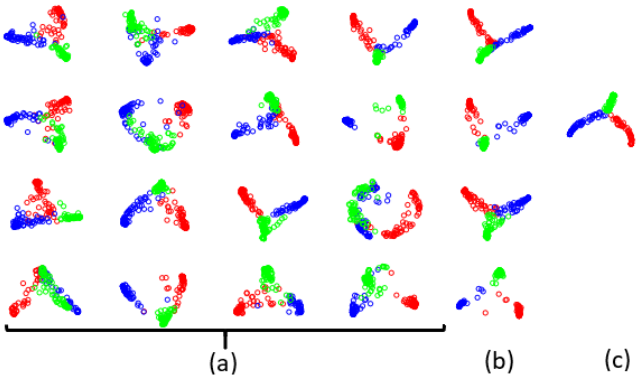

Fig. 2. Individual embedded vectors of samples from 3 clusters from different layers of graph: (a) patches from the coarsest scale, (b) patches from the middle scale and (c) the fine scale.

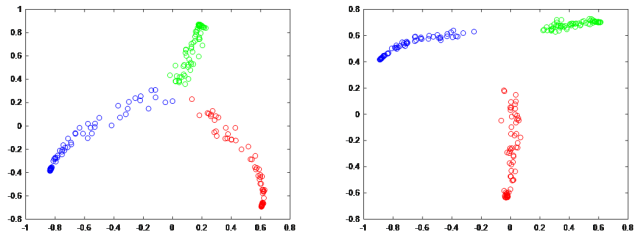

Fig. 3. Embedded vectors of samples from 3 clusters obtained from (a) SSC and (b)MG-SSC.

3.2. AR data set

AR data set [21] contains face images of 126 subjects with 26 images for each person. The images are from frontal view with variations in facial expressions, illumination, and occlusions (scarf and sunglasses). The images are resized to 55×40. For MG-SSC, 3 scales ($s = 3$) are considered and in each scale, each patch is divided into 9 non overlapping patches. The clustering performance of MG-SSC is compared with recent approaches for 5, 7 and 9 subjects from this data set in Table 2. As it can be seen MG-SSC performs the best with a very large gap compared to other methods. This indicates the effectiveness of MG-SSC in dealing with occluded images.

Table 2. Clustering error rate (%) on the AR data set with different number of subjects. The best performance is indicated in bold.

	SSC	LRR[17]	LatLRR[18]	RDLRR[22]	RSI[23]	MG-SSC
5 subjects	31.54	15.84	9.69	5.27	13.08	**0**
7 subjects	33.52	19.89	14.70	9.34	16.24	**0.85**
9 subjects	33.76	21.32	19.68	13.08	20.43	**1.10**

4. CONCLUSION

In this paper, we proposed a novel multi-scale framework for the challenging problem of subspace clustering for occluded and corrupted image data. This approach, dubbed as MG-SSC, bridges the gap between (discriminant but sensitive to error) global representation and (robust but not always informative) local representations of data. The widely used SSC algorithm is applied on extracted patches from different scales to obtain a multi-layered graph structure. The information from these layers is summarized using the embedded low dimensional subspace corresponding to each layer on a Grassmann manifold. Moreover a simple procedure is utilized to naturally assign higher weights to patches from finer scales. The numerical results show a significant improvement over previously suggested algorithms.

5. REFERENCES

[1] René Vidal, "Subspace clustering," *IEEE Signal Processing Magazine*, vol. 28, no. 2, pp. 52–68, 2011.

[2] Jeffrey Ho, Ming-Hsuan Yang, Jongwoo Lim, Kuang-Chih Lee, and David Kriegman, "Clustering appearances of objects under varying illumination conditions," in *null*. IEEE, 2003, p. 11.

[3] Ken-ichi Kanatani, "Motion segmentation by subspace separation and model selection," in *Computer Vision, 2001. ICCV 2001. Proceedings. Eighth IEEE International Conference on*. IEEE, 2001, vol. 2, pp. 586–591.

[4] Amit Gruber and Yair Weiss, "Multibody factorization with uncertainty and missing data using the em algorithm," in *Computer Vision and Pattern Recognition, 2004. CVPR 2004. Proceedings of the 2004 IEEE Computer Society Conference on*. IEEE, 2004, vol. 1, pp. I–I.

[5] Teng Zhang, Arthur Szlam, Yi Wang, and Gilad Lerman, "Hybrid linear modeling via local best-fit flats," *International journal of computer vision*, vol. 100, no. 3, pp. 217–240, 2012.

[6] Guangliang Chen and Gilad Lerman, "Spectral curvature clustering (scc)," *International Journal of Computer Vision*, vol. 81, no. 3, pp. 317–330, 2009.

[7] Ehsan Elhamifar and Rene Vidal, "Sparse subspace clustering: Algorithm, theory, and applications," *IEEE transactions on pattern analysis and machine intelligence*, vol. 35, no. 11, pp. 2765–2781, 2013.

[8] Ulrike Von Luxburg, "A tutorial on spectral clustering," *Statistics and computing*, vol. 17, no. 4, pp. 395–416, 2007.

[9] Mahdi Soltanolkotabi, Emmanuel J Candes, et al., "A geometric analysis of subspace clustering with outliers," *The Annals of Statistics*, vol. 40, no. 4, pp. 2195–2238, 2012.

[10] Shenghua Gao, Kui Jia, Liansheng Zhuang, and Yi Ma, "Neither global nor local: Regularized patch-based representation for single sample per person face recognition," *International Journal of Computer Vision*, vol. 111, no. 3, pp. 365–383, 2015.

[11] Jian Lai and Xudong Jiang, "Classwise sparse and collaborative patch representation for face recognition.," *IEEE Trans. Image Processing*, vol. 25, no. 7, pp. 3261–3272, 2016.

[12] Maryam Abdolali and Mohammad Rahmati, "Multiscale decomposition in low-rank approximation," *IEEE Signal Processing Letters*, vol. 24, no. 7, pp. 1015–1019, 2017.

[13] Fatemeh Abdolali, Reza Aghaeizadeh Zoroofi, Maryam Abdolali, Futoshi Yokota, Yoshito Otake, and Yoshinobu Sato, "Automatic segmentation of mandibular canal in cone beam ct images using conditional statistical shape model and fast marching," *International journal of computer assisted radiology and surgery*, vol. 12, no. 4, pp. 581–593, 2017.

[14] Xiaowen Dong, Pascal Frossard, Pierre Vandergheynst, and Nikolai Nefedov, "Clustering on multi-layer graphs via subspace analysis on grassmann manifolds," *IEEE Transactions on signal processing*, vol. 62, no. 4, pp. 905–918, 2014.

[15] Athinodoros S. Georghiades, Peter N. Belhumeur, and David J. Kriegman, "From few to many: Illumination cone models for face recognition under variable lighting and pose," *IEEE transactions on pattern analysis and machine intelligence*, vol. 23, no. 6, pp. 643–660, 2001.

[16] Ronen Basri and David Jacobs, "Lambertian reflectance and linear subspaces," in *Computer Vision, 2001. ICCV 2001. Proceedings. Eighth IEEE International Conference on*. IEEE, 2001, vol. 2, pp. 383–390.

[17] Guangcan Liu, Zhouchen Lin, Shuicheng Yan, Ju Sun, Yong Yu, and Yi Ma, "Robust recovery of subspace structures by low-rank representation," *IEEE transactions on pattern analysis and machine intelligence*, vol. 35, no. 1, pp. 171–184, 2013.

[18] Guangcan Liu and Shuicheng Yan, "Latent low-rank representation for subspace segmentation and feature extraction," in *Computer Vision (ICCV), 2011 IEEE International Conference on*. IEEE, 2011, pp. 1615–1622.

[19] Chun-Guang Li, Chong You, and René Vidal, "Structured sparse subspace clustering: A joint affinity learning and subspace clustering framework.," *IEEE Trans. Image Processing*, vol. 26, no. 6, pp. 2988–3001, 2017.

[20] René Vidal and Paolo Favaro, "Low rank subspace clustering (lrsc)," *Pattern Recognition Letters*, vol. 43, pp. 47–61, 2014.

[21] Aleix M Martinez, "The ar face database," *CVC Technical Report24*, 1998.

[22] Xian Zhao, Gaoyun An, Yigang Cen, Hengyou Wang, and Ruizhen Zhao, "Robust discriminant low-rank representation for subspace clustering," *Soft Computing*, pp. 1–9, 2018.

[23] Wei Siming and Lin Zhouchen, "Analysis and improvement of low rank representation for subspace segmentation," *arXiv preprint arXiv:1107.1561*, 2011.