Information Fusion 39 (2018) 168–177

Contents lists available at ScienceDirect

Information Fusion

journal homepage: www.elsevier.com/locate/inffus

Full Length Article

Multimodal sparse and low-rank subspace clustering

Mahdi Abavisani*, Vishal M. Patel

Department of Electrical and Computer Engineering, Rutgers, The State University of New Jersey, 94 Brett Road, Piscataway, NJ, 08854, USA

ARTICLE INFO

Article history:
Received 29 September 2016
Revised 12 April 2017
Accepted 14 May 2017
Available online 15 May 2017

Keywords:
Biometrics
Face clustering
Subspace clustering
Multimodal subspace clustering
Sparse subspace clustering
Low-rank representation-based subspace clustering

ABSTRACT

In this paper, we propose multimodal extensions of the recently introduced sparse subspace clustering (SSC) and low-rank representation (LRR) based subspace clustering algorithms for clustering data lying in a union of subspaces. Given multimodal data, our method simultaneously clusters data in the individual modalities according to their subspaces. In our formulation, we exploit the self expressiveness property of each sample in its respective modality and enforce the common representation across the modalities. We modify our model so that it is robust to noise. Furthermore, we kernelize the proposed algorithms to handle nonlinearities in data. The optimization problems are solved efficiently using the alternative direction method of multiplier (ADMM). Experiments on face clustering indicate the proposed method performs favorably compared to state-of-the-art subspace clustering methods.

1. Introduction

In many practical computer vision and image processing applications one has to process very high-dimensional data. In practice, these high-dimensional data can be represented by a low-dimensional subspace. For instance, face images under all possible illumination conditions, handwritten digits with different variations and trajectories of a rigidly moving object in a video can all be represented by low-dimensional subspaces [1–3]. One can view the collection of data from different classes as samples from a union of low-dimensional subspaces. In subspace clustering, the objective is to find the number of subspaces, their dimensions, the segmentation of the data and a basis for each subspace [4].

Various methods have been developed for subspace clustering in the literature. These methods can be categorized into four main groups - algebraic methods [5,6], iterative methods [7,8], statistical methods [9–11], and the methods based on spectral clustering [12–16]. In particular, sparse and low-rank representation-based subspace clustering methods [17–20] have gained a lot of interest in recent years.

Some of the multimodal spectral clustering and segmentation methods developed in recent years include [21–29]. Note that some of these algorithms use dimensionality reduction methods such as Canonical Correlation Analysis (CCA) to project the multi-view data onto a low-dimensional subspace for clustering [22,28]. Also, some of these techniques are specifically designed for two views and cannot be easily generalized to multiple views [25,29].

Various multiview sparse and low-rank representation-based subspace clustering methods have also been proposed in the literature. In particular, a multiview subspace clustering method, called Low-rank Tensor constrained Multiview Subspace Clustering (LT-MSC) was recently proposed in [30]. In the LT-MSC method, all the subspace representations are integrated into a low-rank tensor, which captures the high order correlations underlying multi-view data. In [31], a diversity-induced multiview subspace clustering was proposed in which the Hilbert Schmidt independence criterion was utilized to explore the complementarity of multi-view representations. Recently, [32] proposed a Constrained Multi-view Video Face Clustering (CMVFC) framework in which pairwise constraints are employed in both sparse subspace representation and spectral clustering procedures for multimodal face clustering. A collaborative image segmentation framework, called Multi-task Low-rank Affinity Pursuit (MLAP) was proposed in [21]. In this method, the sparsity-consistent low-rank affinities from the joint decompositions of multiple feature matrices into pairs of sparse and low-rank matrices are exploited for segmentation.

In this paper, we extend the Sparse Subspace Clustering (SSC) [17], Low-rank Representation-based (LRR) [18] subspace clustering and Low-Rank Sparse Subspace Clustering (LRSSC) [19] methods for multimodal data. In our formulation, we exploit the self expressiveness property [17] of each sample in its respective modality and enforce the common representation across the modalities. As a result, we are able to exploit the correlations as well as coupling among different modalities. Furthermore, we kernelize the

* Corresponding author.
E-mail addresses: mahdi.abavisani@rutgers.edu (M. Abavisani), vishal.m.patel@rutgers.edu (V.M. Patel).

http://dx.doi.org/10.1016/j.inffus.2017.05.002
1566-2535/

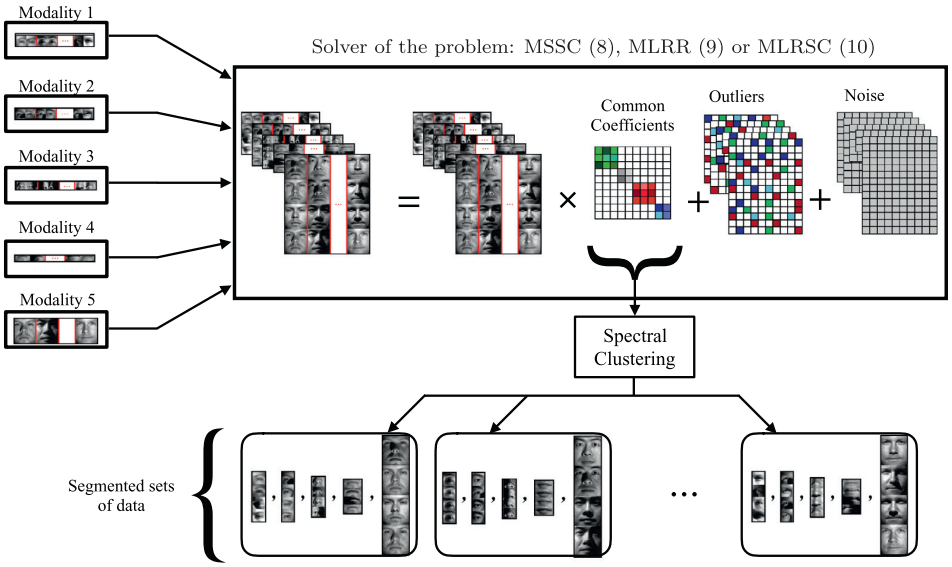


Fig. 1. An overview of the proposed multimodal sparse and low-rank subspace clustering framework.

proposed algorithms to handle nonlinearity in the data samples. The proposed optimization problems are solved using the Alternating Direction Method of Multipliers (ADMM) [33]. Fig. 1 presents an overview of our multimodal subspace clustering framework.

This paper is organized as follows. Section 2 gives a brief background on SSC, LRR and LRSSC algorithms. Details of the proposed multimodal subspace clustering algorithms are given in Section 3. Nonlinear extension of the proposed algorithms are presented in Section 4. Experimental results are presented in Section 5, and finally, Section 6 concludes the paper with a brief summary.

2. Background

In this section, we give a brief background on sparse and low-rank subspace clustering methods such as SSC [17], LRR [18] and LRSC [19].

Let $\mathbf{Y} = [\mathbf{y}_1, \cdots, \mathbf{y}_N] \in \mathbb{R}^{D\times N}$ be a collection of N signals $\{\mathbf{y}_i \in \mathbb{R}^D\}_{i=1}^N$ drawn from a union of n linear subspaces $\mathcal{S}_1 \cup \mathcal{S}_2 \cup \cdots \cup \mathcal{S}_n$ of dimensions $\{d_\ell\}_{\ell=1}^n$ in $\mathbb{R}^D$. Let $\mathbf{Y}_\ell \in \mathbb{R}^{D\times N_\ell}$ be a sub-matrix of $\mathbf{Y}$ of rank d_ℓ with $N_\ell > d_\ell$ points that lie in $\mathcal{S}_\ell$ with $N_1 + N_2 + \cdots + N_n = N$. Given $\mathbf{Y}$, the task of subspace clustering is to cluster the signals according to their subspaces.

2.1. Sparse Subspace Clustering

The SSC algorithm [17], which exploits the fact that noiseless data in a union of subspaces are *self-expressive*, i.e. each data point can be expressed as a *sparse* linear combination of other data points. Hence, SSC aims to find a sparse matrix $\mathbf{C} \in \mathbb{R}^{N\times N}$ by solving the following optimization problem

$$\min \|\mathbf{C}\|_1 \text{ s.t. } \mathbf{Y} = \mathbf{YC},\ \text{diag}(\mathbf{C}) = \mathbf{0} \tag{1}$$

where $\|\mathbf{C}\|_1 = \sum_{i,j} |C_{i,j}|$ is the ℓ_1-norm of $\mathbf{C}$. In the case when the data is contaminated by noise and outliers, one can model the data as $\mathbf{Y} = \mathbf{YC} + \mathbf{N} + \mathbf{E}$, where $\mathbf{N}$ is arbitrary noise and $\mathbf{E}$ is a sparse matrix containing outliers. In this case, the following problem can be solved to estimate the sparse coefficient matrix $\mathbf{C}$

$$\min_{\mathbf{C},\mathbf{E}} \frac{\lambda}{2}\|\mathbf{Y} - \mathbf{YC} - \mathbf{E}\|_F^2 + \|\mathbf{C}\|_1 + \lambda_e\|\mathbf{E}\|_1 \quad \text{s.t. } \text{diag}(\mathbf{C}) = \mathbf{0}, \tag{2}$$

where λ and λ_e are positive regulation parameters [34].

2.2. Low-Rank Representation-based Subspace Clustering

The LRR algorithm [18] for subspace clustering is very similar to the SSC algorithm except that a low-rank representation is found instead of a sparse representation. In particular, in the presence of noisy and occluded data, the following optimization problem is solved

$$\min_{\mathbf{C},\mathbf{E}} \frac{\lambda}{2}\|\mathbf{Y} - \mathbf{YC} - \mathbf{E}\|_F^2 + \|\mathbf{C}\|_* + \lambda_e\|\mathbf{E}\|_{2,1}, \tag{3}$$

where $\|\mathbf{C}\|_*$ is the nuclear-norm of $\mathbf{C}$ which is defined as the sum of its singular values, $\|\mathbf{E}\|_{2,1} = \sum_j \sqrt{\sum_i (E_{i,j})^2}$ is the $\ell_{2,1}$-norm of $\mathbf{E}$ and λ and λ_e are two positive regularization parameters.

2.3. Low-Rank Sparse Subspace Clustering

The representation matrix $\mathbf{C}$ can be simultaneously sparse and low-rank. Thus, LRSSC seeks to find a sparse and low-rank matrix $\mathbf{C}$ by solving the following optimization problem

$$\min_{\mathbf{C},\mathbf{E}} \frac{\lambda}{2}\|\mathbf{Y} - \mathbf{YC} - \mathbf{E}\|_F^2 + \|\mathbf{C}\|_1 + \lambda_r\|\mathbf{C}\|_* + \lambda_e\|\mathbf{E}\|_1 \quad \text{s.t. } \text{diag}(\mathbf{C}) = \mathbf{0} \tag{4}$$

where λ, λ_r and λ_e are positive regularization parameters [19].

In SSC, LRR and LRSSC, once $\mathbf{C}$ is estimated, spectral clustering methods [35] are applied on the affinity matrix $\mathbf{W} = |\mathbf{C}| + |\mathbf{C}|^T$ to obtain the segmentation of the data $\mathbf{Y}$.

3. Multimodal Sparse and Low-Rank Representation-based Subspace Clustering

As discussed earlier, classical subspace clustering methods are specifically designed for unimodal data. These methods cannot be easily extended to the case where we have heterogeneous data. Hence, in what follows, we present a multimodal extension of the sparse and low-rank subspace clustering algorithms. Given N paired data samples $\{(\mathbf{y}_i^1, \mathbf{y}_i^2, \cdots, \mathbf{y}_i^m)\}_{i=1}^N$ from m different modalities, define the corresponding data matrices as $\{\mathbf{Y}^i = [\mathbf{y}_1^i, \mathbf{y}_2^i, \cdots, \mathbf{y}_N^i] \in \mathbb{R}^{D_i\times N}\}_{i=1}^m$, respectively. We assume the m paired

set of sample points are drawn from a union of n linear subspaces in $\{\mathbb{R}^{D_i}\}_{i=1}^m$, respectively.

Given $\{\mathbf{Y}^i\}_{i=1}^m$, the task of multimodal subspace clustering is to simultaneously cluster the signals in distinct modalities according to their subspaces. In our formulation, we exploit the self expressiveness property of each sample in its respective modality, and enforce the common representation across the modalities.

In the case of data contaminated by noise and outliers, the data can be written as

$$\{\mathbf{Y}^i = \mathbf{Y}^i\mathbf{C}^i + \mathbf{N}^i + \mathbf{E}^i\}_{i=1}^m, \tag{5}$$

where $\{\mathbf{C}^i\}_{i=1}^m$, $\{\mathbf{N}^i\}_{i=1}^m$ and $\{\mathbf{E}^i\}_{i=1}^m$ are the corresponding sparse coefficient matrix, noise and error terms, respectively. Essentially based on this model, [30] proposed to integrate the subspace representations $\{\mathbf{C}^i\}_{i=1}^m$ using a low-rank tensor model, while [31] used a diversity induced framework to combine the representation coefficients from different modalities. Similarly, [21] proposed $\ell_{2,1}$ regularization on the concatenated subspace representations to enforce the affinities to have the consistent magnitudes. Finally, [32] proposed to minimize the distances between the normalized affinity matrices that are obtained by subspace clustering from each modality.

The key difference among the proposed method and the above mentioned methods is that in this paper, the subspace representations of different modalities are enforced to be the same while in some of the previous methods, the subspace representations of different modalities are different, but somehow combined by enforcing some type of regularization (i.e. tensor, $\ell_{2,1}$, diversity links, etc.) on the representations. By extracting the common sparse and/or low-rank representation structure of data across different modalities, we are able to exploit the correlations and coupling among different modalities. As a result, we can obtain a more robust subspace sparse and/or low-rank representations. In particular, we model the data as follows

$$\{\mathbf{Y}^i = \mathbf{Y}^i\mathbf{C} + \mathbf{N}^i + \mathbf{E}^i\}_{i=1}^m, \tag{6}$$

where common subspace representation $\mathbf{C}$ is enforced among all modalities. Our model is motivated by [36] and [37,38] in which common sparse representation is enforced for image super-resolution and multimodal biometrics recognition, respectively.

If the errors are sparse, then one can find $\mathbf{C}$ and $\mathbf{E} = \{\mathbf{E}^i\}_{i=1}^m$ by solving the following optimization problem

$$\min_{\mathbf{C},\mathbf{E}} \mathcal{J}(\mathbf{C},\mathbf{E}) + \frac{\lambda}{2}\sum_{i=1}^{2}\|\mathbf{Y}^i - \mathbf{Y}^i\mathbf{C} - \mathbf{E}^i\|_F^2 \quad \text{s.t. } \operatorname{diag}(\mathbf{C}) = 0. \tag{7}$$

Depending on the choice of $\mathcal{J}$, we get different algorithms for multimodal subspace clustering. For instance, if $\mathcal{J}(\mathbf{C},\mathbf{E}) = \|\mathbf{C}\|_1 + \lambda_e\|\mathbf{E}\|_1$, we get multimodal SSC (MSSC), and the resulting optimization problem becomes

$$\min_{\mathbf{C},\mathbf{E}} \|\mathbf{C}\|_1 + \lambda_e\|\mathbf{E}\|_1 + \frac{\lambda}{2}\sum_{i=1}^{2}\|\mathbf{Y}^i - \mathbf{Y}^i\mathbf{C} - \mathbf{E}^i\|_F^2 \quad \text{s.t. } \operatorname{diag}(\mathbf{C}) = 0. \tag{8}$$

When $\mathcal{J}(\mathbf{C},\mathbf{E}) = \|\mathbf{C}\|_* + \lambda_e\|\mathbf{E}\|_1$, we get multimodal LRR (MLRR). Note that in the case of MLRR, the term $\operatorname{diag}(\mathbf{C}) = \mathbf{0}$ in (7) is not required. Hence, we get the following optimization problem

$$\min_{\mathbf{C},\mathbf{E}} \|\mathbf{C}\|_* + \lambda_e\|\mathbf{E}\|_1 + \frac{\lambda}{2}\sum_{i=1}^{2}\|\mathbf{Y}^i - \mathbf{Y}^i\mathbf{C} - \mathbf{E}^i\|_F^2. \tag{9}$$

Finally, when $\mathcal{J}(\mathbf{C},\mathbf{E}) = \|\mathbf{C}\|_1 + \lambda_r\|\mathbf{C}\|_* + \lambda_e\|\mathbf{E}\|_1$, we get multimodal LRSSC (MLRSSC). In some cases, especially when the data is noisy, the term $\operatorname{diag}(\mathbf{C}) = \mathbf{0}$ may make the resulting representation matrix $\mathbf{C}$ not very low-rank. As a result, enforcing rank minimization along with the sparsity constraint with $\operatorname{diag}(\mathbf{C}) = \mathbf{0}$ in MLRSSC may not be that meaningful. Hence, we slightly modify the formulation in (7) for MLRSSC as follows

$$\min_{\mathbf{C},\mathbf{E}} \frac{\lambda}{2}\sum_{i=1}^{m}\|\mathbf{Y}^i - \mathbf{Y}^i\mathbf{A} - \mathbf{E}^i\|_F^2 + \|\mathbf{A}\|_1 + \lambda_r\|\mathbf{C}\|_* + \lambda_e\|\mathbf{E}\|_1 \quad \text{s.t. } \mathbf{A} = \mathbf{C} - \operatorname{diag}(\mathbf{C}). \tag{10}$$

Note that in our formulation, $\mathbf{E}$ is just a compact representation for $\{\mathbf{E}^i\}_{i=1}^m$. As will become apparent later, we solve each $\mathbf{E}^i$ separately since their dimensions may be different due to the different dimensionality of features in each modality (See Fig. 1). Another interesting point to note here is that when $m = 1$, the proposed multimodal algorithms reduce to their unimodal counterparts.

Similar to the unimodal subspace clustering algorithms, once $\mathbf{C}$ is estimated, spectral clustering methods can be applied on the affinity matrix $\mathbf{W} = |\mathbf{C}| + |\mathbf{C}|^T$ to obtain the simultaneous segmentation of the data $\{\mathbf{Y}^i\}_{i=1}^m$. Different steps of the proposed multimodal subspace clustering algorithms are summarized in Algorithm 1.

Algorithm 1 MSSC, MLRR, and MLRSSC Algorithms.

procedure MULTIMODAL SUBSPACE CLUSTERING($\{\mathbf{Y}^i\}_{i=1}^m$, $\lambda_e, \lambda, \lambda_r$, 'Algorithm')
  if Algorithm = MSSC **then** ▷ Obtaining $\mathbf{C}$
    Find $\mathbf{C}$ by solving (8).
  else if Algorithm = MLRR **then**
    Find $\mathbf{C}$ by solving (9).
  else if Algorithm = MLRSSC **then**
    Find $\mathbf{C}$ by solving (10).
  end if
  Normalize the columns of $\mathbf{C}$ as $\mathbf{c}_i \leftarrow \frac{\mathbf{c}_i}{\|\mathbf{c}_i\|_\infty}$.
  Form a similarity graph with N nodes and set the weights on the edges between the nodes by $\mathbf{W} = |\mathbf{C}| + |\mathbf{C}^T|$.
  Apply spectral clustering to the similarity graph.
end procedure
Output: Segmented multimodal data.

3.1. Optimization

We present an approach based on the ADMM method [33] for solving the proposed multimodal subspace clustering problems. Due to the similarity of MSSC, MLRR and MLRSSC problems, we only provide details on the optimization of the MSSC problem.

By introducing the auxiliary variables $\mathbf{U}$, and $\mathbf{Z}$, the MSSC problem (8) can be reformulated as

$$\arg\min_{\mathbf{C},\mathbf{E},\mathbf{U},\mathbf{Z}} \frac{\lambda}{2}\sum_{i=1}^{m}\|\mathbf{Y}^i - \mathbf{Y}^i\mathbf{C} - \mathbf{E}^i\|_F^2 + \|\mathbf{Z}\|_1 + \lambda_e\|\mathbf{U}\|_1 \quad \text{s.t. } \mathbf{C} = \mathbf{Z}, \mathbf{E} = \mathbf{U}, \ \operatorname{diag}(\mathbf{C}) = \mathbf{0}. \tag{11}$$

Let $f_{\alpha_C,\alpha_E}(\mathbf{C},\mathbf{E},\mathbf{Z},\mathbf{U};\mathbf{A}_E,\mathbf{A}_C)$ be the augmented Lagrangian function defined as

$$\begin{aligned}\arg\min_{\mathbf{C},\mathbf{E},\mathbf{U},\mathbf{Z}} &\frac{\lambda_n}{2}\sum_{i=1}^{m}\|\mathbf{Y}^i - \mathbf{Y}^i\mathbf{C} - \mathbf{E}^i\|_F^2 \\ &+ \|\mathbf{Z}\|_1 + \frac{\alpha_C}{2}\|\mathbf{C} - (\mathbf{Z} - \operatorname{diag}(\mathbf{Z}))\|_F^2 \\ &+ \langle\mathbf{A}_C, \mathbf{C} - (\mathbf{Z} - \operatorname{diag}(\mathbf{Z}))\rangle \\ &+ \lambda_e\sum_{i=1}^{m}\|\mathbf{U}^i\|_1 + \frac{\alpha_E}{2}\sum_{i=1}^{m}\|\mathbf{E}^i - \mathbf{U}^i\|_F^2\end{aligned} \tag{12}$$

$$+\sum_{i=1}^{m}\langle \mathbf{A_E}^i, \mathbf{E}^i - \mathbf{U}^i\rangle,$$

where $\mathbf{A}_C$ and $\mathbf{A}_E$ are the multipliers of the constrains, α_C and α_E are positive parameters and $\langle \mathbf{A}, \mathbf{B}\rangle$ denotes trace($\mathbf{A}^T\mathbf{B}$). The resulting problem can be solved using the Augmented Lagrangian Method (ALM) [39] by keeping multipliers fixed, and updating **C, E, Z, U**, and then updating multipliers $\mathbf{A}_C$ and $\mathbf{A}_E$ while keeping the other terms fixed. This process is repeated until convergence or maximum number of iterations is reached.

3.1.1. Update step for C

Fixing $\mathbf{E}_k$, $\mathbf{Z}_k$, and $\mathbf{U}_k$, $\mathbf{C}_{k+1}$ can be obtained by minimizing f_{α_C,α_E} with respect to **C**. Therefore, $\mathbf{C}_{k+1}$ is updated by solving the following linear system of equations

$$\left(\sum_{i=1}^{m}\lambda_n \mathbf{Y}^{i^T}\mathbf{Y}^i + \alpha_C\mathbf{I}\right)\mathbf{C}_{k+1} = \left(\sum_{i=1}^{m}\lambda_n \mathbf{Y}^{i^T}\left(\mathbf{Y}^i - \mathbf{E}^i\right)\right) + \alpha_C(\mathbf{Z}_k + \text{diag}(\mathbf{Z}_k)) - \mathbf{A}_{C,k}, \tag{13}$$

where $\mathbf{I}$ is an $N \times N$ identity matrix. When N is not very large, one can simply apply matrix inversion to update $\mathbf{C}_{k+1}$ from (13). For large values of N, gradient-based methods can be used to solve for $\mathbf{C}_{k+1}$.

3.1.2. Update step for E

As different modalities can have features with different dimensions, $\mathbf{E}^i$s are updated separately by minimizing f_{α_C,α_E} with respect to $\mathbf{E}^i$ as follows

$$\mathbf{E}^i_{k+1} = (1+\alpha_E)^{-1}\left(\mathbf{Y}^i - \mathbf{Y}^i\mathbf{C}_{k+1} + \alpha_E\mathbf{U}^i_k - \mathbf{A}^i_{E,k}\right),$$

where $\mathbf{A}^i_{E,k}$ is the kth update of the ith modality's multiplier.

3.1.3. Update step for Z

The variable **Z** can be updated as follows

$$\mathbf{Z}_{k+1} = \mathbf{J} - \text{diag}(\mathbf{J}),$$

where

$$\mathbf{J} \triangleq \mathcal{S}_{\frac{2}{\alpha_C}}\left(\mathbf{C}_{k+1} + \tfrac{2\mathbf{A}_{C,k}}{\alpha_C}\right),$$
$$\mathcal{S}_\eta(\nu) = (|\nu| - \eta)_+ + \text{sgn}(\nu),$$
$$(.)_+ = \begin{cases} (|\nu| - \eta), & |\nu| - \eta \geq 0 \\ 0, & \text{Otherwise.} \end{cases}$$

3.1.4. Update step for U

The update step for **U** takes the following form

$$\mathbf{U}^i_{k+1} = \mathcal{S}_{\frac{\lambda_E}{\alpha_E}}\left(\mathbf{E}^i_{k+1} + \alpha_E^{-1}\mathbf{A}^i_{E,k},\right),$$

where $\mathcal{S}_\eta$ is the shrinkage-thresholding operator defined in the previous step.

3.1.5. Update steps for $\mathbf{A}_E$ and $\mathbf{A}_C$

Finally, the multipliers are updated by gradient ascent with step sizes of α_C and α_E as follows

$$\mathbf{A}_{C,k+1} = \mathbf{A}_{C,k} + \alpha_C(\mathbf{C}_{k+1} - \mathbf{Z}_{k+1}),$$
$$\mathbf{A}^i_{E,k+1} = \mathbf{A}^i_{E,k} + \alpha_E\left(\mathbf{E}^i_{k+1} - \mathbf{U}^i_{k+1}\right).$$

3.2. Computational complexity

In this section we analyze the computational complexity of the proposed multimodal subspace clustering algorithms. We denote the number of available data points in each modality as N, the dimension of multimodal features as $\{D^i\}_{i=1}^m$ with $D_t = \sum_{i=1}^m D^i$, and the number of subspaces as n. We also assume that the needed number of iterations to reach the convergence in solving the problems (8), (9) and (10) are t_1, and spectral clustering algorithm at the final step of the Algorithm 1 needs t_2 iterations.

In general, matrix multiplication of an $M \times N$ matrix with an $N \times N$ matrix has the complexity of $O(MN^2)$, and matrix addition of two $M \times N$ matrices has the complexity of $O(MN)$. In addition, both singular value decomposition (SVD) and matrix inversion of an $N \times N$ matrix has the complexity of $O(N^3)$.

The first step of the MSSC algorithm involves updating **C**, which requires a matrix inversion, matrix multiplications and addiction operations. However, among the operations for updating **C**, the matrix inversion with the complexity of $O(N^3)$, and the multiplications with the Gram matrices with the computational complexities of $\{O(D_iN^2)\}_{i=1}^m$ can be calculated in advance, and can be used directly in the iterations. Therefore, assuming that the inverse matrix and the Gram matrices are available, updating **C** has the dominant complexity of $O(N^3 + D_tN^2)$ in each iteration. In the next step, updating each $\mathbf{E}^i$ has the dominant complexity of $O(D_iN^2)$. Updating **Z** has the complexity of $O(N^2)$ as it requires a matrix addition and thresholding each element for computing **J**. Similarly, update step for **U** requires $O(D_tN)$ computations. Afterward, updating multipliers $\mathbf{A}_C$, and $\mathbf{A}_E$ have the complexities of $O(N^2)$ and $O(D_tN)$, respectively. Therefore, as the coefficient matrix is obtained after t_1 iterations, updating steps are iterated t_1 times, which results in the overall complexity of $O(t_1(D_tN^2 + N^3))$. Finally, the spectral clustering step has the computational complexity of $O(t_2nN)$. Therefore, the overall computational complexity of the MSSC algorithm including the inversion task at the beginning of the algorithm is $O(N^3 + t_1(D_tN^2 + N^3) + t_2nN)$.

The computations in the MLRR and the MLRSSC algorithms are very similar to the MSSC algorithm, except that they have an additional step of the SVD where they calculate **Z**. However, their dominant complexities are in the same order as with the MSSC algorithm.

4. Non-Linear Multimodal Subspace Clustering

While the linear multimodal subspace clustering models (8), (9) and (10) are good approximations, in practice many datasets are better modeled by non-linear manifolds. One approach to dealing with nonlinear manifolds is to use kernel methods. Kernel-based sparse representations have been exploited before in the context of sparse coding [40], dictionary learning [41], compressed sensing [42], and subspace clustering [20,43]. It has been shown that the non-linear mapping using the kernel trick can group the data with the same distribution and make them linearly separable. In this section, we present nonlinear extensions of the proposed multimodal subspace clustering algorithms using the kernel trick.

Let $\Phi : \mathbb{R}^D \to \mathcal{H}$ be the mapping from the input space to the reproducing kernel Hilbert space $\mathcal{H}$. The kernel function $\kappa : \mathbb{R}^D \times \mathbb{R}^D \to \mathbb{R}$ is defined as the inner product $\boldsymbol{\kappa}(\mathbf{x}_i, \mathbf{x}_j) = \langle \Phi(\mathbf{x}_i), \Phi(\mathbf{x}_j)\rangle$. Then, the kernel extension of (7) without the sparse noise term **E** can be formulated as

$$\min_{\mathbf{C}} \mathcal{J}(\mathbf{C}) + \frac{\lambda}{2}\sum_{i=1}^{2}\|\Phi(\mathbf{Y}^i) - \Phi(\mathbf{Y}^i)\mathbf{C}\|_F^2 \quad \text{s.t. } \text{diag}(\mathbf{C}) = 0, \tag{14}$$

where $\Phi(\mathbf{Y}^i) = [\Phi(\mathbf{y}_1^i), \Phi(\mathbf{y}_2^i), \cdots, \Phi(\mathbf{y}_N^i)]$. This problem can be rewritten as

$$\min_{\mathbf{C}} \frac{\lambda}{2} \sum_{i=1}^{m} Tr(\mathcal{K}_{\mathbf{Y}^i\mathbf{Y}^i} - 2\mathcal{K}_{\mathbf{Y}^i\mathbf{Y}^i}\mathbf{C} + \mathbf{C}^T\mathcal{K}_{\mathbf{Y}^i\mathbf{Y}^i}\mathbf{C}) + \mathcal{J}(\mathbf{C}) \text{ s.t. } \text{diag}(\mathbf{C}) = 0, \tag{15}$$

where $[\mathcal{K}_{\mathbf{Y}^i\mathbf{Y}^i}]_{k,l} = [\langle \Phi(\mathbf{Y}^i), \Phi(\mathbf{Y}^i) \rangle]_{k,l} = \kappa(\mathbf{y}_k^i, \mathbf{y}_l^i)$, and $Tr(.)$ denotes trace operation. Similar to the linear multimodal subspace clustering methods, we apply the ADMM method to efficiently solve the problem for kernel multimodal sparse and low-rank subspace clustering. We denote the nonlinear versions of MSSC, MLRR and MLRSSC as KMSSC, KMLRR and KMLRSSC, respectively.

5. Experimental results

We evaluate the performance of our multimodal subspace clustering algorithms on five publicly available face datasets. We compare the performance of our method with several state-of-the-art subspace clustering methods such as SSC [17], LRR [15], and LRSC [16] by concatenating features from different modalities and then feeding them into these unimodal algorithms. We denote these methods as SSC-C, LRR-C and LRSC-C. In addition, we compare the performance of our method with three recently introduced state-of-the-art multimodal subspace clustering algorithms - MLAP [21], CMVFC [32], and LT-MSC [30]. Cross validation is used for parameter selection in all the experiments. Note that the MLAP algorithm requires all the modalities to have the same dimension. Therefore, the dimensions of different modalities are reduced to a common dimension (i.e. the smallest dimension among all modalities) using principal component analysis (PCA). For the experiments with the kernel multimodal subspace clustering algorithms such as KMSSC, KMLRR and KMLRSSC, we use the Gaussian kernel $\kappa(\mathbf{x}, \mathbf{y}) = \exp(-\sigma\|\mathbf{x} - \mathbf{y}\|^2)$, where σ is the parameter of the kernel function. Subspace clustering error is used to measure the performance of different algorithms. It is defined as

$$\text{subspace clustering error} = \frac{\text{\# of misclassified points}}{\text{total \# of points}} \times 100.$$

5.1. Face clustering using facial components

In the first set of experiments, we use the Extended Yale B [44], and AR face [45] datasets. We extracted four weak modalities from the face images: left and right periocular, mouth and nose regions. This was done by applying rectangular masks as shown in Fig. 2, and cropping out the respective regions. These facial components, along with the whole face, were taken as different modalities for testing our multimodal subspace clustering methods. Simple pixel intensity values were used as features for all of them.

5.1.1. Subspace clustering of the Extended Yale B dataset

The Extended Yale B dataset [44] consists of 192 × 168 size images of 38 individuals. The dataset contains 64 frontal images of each subject under varying illumination conditions. The performance of SSC, LRR and LRSC on the individual facial components is summarized in Table 2. It can be seen from this table that among all five modalities, face gives the best performance. This is not surprising as the other modalities such as mouth, nose and eyes are considered as weak modalities, and they are not as stable as faces [46]. Overall LRR and LRSC methods seem to perform better than SSC on this dataset using individual modalities.

The first and sixth rows of Table 1 summarize the results obtained by different multimodal subspace cluttering methods on the Extended Yale B dataset. Once the data from different modalities are concatenated, the dimension of the resulting multimodal vector is very large. We reduce its dimension by using a random projection matrix. We denote the resulting methods as C-RP LRR and C-RP SSC. It can be seen from this table that our proposed multimodal methods perform significantly better than MLAP, CMVFC, and LT-MSC. Furthermore, it is interesting to see that the fusion results of our multimodal methods are much better than the ones obtained using single modalities. This can be clearly seen by comparing Table 2 with the first and sixth rows of Table 1. This experiment clearly shows the significance of our common sparse and low-rank representation-based methods for subspace clustering. Also, KMSSC, KMLRR and KMLRSSC further improve the performance over MSSC, MLRR and MLRSSC, respectively.

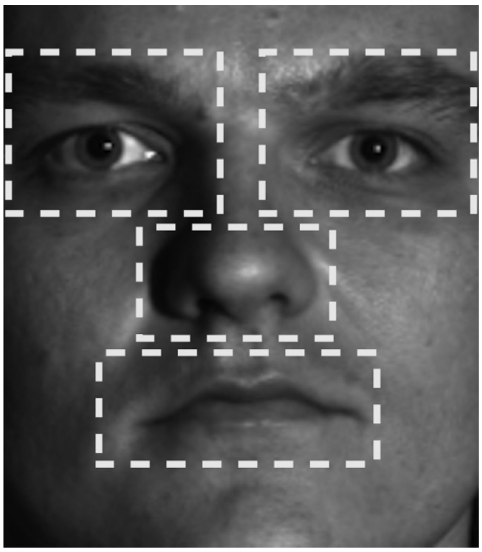


Fig. 2. Face masks used to crop out different facial components.

In Fig. 3, we show the recovered common representations corresponding to the MSSC, MLRR and MLRSSC methods. Only the images from the first four subjects are used in this experiment for better visualization. As can be seen from this figure, that the recovered coefficient matrices have block diagonal structures. In particular, the coefficient matrix corresponding to the MSSC algorithm (shown in Fig. 3 (a)) is very sparse. On the other hand, the coefficient matrix corresponding to the MLRR algorithm (shown in Fig. 3 (b)) has many nonzero coefficients that are grouped together in a given block, which essentially corresponds to low rankness of the common coefficient matrix. Since the MLRSSC algorithm provides a trade-off between sparsity and low-rank structure of the coefficient matrix, it has more non-zero coefficients that are grouped together than the matrix corresponding to the MSSC algorithm. This can be clearly seen by comparing Fig. 3(a) with Fig. 3(c).

5.1.2. Subspace clustering of the AR face dataset

The AR face dataset [45] consists of faces from 116 individuals with varying illumination, expression and occlusion conditions, captured in two sessions. In this experiment, we choose 14 images per person from the publicly available cropped dataset.[1] These images correspond to different illumination and expression variations. The performance of unimodal methods on individual components is summarized in Table 3. It is interesting to see that the performance of different methods using individual components is much

[1] Available at http://www2.ece.ohio-state.edu/~aleix/ARdatabase.html.

Table 1
Multimodal subspace clustering performance of different methods.

Experiment	Used features	SSC-C [17]	LRR-C [15]	LRSC-C [16]	MLAP [21]	CMVFC [32]	LT-MSC [30]	C-RP LRR
1 - Yale B Facial Components	Pixels	22.07	21.45	26.73	26.94	35.31	20.71	24.79
2 - AR Facial Components	Pixels	22.36	49.14	47.35	54.53	38.64	44.07	55.08
3 - Fusion of Features	Multiple features	25.14	20.13	25.45	23.63	34.61	18.76	22.69
4 - UMD-AA01	Alexnet "fc7"	24.49	46.35	39.35	34.69	27.73	30.62	36.73
5 - VIS NIR	Pixels	37.37	55.74	54.79	58.59	38.13	50.50	58.41
		C-RP SSC	MSSC	MLRR	MLRSSC	KMSSC	KMLRR	KMLRSSC
6 - Yale B Facial Components	Pixels	29.54	18.73	19.02	18.52	**12.67**	15.78	13.47
7 - AR Facial Components	Pixels	33.07	17.35	38.43	17.52	**10.58**	32.78	16.85
8 - Fusion of Features	Multiple features	30.82	23.36	18.61	18.83	23.25	**17.20**	18.43
9 - UMD-AA01	Alexnet "fc7"	23.12	22.45	32.56	27.11	**22.16**	26.23	27.89
10 - VIS NIR	Pixels	38.89	36.16	52.52	34.34	**30.30**	46.97	**30.30**

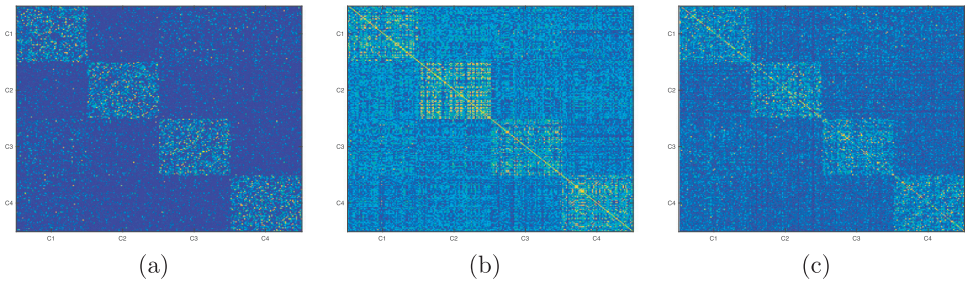


Fig. 3. Common coefficient matrices corresponding to different multimodal subspace clustering methods. Only the images from the first four subjects are used in this experiment for better visualization. **Ci** denotes coefficients of all the samples belonging to the cluster *i*. (a) The coefficient matrix corresponding to the MSSC algorithm. (b) The coefficient matrix corresponding to the MLRR algorithm. (c) The coefficient matrix corresponding to the MLRSSC algorithm.

Table 2
Clustering errors on the individual facial components of the Extended Yale B dataset.

	Left Eye	Right Eye	Nose	Mouth	Face
SSC [17]	33.91	30.49	54.74	43.48	23.76
LRR [15]	**26.28**	27.39	56.46	**31.81**	**22.52**
LRSC [16]	29.62	**25.86**	**51.93**	32.30	23.96

Table 3
Clustering errors on the individual facial components of the AR database.

	Left Eye	Right Eye	Nose	Mouth	Face
SSC [17]	**43.92**	**37.50**	72.78	68.07	**19.64**
LRR [15]	54.42	52.36	**61.79**	**61.21**	43.77
LRSC [16]	62.43	62.93	64.36	65.57	40.57

worse than using the entire face. This is mainly due to the fact that the AR dataset contains faces with various expressions. As a result, the weak modalities do not work well on this dataset.

The second and seventh rows of Table 1 summarize the results obtained by different multimodal subspace clustering methods on the AR face dataset. Although the facial components in the AR face dataset provide poor results individually, their fusion significantly enhances the performance of different subspace clustering methods. The KMSSC algorithm produces the best results on this dataset. Again this experiment shows the significance of our multimodal fusion method for subspace clustering. It is also interesting to note that MLRSSC algorithm provides a close performance to MSSC, but its nonlinear counterpart KMLRSSC cannot reach the performance of KMSSC. This can mainly happen because of sparse error subtraction in proposed linear methods that can significantly help satisfying low-rank constraints such as in MLRSSC.

Table 4
Results on the Yale B dataset: clustering errors using different facial features.

	Pixels	LBP	Gabor	HOG	PCA
SSC [17]	23.76	41.58	33.66	27.76	24.71
LRR [15]	**22.52**	**27.31**	**20.66**	**19.05**	**18.81**
LRSC [16]	23.96	33.74	36.13	33.20	20.79

5.2. Face clustering using different features

We extract different features from the face images of the Extended Yale B dataset and use them as different modalities. We extract the local binary pattern (LBP), Gabor, histogram of oriented gradients (HOG) and PCA features. Similar experiments have been conducted in [30] and [32] for face clustering.

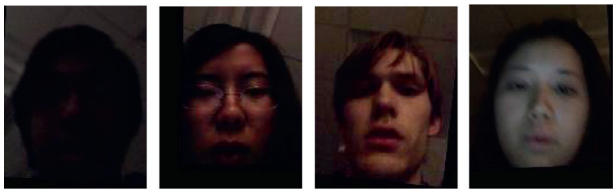
Table 4 compares the performance of different subspace clustering methods on the individual features. For comparison, results corresponding to pixels are also copied from Table 2. This table clearly shows that extracting discriminative and robust features first and then applying subspace clustering algorithms can provide better performance over just using pixel values as features.

The results obtained by different multimodal subspace clustering methods are summarized in the third and eighth rows of Table 1. We observe that almost all methods perform much better when discriminative features are used as different modalities. Furthermore, when different features are fused using our method, their performance is significantly enhanced. Also, nonlinear kernel methods improve the performance over their linear counterparts.

5.2.1. Mobile phone facial images clustering

The UMD-AA01 dataset [47] is collected on mobile devices for the original purpose of active authentication, but as it contains various ambient conditions, we use it for multimodal experiments in this paper. This dataset contains facial images of 50 users over 3 sessions corresponding to different illumination conditions. In each

(a) Session one.



(b) Session two.

(c) Session three.

Fig. 4. Sample images from different sessions in the UMD-AA01 datasets. Each session has been considered as a modality in this paper.

Table 5
Clustering errors on the individual sessions of the UMD-AA01 dataset.

	Session 1	Session 2	Session 3
SSC [17]	**37.32**	**46.36**	**40.82**
LRR [15]	41.98	47.52	48.10
LRSC [16]	44.31	48.98	44.60

session more than 750 images have been taken from each face. We randomly selected seven samples per person in each session and used them in the experiments. We used the normalization method introduced in [48], then extracted deep features corresponding to the "fc7" layer from the Alexnet convolutional neural network [49]. Fig. 4 shows some sample images from this dataset.

Table 5 reports the performance of various unimodal subspace clustering methods on the UMD-AA01 dataset. The performance of multimodal methods is also shown in the fourth and ninth rows of Table 1. As can be seen from this table the use of multimodal data can improve the subspace clustering performance over their unimodal counterparts.

5.3. Visible and infrared face images clustering

In this set of experiments, we use visible and infrared faces as different modalities. Long Distance Heterogeneous Face Database (LDHF) database [50] consists of visible and near-infrared face images of 100 individuals (70 males and 30 females). The face images were captured in both daytime and nighttime at different standoffs (e.g., 1 m, 60 m 100 m, and150 m) resulting in four VIS-NIR pairs per subject. Sample image pairs from this dataset are shown in Fig. 5. In this experiment, the face area is cropped and resized to a fixed size of 100 × 100 pixels. We simply use the pixel intensities as features.

Table 6
Results on VIS-NIR: clustering errors using visible and near infrared images.

	Visible	Near-infrared
SSC [17]	**42.17**	**49.49**
LRR [15]	57.45	61.44
LRSC [16]	58.83	60.85

Results corresponding to different unimodal subspace clustering methods are reported in Table 6. It can be seen from the table that generally visible images provide better performance in terms of clustering error. In addition, this table shows that LRR has a poor performance on this dataset. This can be explained by the fact that in this dataset, we are dealing with too many number of subjects with a few samples from each subject. It has been observed that increasing the number of subjects makes subspace clustering difficult [17].

The fifth and tenth rows of the Table 1 provide clustering errors of multimodal subspace clustering methods on the VIS-NIR dataset. We can observe that the proposed MSSC, MLRSSC, and their kernel extensions provide the best results. An interesting observation from the Table 1 is that the LT-MSC method, which is a linear low-rank representation-based method, has a slightly better performance on the VIS-NIR dataset compared to the MLRR method. Similar trend is also observed on the other datasets as well. However, it should be noted that the LT-MSC needs m more parameters to select for balancing the representations from the m modalities. While this is not the case in our MLRR method. It is interesting to note that the MLRR and KMLRR algorithms do provide significant improvements over the unimodal LRR method and the other low-rank representation-based methods.

The fact that low-rank representation-based methods in this experiment are showing weaker performances compared to the sparsity-based methods can be explained by the fact that there are a large number of subjects and low number of samples per subject in VIS-NIR dataset. Fig. 6 shows the first 12 largest singular values corresponding to one subject's data in the Extended Yale B, AR, session one in UMD-AA01 and VIS datasets. It is clear from this figure that samples in all four datasets do lie in a lower dimensional subspaces since the singular values drop quickly. In particular, each subject in the Extended Yale B dataset, AR dataset, UMD-AA01 dataset and VIS dataset, correspond to a subspace of dimension 9, 4, 4 and 3, respectively. However, considering the number of samples in each cluster one can see VIS dataset cannot show a low-rank structure as much as other datasets can show.

5.4. Impact of illumination variation

In this section, we compare the effect of illumination variations on the performance of different multimodal subspace clustering methods. We split the Yale B dataset according to different illumination variations. We choose one of the images per subject as a reference image, and the other images will be divided into four subsets according to the light angle difference from the reference. Fig. 7 shows the variation within different subsets. We apply the same rectangular masks shown in Fig. 2 for extracting the facial components.

Table 7 compares the performances of various methods on the different subsets. As expected, as illumination variations become intense, the performance of different methods drop significantly. It is interesting to see that the nonlinear methods show less dependency on the amount of variations in the sample sets. This is because kernel methods can find non-linear relations between the

(a) (b) (c) (d)

Fig. 5. Sample images from the LDHF dataset at different standoffs (a) 1m, (b) 60m, (c) 100m and (d) 150m. Visible and near-infrared images are shown in the first and the second row, respectively.

Table 7
Multimodal subspace clustering performance of different methods vs illumination variation in the data points of the Yale B face dataset.

	SSC-C [17]	LRR-C [15]	LRSC-C [16]	MLAP [21]	CMVFC [32]	LT-MSC [30]	C-RP LRR
1 - Subset 1	19.55	36.27	21.80	21.99	12.40	25.43	23.30
2 - Subset 2	37.97	46.99	25.94	24.24	18.47	34.98	28.57
3 - Subset 3	39.47	70.11	61.27	56.39	33.64	52.31	60.52
4 - Subset 4	43.23	74.06	66.72	60.33	34.39	56.52	65.03
	C-RP SSC	MSSC	MLRR	MLRSSC	KMSSC	KMLRR	KMLRSSC
5 - Subset 1	18.23	9.58	21.42	8.76	8.22	18.98	**6.39**
6 - Subset 2	34.86	16.35	23.49	16.13	**13.27**	22.34	14.47
7 - Subset 3	36.53	28.57	48.49	27.43	24.97	25.43	**23.49**
8 - Subset 4	45.48	33.83	54.50	32.71	31.22	36.27	**31.07**

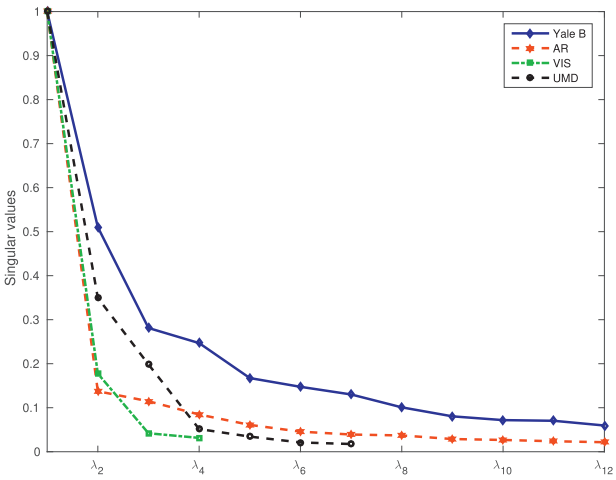
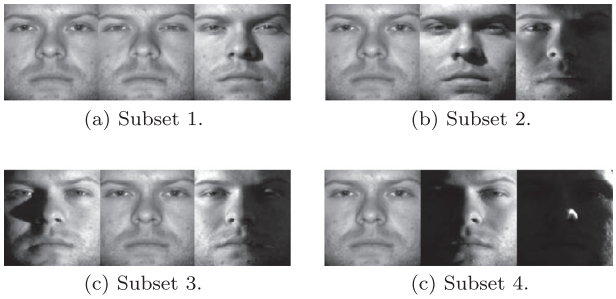


Fig. 6. First largest singular values of samples corresponding to the first person in Yale B, AR, session one in UMD-AA01 and VIS datasets.



(a) Subset 1. (b) Subset 2.

(c) Subset 3. (c) Subset 4.

Fig. 7. Illumination variation within the selected subsets in the Yale B dataset.

samples, while linear methods cannot easily deal with these variations.

5.5. Runtime comparisons

In order to compare the computational complexity of different multimodal subspace clustering methods, we measure the running time of different algorithms. Since all the compared and proposed methods are iterative algorithms, many factors such as step size of gradient descent, maximum number of iterations and choice of regulation parameters can affect their running time. Thus, we report the running time of the experiments on a specific dataset. In particular, we measure the runtime of the methods on the UMD-AA01 dataset with the same settings that resulted in the reported clustering errors in the fourth and ninth rows of Table 1. For the methods with publicly available software packages, we use their published codes. Regarding the nonlinear methods and random projection methods, calculations of finding Gram matrices and extracting the projected features are also included in the reported runtimes. Besides, each experiment is conducted 10 times, and the average runtime is reported. All the simulations were done in Matlab on an Intel® Xeon(R) 16-core machine with 3.0 GHz CPU and 32GB RAM, running Linux Ubuntu 14.04. Table 8 compares the runtime time of different methods. As can be seen from this table, the proposed methods are computationally efficient compared to some of the other subspace clustering methods.

5.6. Convergence

To empirically show the convergence of our method, in Fig. 8 (a) and (b), we show the objective function vs iteration plots of the ADMM method for solving the MSSC and KMSSC problems, respectively with the experiments on the AR dataset. As can be seen from this figure, the proposed algorithms do converge in a few it-

Table 8
Runtime of different multimodal subspace clustering algorithms on the UMD-AA01 dataset.

Method:	SSC-C [17]	LRR-C [15]	LRSC-C [16]	MLAP [21]	CMVFC [32]	LT-MSC [30]	C-RP LRR
Time (Seconds)	45.05	3.85	**0.29**	20.49	167.42	1.18	1.72
Method:	C-RP SSC	MSSC	MLRR	MLRSSC	KMSSC	KMLRR	KMLRSSC
Time (Seconds)	36.01	16.26	1.63	2.30	3.73	23.20	2.66

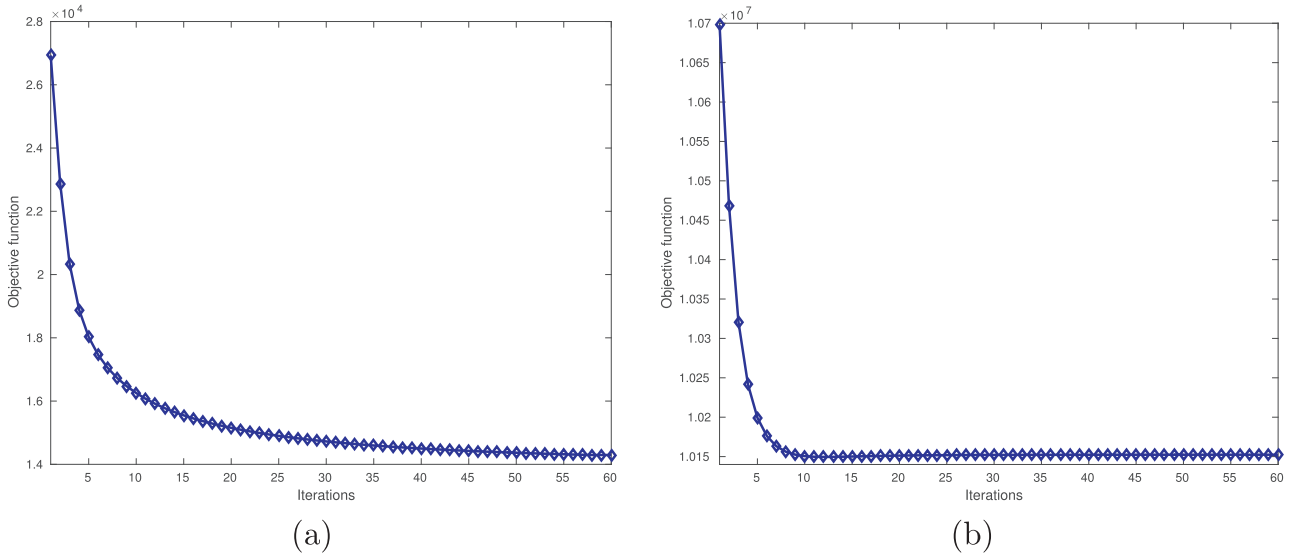


Fig. 8. Objective function of proposed algorithms versus iterations. (a) Convergence plot of the MSSC algorithm. (b) Convergence plot of the KMSSC algorithm.

erations. Experiments have shown that the MLRR, KMLRR, MLRSSC and KMLRSSC algorithms also converge in a few iterations.

6. Conclusion

We introduced multimodal extensions of the classical SSC, LRR and LRSSC methods for subspace clustering. The proposed optimization algorithms are efficiently solved using the ADMM method. Furthermore, using the kernel trick, we made the proposed multimodal subspace clustering methods nonlinear. Extensive experiments on face clustering using publicly available datasets showed that the proposed methods can perform better than many state-of-the-art multimodal subspace clustering methods.

Acknowledgment

This work was supported by the NSF grant 1618677.

References

[1] R. Basri, D.W. Jacobs, Lambertian reflectance and linear subspaces, IEEE Trans. Pattern Anal. Mach. Intell. 25 (2) (2003) 218–233.
[2] T. Hastie, P.Y. Simard, Metrics and models for handwritten character recognition, Stat. Sci. (1998) 54–65.
[3] J.P. Costeira, T. Kanade, A multibody factorization method for independently moving objects, Int. J. Comput. Vision 29 (3) (1998) 159–179.
[4] R. Vidal, Subspace clustering, IEEE Signal Process. Mag. 28 (2) (2011) 52–68, doi:10.1109/MSP.2010.939739.
[5] R. Vidal, Y. Ma, S. Sastry, Generalized principal component analysis (gpca), IEEE Trans. Pattern Anal. Mach. Intell. 27 (12) (2005) 1–15.
[6] T.E. Boult, L.G. Brown, Factorization-based segmentation of motions, in: IEEE Workshop on Visual Motion, 1991, pp. 179–186.
[7] J. Ho, M.H. Yang, J. Lim, K. Lee, D. Kriegman, Clustering appearances of objects under varying illumination conditions, in: IEEE Conference on Computer Vision and Pattern Recognition, 2003.
[8] T. Zhang, A. Szlam, G. Lerman, Median k-flats for hybrid linear modeling with many outliers, Workshop on Subspace Methods, 2009.
[9] S. Rao, R. Tron, R. Vidal, Y. Ma, Motion segmentation in the presence of outlying, incomplete, or corrupted trajectories, IEEE Trans. Pattern Anal. Mach. Intell. 32 (10) (2010) 1832–1845.
[10] H. Derksen, Y. Ma, W. Hong, J. Wright, Segmentation of multivariate mixed data via lossy coding and compression, SPIE Visual Communications and Image Processing, 6508, 2007.
[11] A.Y. Yang, S.R. Rao, Y. Ma, Robust statistical estimation and segmentation of multiple subspaces, in: IEEE Conference on Computer Vision and Pattern Recognition, 2006. 99–99
[12] A. Goh, R. Vidal, Segmenting motions of different types by unsupervised manifold clustering, in: IEEE Conference on Computer Vision and Pattern Recognition, 2007.
[13] J. Yan, M. Pollefeys, A general framework for motion segmentation: Independent, articulated, rigid, non-rigid, degenerate and non-degenerate, in: European Conference on Computer Vision, 2006.
[14] E. Elhamifar, R. Vidal, Sparse subspace clustering, in: IEEE Conference on Computer Vision and Pattern Recognition, 2009, pp. 2790–2797.
[15] G. Liu, Z. Lin, Y. Yu, Robust subspace segmentation by low-rank representation, in: International Conference on Machine Learning, 2010.
[16] P. Favaro, R. Vidal, A. Ravichandran, A closed form solution to robust subspace estimation and clustering, in: IEEE Conference on Computer Vision and Pattern Recognition, 2011.
[17] E. Elhamifar, R. Vidal, Sparse subspace clustering: algorithm, theory, and applications, IEEE Trans. Pattern Anal. Mach. Intell. 35 (11) (2013) 2765–2781.
[18] G. Liu, Z. Lin, S. Yan, J. Sun, Y. Yu, Y. Ma, Robust recovery of subspace structures by low-rank representation, IEEE Trans. Pattern Anal. Mach. Intell. 35 (1) (2013) 171–184.
[19] Y.X. Wang, H. Xu, C. Leng, Provable subspace clustering: when lrr meets ssc, in: C. Burges, L. Bottou, M. Welling, Z. Ghahramani, K. Weinberger (Eds.), Advances in Neural Information Processing Systems, 2013, pp. 64–72.
[20] V.M. Patel, H.V. Nguyen, R. Vidal, Latent space sparse and low-rank subspace clustering, IEEE J. Sel. Top. Signal Process. 9 (4) (2015) 691–701.
[21] B. Cheng, G. Liu, J. Wang, Z. Huang, S. Yan, Multi-task low-rank affinity pursuit for image segmentation, in: IEEE International Conference on Computer Vision, 2011, pp. 2439–2446.
[22] K. Chaudhuri, S.M. Kakade, K. Livescu, K. Sridharan, Multi-view clustering via canonical correlation analysis, in: International Conference on Machine Learning, 2009, pp. 129–136.
[23] A. Kumar, P. Rai, H. Daume, Co-regularized multi-view spectral clustering, in: Advances in Neural Information Processing Systems, 2011, pp. 1413–1421.
[24] X. Zhao, N. Evans, J.L. Dugelay, A subspace co-training framework for multi-view clustering, Pattern Recognit. Lett. 41 (2014) 73–82.
[25] S. Bickel, T. Scheffer, Multi-view clustering., in: IEEE International Conference on Data Mining, 4, 2004, pp. 19–26.

[26] M. White, X. Zhang, D. Schuurmans, Y. liang Yu, Convex multi-view subspace learning, in: Advances in Neural Information Processing Systems, 2012, pp. 1673–1681.
[27] H. Wang, C. Weng, J. Yuan, Multi-feature spectral clustering with minimax optimization, in: IEEE Conference on Computer Vision and Pattern Recognition, 2014, pp. 4106–4113.
[28] R. Xia, Y. Pan, L. Du, J. Yin, Robust multi-view spectral clustering via low-rank and sparse decomposition, in: Proceedings of the Twenty-Eighth AAAI Conference on Artificial Intelligence, 2014, pp. 2149–2155.
[29] V.R. de Sa, Spectral clustering with two views, ICML Workshop on Learning With Multiple Views, 2005.
[30] C. Zhang, H. Fu, S. Liu, G. Liu, X. Cao, Low-rank tensor constrained multiview subspace clustering, in: IEEE International Conference on Computer Vision, IEEE, 2015, pp. 1582–1590.
[31] X. Cao, C. Zhang, H. Fu, S. Liu, H. Zhang, Diversity-induced multi-view subspace clustering, in: IEEE Conference on Computer Vision and Pattern Recognition, 2015a, pp. 586–594.
[32] X. Cao, C. Zhang, C. Zhou, H. Fu, H. Foroosh, Constrained multi-view video face clustering, IEEE Trans. Image Process. 24 (11) (2015b) 4381–4393.
[33] S. Boyd, N. Parikh, E. Chu, B. Peleato, J. Eckstein, Distributed optimization and statistical learning via the alternating direction method of multipliers, Found. Trends Mach. Learn. 3 (1) (2011) 1–122.
[34] M. Soltanolkotabi, E. Elhamifar, E.J. Candes, et al., Robust subspace clustering, Ann. Stat. 42 (2) (2014) 669–699.
[35] U. Von Luxburg, A tutorial on spectral clustering, Stat. Comput. 17 (4) (2007) 395–416.
[36] J. Yang, J. Wright, T.S. Huang, Y. Ma, Image super-resolution via sparse representation, IEEE Trans. Image Process. 19 (11) (2010) 2861–2873.
[37] H. Zhang, V.M. Patel, R. Chellappa, Multitask multivariate common sparse representations for robust multimodal biometrics recognition, in: IEEE International Conference on Image Processing, 2015a, pp. 202–206.
[38] H. Zhang, V.M. Patel, R. Chellappa, Robust multimodal recognition via multitask multivariate low-rank representations, in: IEEE International Conference and Workshops on Automatic Face and Gesture Recognition, 2015b, pp. 1–8.
[39] M.R. Hestenes, Multiplier and gradient methods, J. Optim. Theory Appl. 4 (5) (1969) 303–320.
[40] S. Gao, I.W. Tsang, L.T. Chia, Kernel sparse representation for image classification and face recognition, in: European Conference on Computer Vision, 6314, 2010.
[41] H.V. Nguyen, V.M. Patel, N.M. Nasrabadi, R. Chellappa, Design of non-linear kernel dictionaries for object recognition, IEEE Trans. Image Process. 22 (12) (2013) 5123–5135.
[42] H. Qi, S. Hughes, Using the kernel trick in compressive sensing: Accurate signal recovery from fewer measurements, in: IEEE International Conference on Acoustics, Speech and Signal Processing, 2011, pp. 3940–3943.
[43] V.M. Patel, R. Vidal, Kernel sparse subspace clustering, in: IEEE International Conference on Image Processing, 2014.
[44] A.S. Georghiades, P.N. Belhumeur, D.J. Kriegman, From few to many: illumination cone models for face recognition under variable lighting and pose, IEEE Trans. Pattern Anal. Mach. Intell. 23 (6) (2001) 643–660.
[45] A.M. Martinez, R. Benavente, AR Face Database, Technical Report 24, CVC technical report, 1998.
[46] S. Shekhar, V.M. Patel, N.M. Nasrabadi, R. Chellappa, Joint sparse representation for robust multimodal biometrics recognition, IEEE Trans. Pattern Anal. Mach. Intell. 36 (1) (2014) 113–126.
[47] M.E. Fathy, V.M. Patel, R. Chellappa, Face-based active authentication on mobile devices, in: IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), 2015, pp. 1687–1691.
[48] H. Wang, S.Z. Li, Y. Wang, Face recognition under varying lighting conditions using self quotient image, in: IEEE International Conference on Automatic Face and Gesture Recognition, 2004, pp. 819–824.
[49] A. Krizhevsky, I. Sutskever, G.E. Hinton, Imagenet classification with deep convolutional neural networks, in: Advances in Neural Information Processing Systems, 2012, pp. 1097–1105.
[50] D. Kang, H. Han, A.K. Jain, S.-W. Lee, Nighttime face recognition at large standoff: cross-distance and cross-spectral matching, Pattern Recognit. 47 (12) (2014) 3750–3766.