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 multiview 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 multiview 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 multiview representations. Recently, [32] proposed a Constrained Multiview 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/

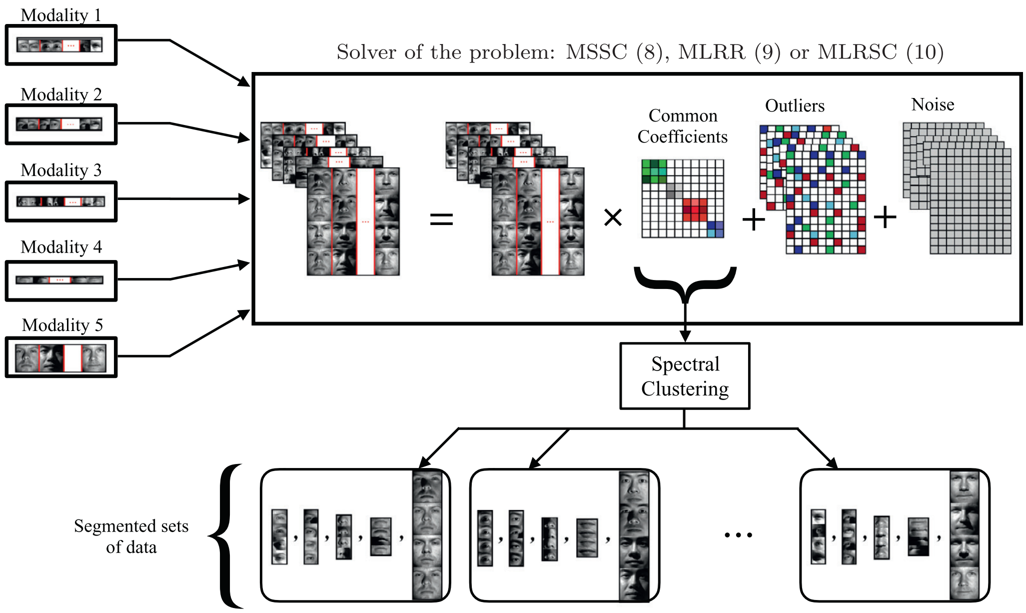

Fig. 1. An overview of the proposed multimodal sparse and low-rank subspace clustering framework.

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

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

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 $\mathbf{C}$

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

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

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

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