# A Low Tensor-Rank Representation Approach for Clustering of Imaging Data

Tong Wu, *Member, IEEE,* and Waheed U. Bajwa, *Senior Member, IEEE*

*Abstract*—This paper proposes an algorithm for clustering of two-dimensional data. Instead of "flattening" data into vectors, the proposed algorithm keeps samples as matrices and stores them as lateral slices in a third-order tensor. It is then assumed that the samples lie near a union of free submodules and their representations under this model are obtained by imposing a low tensor-rank constraint and a structural constraint on the representation tensor. Clustering is carried out using an affinity matrix calculated from the representation tensor. Effectiveness of the proposed algorithm and its superiority over existing methods are demonstrated through experiments on two image datasets.

*Index Terms*—Clustering, free submodule, tensor multi-rank

## I. INTRODUCTION

**C**LUSTERING of two-dimensional data, e.g., images, has attracted increasing attention recently. One particular approach that has proven to be remarkably effective at clustering images involves treating imaging data as lying near a *union of subspaces* (UoS) and clustering the images based on their closeness to individual subspaces. This approach, which is broadly referred to as *subspace clustering*, has resulted in state-of-the-art clustering performance in many settings; an incomplete list of works in this regard includes [1]–[9].

Traditional literature on subspace clustering, despite its effectiveness, ignores spatial aspects of images. Indeed, works like [1]–[9] require flattening of images into vectors before clustering can be carried out. However, there is an increasing realization that exploitation of the spatial aspects of imaging data using multilinear algebra tools [10]–[14] can lead to improved performance of many tasks [15]–[22]. Our goal in this paper is to leverage techniques from both multilinear algebra and abstract algebra for improved image clustering.

The main contribution of this paper is a clustering algorithm that accounts for the two-dimensional structure of images. The algorithm is based on a tensor factorization method proposed in [12], which rests on the notion of the t-product [12], [13]—a generalization of matrix multiplication for third-order tensors. The key idea of our work is to collect images as lateral slices in a three-dimensional array (i.e., tensor) and model them as lying near a *union of free submodules* (UoFS) [18]–[20]. We then represent the data tensor as a t-product of the tensor itself and a low-rank, structured coefficient tensor; an affinity matrix can be built afterward from the structured, low-tensor rank representations for clustering.

In terms of prior work, this paper is closely related to [18], [20]. Both these works are also based on the UoFS

model. However, [18] imposes a sparsity constraint on the representation tensor, while [20] requires each frontal slice of the representation tensor to be low rank. In contrast, we require the representation tensor to be both structured and low tensor rank. In order to impose the low tensor-rank condition, we make use of the *tensor nuclear norm* (TNN) [13]. In order to structurally constrain the representation tensor, we rely on the heuristic that inner products between images belonging to different free submodules should be small. The resulting algorithm, termed *structure-constrained low-rank submodule clustering* (SCLRSmC), outperforms [18], [20] in experiments on real-world image datasets. We conclude by noting that TNN has been utilized in [9] for clustering of multi-view data [23]. The algorithm of [9], however, is still based on the UoS model and reduces to a variant of [4], [5] for single-view data.

*Notation:* We use calligraphic uppercase, bold uppercase, bold lowercase, and non-bold letters for tensors, matrices, vectors, and scalars, respectively. For a third-order tensor $\mathcal{A}$, $a_{i,j,k}$ denotes its $(i, j, k)$-th element, $\mathcal{A}(i, :, :)$, $\mathcal{A}(:, i, :)$ and $\mathcal{A}(:, :, i)$ or $\mathbf{A}^{(i)}$ denote its $i$-th horizontal, lateral and frontal slices, respectively, while the $(i, j)$-th mode-1, mode-2 and mode-3 *fiber* (or *tube*) are represented by $\mathcal{A}(:, i, j)$, $\mathcal{A}(i, :, j)$ and $\mathcal{A}(i, j, :)$, respectively. We use $\widehat{\mathcal{A}} = \text{fft}(\mathcal{A}, [\,], 3)$ to denote the (orthonormal) Fourier transform along mode-3 of $\mathcal{A}$. The inner product of two tensors $\mathcal{A}, \mathcal{B} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$ is defined as $\langle \mathcal{A}, \mathcal{B} \rangle = \sum_{i,j,k} a_{i,j,k} b_{i,j,k}$. For a matrix $\mathbf{A}$, its nuclear, $\ell_1$, and $\ell_\infty$ norms are denoted by $\|\mathbf{A}\|_*$ (sum of singular values), $\|\mathbf{A}\|_1 = \sum_{i,j} |a_{i,j}|$ and $\|\mathbf{A}\|_\infty = \max_{i,j} |a_{i,j}|$, respectively. Further, the Frobenius and infinity norms of a tensor $\mathcal{A} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$ are defined as $\|\mathcal{A}\|_F = \sqrt{\sum_{i,j,k} a_{i,j,k}^2}$ and $\|\mathcal{A}\|_\infty = \max_i \|\mathbf{A}^{(i)}\|_\infty$, respectively.

## II. TECHNICAL BACKGROUND

In this section, we briefly describe the tensor algebra facts that will be used throughout this paper; further details about many of these facts can be found in [12], [13].

**Definition 1** (Identity tensor)**.** The identity tensor $\mathcal{I} \in \mathbb{R}^{n_1 \times n_1 \times n_3}$ is a tensor whose first frontal slice is an $n_1 \times n_1$ identity matrix and all other frontal slices are zero matrices.

**Definition 2** (Tensor transpose)**.** Let $\mathcal{A} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$, then $\mathcal{A}^T$ is an $n_2 \times n_1 \times n_3$ tensor obtained by transposing each of $\mathcal{A}$'s frontal slices and then reversing the order of the transposed frontal slices 2 through $n_3$, i.e., the transposed frontal slices indexed $2, \ldots, n_3$ become frontal slices $n_3, \ldots, 2$ in $\mathcal{A}^T$.

**Definition 3** (t-product [13])**.** Let $\mathcal{A} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$ and $\mathcal{B} \in \mathbb{R}^{n_2 \times n_4 \times n_3}$ be two tensors, then the t-product $\mathcal{A} * \mathcal{B}$ is an $n_1 \times$

$n_4 \times n_3$ tensor $\mathcal{C}$ whose $(i, p)$-th tube $\mathcal{C}(i, p, :)$ is $\mathcal{C}(i, p, :) = \sum_{j=1}^{n_2} \mathcal{A}(i, j, :) \circ \mathcal{B}(j, p, :)$, where $i = 1, \ldots, n_1$, $p = 1, \ldots, n_4$, and $\circ$ denotes circular convolution between two fibers.

The t-product is analogous to matrix multiplication: the circular convolution between tubes replaces scalar multiplication between matrix elements. Based on the relationship between circular convolution and the *discrete Fourier transform* (DFT), the t-product in the spatial domain corresponds to matrix multiplication of the frontal slices in the Fourier domain; that is, $\widehat{\mathbf{C}}^{(k)} = \widehat{\mathbf{A}}^{(k)} \widehat{\mathbf{B}}^{(k)}$, $k = 1, \ldots, n_3$, where $\widehat{\mathbf{A}}^{(k)}$ denotes the $k$-th frontal slice of the tensor $\widehat{\mathcal{A}}$.

**Definition 4** (Orthogonal tensor)**.** A tensor $\mathcal{Q} \in \mathbb{R}^{n_1 \times n_1 \times n_3}$ is orthogonal if $\mathcal{Q} * \mathcal{Q}^T = \mathcal{Q}^T * \mathcal{Q} = \mathcal{I}$.

Next, we define the *tensor singular value decomposition* (t-SVD) of third-order tensors.

**Definition 5** (t-SVD [12])**.** Let $\mathcal{A} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$ be a tensor, then its t-SVD is given by $\mathcal{A} = \mathcal{U} * \mathcal{S} * \mathcal{V}^T$, where $\mathcal{U} \in \mathbb{R}^{n_1 \times n_1 \times n_3}$ and $\mathcal{V} \in \mathbb{R}^{n_2 \times n_2 \times n_3}$ are orthogonal tensors, while $\mathcal{S} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$ is an f-diagonal tensor, defined as one whose frontal slices are diagonal matrices.

Note that the t-SVD of a tensor can be readily computed using SVDs of frontal slices of the Fourier tensor. Let $\text{SVD}(\widehat{\mathbf{A}}^{(k)}) = \widehat{\mathcal{U}}(:, :, k) \widehat{\mathcal{S}}(:, :, k) \widehat{\mathcal{V}}(:, :, k)$, $k = 1, \ldots, n_3$, then

$$\mathcal{U} = \text{ifft}(\widehat{\mathcal{U}}, [\,], 3), \quad \mathcal{S} = \text{ifft}(\widehat{\mathcal{S}}, [\,], 3), \quad \mathcal{V} = \text{ifft}(\widehat{\mathcal{V}}, [\,], 3), \quad (1)$$

where $\text{ifft}(\widehat{\mathcal{U}}, [\,], 3)$ denotes inverse DFT along mode-3 of $\widehat{\mathcal{U}}$.

**Definition 6** (Tensor multi-rank [13])**.** The tensor multi-rank of $\mathcal{A} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$ is a vector $\mathbf{p} \in \mathbb{R}^{n_3}$ with the $k$-th element equal to the rank of the $k$-th frontal slice of $\widehat{\mathcal{A}}$.

Finally, the *tensor nuclear norm* (TNN) of $\mathcal{A} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$ is defined as the sum of the singular values of all the frontal slices of $\widehat{\mathcal{A}}$ [17], i.e., $\|\mathcal{A}\|_{\circledast} = \sum_{i=1}^{\min(n_1, n_2)} \sum_{k=1}^{n_3} |\widehat{s}_{i,i,k}|$, where $\widehat{s}_{i,i,k}$ denotes the $(i, i, k)$-th entry of the tensor $\widehat{\mathcal{S}}$.

## III. STRUCTURE-CONSTRAINED LOW-RANK SUBMODULE CLUSTERING OF IMAGING DATA

The basic problem corresponds to a collection of $N$ images $\mathbb{Y} = \{\mathbf{Y}_j \in \mathbb{R}^{n_1 \times n_3}\}_{j=1}^{N}$ that belong to $L$ distinct categories. The challenge is to segment $\mathbb{Y}$ into $L$ subcollections (i.e., clusters). Clustering algorithms such as [1]–[9] approach this problem by *vectorizing* $\mathbf{Y}_j$'s into $\mathbf{y}_j \in \mathbb{R}^M$, $M = n_1 n_3$, and assuming that the $\mathbf{y}_j$'s belong to a union of $L$ low-dimensional subspaces in $\mathbb{R}^M$. While this vectorization approach works well in many settings, it is susceptible to errors in challenging environments since it does not explicitly take into account the spatial aspects of images. In this paper, we move away from the union-of-subspaces model and instead rely on a *union-of-free-submodules* (UoFS) model that arranges $\mathbf{Y}_j$'s as lateral slices of a tensor $\mathcal{Y} \in \mathbb{R}^{n_1 \times N \times n_3}$.

To motivate the UoFS model, we briefly discuss some abstract algebra concepts in relation to tensors of size $n_1 \times N \times n_3$. Let $\mathbb{K}_{n_3}$ denote the set of all tubes in $\mathbb{R}^{1 \times 1 \times n_3}$. Then $\mathbb{K}_{n_3}$ forms a commutative ring under regular addition and the t-product, with identity element given by $\vec{e} = [1, 0, 0, .., 0]$
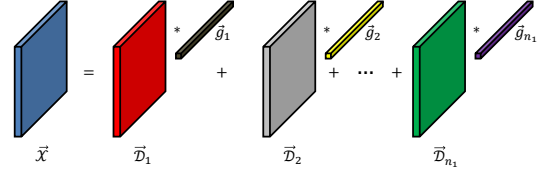


Fig. 1. Any element of the free module $\mathbb{K}_{n_3}^{n_1}$ over the commutative ring $\mathbb{K}_{n_3}$ can be expressed as a t-linear combination of the generating set.

[12]. Next, let $\mathbb{K}_{n_3}^{n_1}$ denote the set of $n_1 \times 1 \times n_3$ lateral slices. We can think of the members of $\mathbb{K}_{n_3}^{n_1}$ as vectors of length $n_1$ in which each element itself is a $1 \times 1 \times n_3$ tube. It then follows that $\mathbb{K}_{n_3}^{n_1}$ forms a *free module* of dimension $n_1$ over $\mathbb{K}_{n_3}$ since one can construct a generating set $\{\vec{\mathcal{D}}_i \in \mathbb{K}_{n_3}^{n_1}\}_{i=1}^{n_1}$ such that any $\vec{\mathcal{X}} \in \mathbb{K}_{n_3}^{n_1}$ can be uniquely represented as a *t-linear combination* of the $\vec{\mathcal{D}}_i$'s: $\vec{\mathcal{X}} = \sum_{i=1}^{n_1} \vec{\mathcal{D}}_i * \vec{g}_i$ for $\vec{g}_i \in \mathbb{K}_{n_3}$ [11] (see Fig. 1). Thus, the free module $\mathbb{K}_{n_3}^{n_1}$ over the ring $\mathbb{K}_{n_3}$ generalizes the concept of a vector space over a field. The advantage of this representation is that it allows data to be generated from shifted copies of the generating set, making any resulting method robust to spatial shifts [18], [19].

We now return to the tensor $\mathcal{Y} \in \mathbb{R}^{n_1 \times N \times n_3}$ and consider each $\mathbf{Y}_j$ as an element $\vec{\mathcal{Y}}_j \in \mathbb{K}_{n_3}^{n_1}$. Now consider the analogue of a vector subspace in the context of a free module, viz., a free submodule. Specifically, an $s$-dimensional free submodule $\mathcal{S}_{n_3}^{n_1}$ of $\mathbb{K}_{n_3}^{n_1}$ is a subset of $\mathbb{K}_{n_3}^{n_1}$ in which each element can be represented as a t-linear combination of $s < n_1$ elements of $\mathbb{K}_{n_3}^{n_1}$. Our assumption is that the images, viewed as lateral slices, belong to a union of $L$ free submodules $\{{}^{\ell}\mathcal{S}_{n_3}^{n_1}\}_{\ell=1}^{L}$ of dimensions $\{s_\ell < n_1\}_{\ell=1}^{L}$. Our goal then is to segment images into clusters with each cluster corresponding to a submodule.

### A. Structure-constrained Low-rank Submodule Clustering

Our approach to clustering images under the UoFS model involves exploitation of the self-expressiveness property of free submodules: an image belonging to a free submodule can be represented as a t-linear combination of other images belonging to the same free submodule.[1] Mathematically, this means there exists a tensor $\mathcal{Z} \in \mathbb{R}^{N \times N \times n_3}$ such that $\mathcal{Y} \approx \mathcal{Y} * \mathcal{Z}$. In addition, since images belonging to one submodule should struggle to represent images belonging to another submodule, we should look for $\mathcal{Z}$'s that are permutations of f-block diagonal tensors (i.e., whose frontal slices are block-diagonal matrices). Once we obtain a representation tensor, say, $\mathcal{Z}^*$, that satisfies these requirements, we can build an affinity matrix $\mathbf{W}$ with $(i, j)$-th entry $w_{i,j} = \|\mathcal{Z}^*(i, j, :)\|_F + \|\mathcal{Z}^*(j, i, :)\|_F$ and use spectral clustering [24] to obtain final clustering.[2] The challenge then is developing a method that leverages the self-expressiveness property to learn a representation tensor $\mathcal{Z}$ that enforces the UoFS structure on the lateral slices of $\mathcal{Y}$.

There exist a number of recipes in the literature on subspace clustering for learning of self-expressive representations. Among these, the low-rank constraint of [4] and the structure

---

[1]This is similar to the self-expressiveness property that is exploited in the literature on subspace clustering. Note that an implicit assumption in use of such properties is that the number of linearly independent images in each cluster exceeds dimensionality of the underlying geometric structure.

[2]In this paper, we assume the number of clusters $L$ is known a priori.

constraint of [6] produce some of the best results. We extend these ideas to free submodule clustering and learn a $\mathcal{Z}$ that not only approximates the data tensor ($\mathcal{Y} \approx \mathcal{Y} * \mathcal{Z}$), but that also has a small multi-rank (measured in $\ell_1$ norm) as well as a pronounced f-block diagonal structure. To obtain $\mathcal{Z}$ with a low multi-rank, we minimize its TNN $\|\mathcal{Z}\|_\circledast$. This is because TNN is the tightest convex relaxation of $\ell_1$ norm of the tensor multi-rank [17]. To explicitly impose the f-block diagonal structure on $\mathcal{Z}$, we use the intuition that images belonging to different free submodules should have lower correlations. We capture this intuition in terms of a *dissimilarity* matrix $\mathbf{M} \in [0,1]^{N \times N}$ whose $(i,j)$-th entry is defined as follows: let $\widetilde{\mathcal{Y}}$ be the tensor obtained by normalizing each lateral slice of $\mathcal{Y}$ by its corresponding Frobenius norm $\|\mathcal{Y}(:,i,:)\|_F$; then, $m_{i,j} = 1 - \exp\left(-\frac{1-|\langle \widetilde{\mathcal{Y}}(:,i,:), \widetilde{\mathcal{Y}}(:,j,:)\rangle|}{\sigma}\right)$, where $\sigma$ denotes the empirical average of all $1-|\langle \widetilde{\mathcal{Y}}(:,i,:), \widetilde{\mathcal{Y}}(:,j,:)\rangle|$'s. Integrating the goals of a low multi-rank, f-block diagonal structure, and small approximation error into a single objective function, we finally pose the problem of *structure-constrained low-rank submodule clustering* (SCLRSmC) as follows:

$$\min_{\mathcal{Z}} \|\mathcal{Z}\|_\circledast + \lambda_1 \sum_{k=1}^{n_3} \|\mathbf{M} \odot \mathbf{Z}^{(k)}\|_1 + \lambda_2 \|\mathcal{Y} - \mathcal{Y} * \mathcal{Z}\|_F^2, \quad (2)$$

where $\odot$ denotes elementwise multiplication, while $\lambda_1$ and $\lambda_2$ are penalty parameters.

### B. An Augmented Lagrangian Method for SCLRSmC

To solve (2) in an efficient manner, we resort to "variable splitting" of $\mathcal{Z}$, which transforms (2) into the following:

$$\min_{\mathcal{C}, \mathcal{Q}, \mathcal{Z}} \|\mathcal{C}\|_\circledast + \lambda_1 \sum_{k=1}^{n_3} \|\mathbf{M} \odot \mathbf{Q}^{(k)}\|_1 + \lambda_2 \|\mathcal{Y} - \mathcal{Y} * \mathcal{Z}\|_F^2$$
$$\text{subject to} \quad \mathcal{Z} = \mathcal{C}, \text{ and } \mathcal{Z} = \mathcal{Q}. \quad (3)$$

This constrained problem can now be solved using augmented Lagrangian (AL) methods [25]. Specifically, the AL of (3) is

$$\mathcal{L}(\mathcal{C}, \mathcal{Q}, \mathcal{Z}, \mathcal{G}_1, \mathcal{G}_2, \mu) = \|\mathcal{C}\|_\circledast + \lambda_1 \sum_{k=1}^{n_3} \|\mathbf{M} \odot \mathbf{Q}^{(k)}\|_1$$
$$+ \lambda_2 \|\mathcal{Y} - \mathcal{Y} * \mathcal{Z}\|_F^2 + \langle \mathcal{G}_1, \mathcal{Z} - \mathcal{C} \rangle + \langle \mathcal{G}_2, \mathcal{Z} - \mathcal{Q} \rangle$$
$$+ \frac{\mu}{2}\left(\|\mathcal{Z} - \mathcal{C}\|_F^2 + \|\mathcal{Z} - \mathcal{Q}\|_F^2\right), \quad (4)$$

where the tensors $\mathcal{G}_1$ and $\mathcal{G}_2$ comprise Lagrange multipliers and $\mu \geq 0$ is the AL penalty parameter [25]. We can now use any inexact AL technique [26], [27] and solve (2) by iteratively minimizing the AL $\mathcal{L}(\mathcal{C}, \mathcal{Q}, \mathcal{Z}, \mathcal{G}_1, \mathcal{G}_2, \mu)$ over one tensor at a time while keeping the others fixed.

**Updating $\mathcal{C}$:** Keeping other tensors in (4) fixed, the subproblem of updating $\mathcal{C}$ at the $(t+1)$-th iteration has the form

$$\mathcal{C}^{(t+1)} = \arg\min_{\mathcal{C}} \|\mathcal{C}\|_\circledast + \langle \mathcal{G}_1^{(t)}, \mathcal{Z}^{(t)} - \mathcal{C} \rangle + \frac{\mu^{(t)}}{2}\|\mathcal{Z}^{(t)} - \mathcal{C}\|_F^2$$
$$= \arg\min_{\mathcal{C}} \|\mathcal{C}\|_\circledast + \frac{\mu^{(t)}}{2}\left\|\mathcal{C} - \left(\mathcal{Z}^{(t)} + \frac{\mathcal{G}_1^{(t)}}{\mu^{(t)}}\right)\right\|_F^2. \quad (5)$$

This subproblem can be solved using [9, Theorem 2].

**Updating $\mathcal{Q}$:** Fixing other tensors in (4), the subproblem of updating $\mathcal{Q}$ can be expressed as

$$\mathcal{Q}^{(t+1)} = \arg\min_{\mathcal{Q}} \lambda_1 \sum_{k=1}^{n_3} \|\mathbf{M} \odot \mathbf{Q}^{(k)}\|_1 + \langle \mathcal{G}_2^{(t)}, \mathcal{Z}^{(t)} - \mathcal{Q} \rangle$$
$$+ \frac{\mu^{(t)}}{2}\|\mathcal{Z}^{(t)} - \mathcal{Q}\|_F^2. \quad (6)$$

We can decompose (6) into $n_3$ independent subproblems, with the $k$-th frontal slice $\mathbf{Q}^{(k)^{(t+1)}}$ of $\mathcal{Q}$ being updated as

$$\min_{\mathbf{Q}} \lambda_1 \|\mathbf{M} \odot \mathbf{Q}\|_1 + \frac{\mu^{(t)}}{2}\left\|\mathbf{Q} - (\mathbf{Z}^{(k)^{(t)}} + \frac{\mathbf{G}_2^{(k)^{(t)}}}{\mu^{(t)}})\right\|_F^2, \quad (7)$$

which has a closed-form solution given in [6, Proposition 3].

**Updating $\mathcal{Z}$:** The subproblem of updating $\mathcal{Z}$ has the form

$$\mathcal{Z}^{(t+1)} = \arg\min_{\mathcal{Z}} \lambda_2 \|\mathcal{Y} - \mathcal{Y} * \mathcal{Z}\|_F^2 + \langle \mathcal{G}_1^{(t)}, \mathcal{Z} - \mathcal{C}^{(t+1)} \rangle +$$
$$\langle \mathcal{G}_2^{(t)}, \mathcal{Z} - \mathcal{Q}^{(t+1)} \rangle + \frac{\mu^{(t)}}{2}(\|\mathcal{Z} - \mathcal{C}^{(t+1)}\|_F^2 + \|\mathcal{Z} - \mathcal{Q}^{(t+1)}\|_F^2)$$
$$= \arg\min_{\mathcal{Z}} \lambda_2 \|\mathcal{Y} - \mathcal{Y} * \mathcal{Z}\|_F^2$$
$$+ \frac{\mu^{(t)}}{2}\left(\|\mathcal{Z} - \mathcal{P}_1^{(t+1)}\|_F^2 + \|\mathcal{Z} - \mathcal{P}_2^{(t+1)}\|_F^2\right), \quad (8)$$

where $\mathcal{P}_1^{(t+1)} = \mathcal{C}^{(t+1)} - \frac{\mathcal{G}_1^{(t)}}{\mu^{(t)}}$ and $\mathcal{P}_2^{(t+1)} = \mathcal{Q}^{(t+1)} - \frac{\mathcal{G}_2^{(t)}}{\mu^{(t)}}$. We now transform (8) into the Fourier domain to obtain

$$\widehat{\mathcal{Z}}^{(t+1)} = \arg\min_{\widehat{\mathcal{Z}}} \lambda_2 \|\widehat{\mathcal{Y}} - \widehat{\mathcal{Y}} \otimes \widehat{\mathcal{Z}}\|_F^2$$
$$+ \frac{\mu^{(t)}}{2}\left(\|\widehat{\mathcal{Z}} - \widehat{\mathcal{P}}_1^{(t+1)}\|_F^2 + \|\widehat{\mathcal{Z}} - \widehat{\mathcal{P}}_2^{(t+1)}\|_F^2\right), \quad (9)$$

where $\otimes$ denotes slicewise multiplication. We again break (9) into $n_3$ subproblems, with update of the $k$-th frontal slice of $\widehat{\mathcal{Z}}$ given by $\widehat{\mathbf{Z}}^{(k)^{(t+1)}} = \arg\min_{\widehat{\mathbf{Z}}^{(k)}} f(\widehat{\mathbf{Z}}^{(k)})$ with $f(\widehat{\mathbf{Z}}^{(k)}) = \lambda_2 \|\widehat{\mathbf{Y}}^{(k)} - \widehat{\mathbf{Y}}^{(k)}\widehat{\mathbf{Z}}^{(k)}\|_F^2 + \frac{\mu^{(t)}}{2}(\|\widehat{\mathbf{Z}}^{(k)} - \widehat{\mathbf{P}}_1^{(k)^{(t+1)}}\|_F^2 + \|\widehat{\mathbf{Z}}^{(k)} - \widehat{\mathbf{P}}_2^{(k)^{(t+1)}}\|_F^2)$. It can then be shown that

$$\widehat{\mathbf{Z}}^{(k)^{(t+1)}} = (2\lambda_2 \widehat{\mathbf{Y}}^{(k)^T}\widehat{\mathbf{Y}}^{(k)} + 2\mu^{(t)}\mathbf{I})^{-1}$$
$$\left(2\lambda_2 \widehat{\mathbf{Y}}^{(k)^T}\widehat{\mathbf{Y}}^{(k)} + \mu^{(t)}(\widehat{\mathbf{P}}_1^{(k)^{(t+1)}} + \widehat{\mathbf{P}}_2^{(k)^{(t+1)}})\right). \quad (10)$$

After updating each frontal slice of $\widehat{\mathcal{Z}}^{(t+1)}$, the final update for $\mathcal{Z}^{(t+1)}$ can be expressed as $\mathcal{Z}^{(t+1)} = \text{ifft}(\widehat{\mathcal{Z}}^{(t+1)}, [\ ], 3)$. The overall algorithm is summarized in Algorithm 1.

## IV. EXPERIMENTAL RESULTS

We evaluate the clustering performance of SCLRSmC on the UCSD dynamic scenes dataset[3] and the MNIST dataset[4]. The reported results, corresponding to $\rho = 1.9$, $\mu^{(0)} = 0.1$, $\mu_{\max} = 10^{10}$, and $\epsilon = 10^{-5}$, are compared with those obtained using several state-of-the-art clustering methods, namely, SLRSmC [20], SSmC [18], SSC [5], LRR [4], and SC-LRR [6]. Note that the parameters in each of these methods are tuned to achieve the best clustering performance.

The UCSD dynamic scene dataset consists of 18 video sequences and manifests a variety of challenges since the

**Algorithm 1** An Inexact AL Method for SCLRSmC
***
**Input:** Data $\mathcal{Y}$, matrix $\mathbf{M}$, and parameters $\lambda_1, \lambda_2, \rho, \mu_{\max}, \epsilon$.
**Initialize:** Penalty parameter $\mu^{(0)}$, tensors $\mathcal{C}^{(0)} = \mathcal{Q}^{(0)} = \mathcal{Z}^{(0)} = \mathcal{G}_1^{(0)} = \mathcal{G}_2^{(0)} \leftarrow 0 \in \mathbb{R}^{N \times N \times n_3}$, and $t \leftarrow 0$.

1: **while** not converged **do**
2:    Fix other tensor variables and update $\mathcal{C}$ by solving (5).
3:    Fix other tensor variables and update $\mathcal{Q}$ by solving (6).
4:    Fix other tensor variables and update $\mathcal{Z}$ by solving (8).
5:    $\mathcal{G}_1^{(t+1)} \leftarrow \mathcal{G}_1^{(t)} + \mu^{(t)}(\mathcal{Z}^{(t+1)} - \mathcal{C}^{(t+1)})$.
6:    $\mathcal{G}_2^{(t+1)} \leftarrow \mathcal{G}_2^{(t)} + \mu^{(t)}(\mathcal{Z}^{(t+1)} - \mathcal{Q}^{(t+1)})$.
7:    $\mu^{(t+1)} \leftarrow \min(\mu_{\max}, \rho\mu^{(t)})$.
8:    **break if**

$$\max \begin{Bmatrix} \|\mathcal{Z}^{(t+1)} - \mathcal{C}^{(t+1)}\|_\infty, \|\mathcal{Z}^{(t+1)} - \mathcal{Q}^{(t+1)}\|_\infty, \\ \|\mathcal{Z}^{(t+1)} - \mathcal{Z}^{(t)}\|_\infty, \|\mathcal{C}^{(t+1)} - \mathcal{C}^{(t)}\|_\infty, \\ \|\mathcal{Q}^{(t+1)} - \mathcal{Q}^{(t)}\|_\infty \end{Bmatrix} < \epsilon.$$

9:    $t \leftarrow t + 1$.
10: **end while**
***
**Output:** Representation tensor $\mathcal{Z}^* = \mathcal{Z}^{(t+1)}$.

videos involve significant camera motion and highly dynamic backgrounds such as water, smoke, and fire. We select 10 sequences from the dataset, labeled birds, boats, bottle, chopper, cyclists, flock, hockey, landing, ocean, and skiing. Each of these sequences comprises 30 to 246 images, which we downsample to $90 \times 135$. Our results correspond to random selection of $L \in \{3, 4, 5, 6, 7\}$ sequences from the 10 categories, which are then input to the clustering algorithms. The final results are shown in Table I, which correspond to an average of 20 randomly selected sequences for each value of $L$. Clearly, SCLRSmC outperforms other methods by a large margin, which demonstrates its effectiveness.

TABLE I
CLUSTERING ERROR (%) ON THE UCSD DYNAMIC SCENES DATASET
WITH DIFFERENT NUMBER OF CLUSTERS ($L$)

| $L$ | SCLRSmC | SLRSmC | SSmC | SSC | LRR | SC-LRR |
|---|---|---|---|---|---|---|
| 3 | **1.27** | 10.69 | 21.59 | 3.11 | 9.28 | 7.14 |
| 4 | **3.83** | 16.81 | 29.29 | 16.91 | 18.86 | 10.9 |
| 5 | **4.59** | 26.36 | 33.16 | 17.35 | 20.46 | 20.23 |
| 6 | **3.77** | 25.55 | 27.93 | 18.57 | 19.80 | 20.41 |
| 7 | **4.11** | 28.83 | 31.72 | 19.71 | 21.24 | 25.25 |

Our next set of results corresponds to the relatively easier MNIST dataset [28], which comprises 70,000 *centered* $28 \times 28$ images of handwritten digits. Subspace clustering methods are ideal for this dataset since it can be argued that vectorized images of each digit in this dataset approximately lie near a subspace [29]. Similar to [7], we consider clustering of digits $\{2, 4, 8\}$ and randomly select 100 images of each of these digits (i.e., $N = 300$). The results, averaged over 20 random trials, are presented in the first row of Table II. Within the class of submodule clustering methods (SCLRSmC, SLRSmC, and SSmC), we observe that SCLRSmC outperforms the other two methods. At the same time, SC-LRR and SSC—both of which are based on the UoS model—outperform SCLRSmC. We conjecture that the reason for this is perfect spatial alignment of MNIST digits. Specifically, the approximation $\mathcal{Y} \approx \mathcal{Y} * \mathcal{Z}$

means SCLRSmC is trying to represent every pixel in a given row of an image as a linear combination of *all* pixels in the corresponding rows of other images. While this imparts robustness to spatial shifts, it also means SCLRSmC solves a significantly harder problem in the case of aligned images. Indeed, in the case of perfect alignment, every pixel in an image should be representable as a linear combination of corresponding pixels in other images. This is precisely the problem solved by subspace clustering, which likely explains its superior performance in the case of aligned images.

TABLE II
CLUSTERING ERROR (%) ON THE MNIST HANDWRITTEN DIGIT DATASET

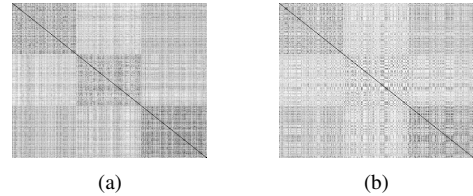| | SCLRSmC | SLRSmC | SSmC | SSC | LRR | SC-LRR |
|---|---|---|---|---|---|---|
| Aligned | 10.5 | 24.7 | 28.88 | 5.57 | 19.58 | **3.42** |
| Shifted | **15.5** | 21.83 | 28.88 | 42.75 | 46.42 | 36.82 |



(a)          (b)

Fig. 2. Visualization of the dissimilarity matrix $\mathbf{M}$ for (a) aligned and (b) spatially shifted MNIST images ($L = 3$). Here, lighter shades of gray represent higher dissimilarity among images.

In order to validate our earlier conjecture as well as test the robustness of SCLRSmC against spatial shifts, we modify the MNIST dataset by adding a random (left or right) 6-pixel horizontal shift to each image [18]. Note that these spatial shifts also reduce correlations among images belonging to the same category. This, in turn, changes the dissimilarity matrix $\mathbf{M}$, which plays a central role in SCLRSmC. It can, however, be seen from Fig. 2 that these changes are not catastrophic (compare, e.g., Fig. 2(a) and Fig. 2(b)). In particular, while the block diagonal structure of $\mathbf{M}$ gets less pronounced in the case of spatially shifted images, it is not completely eliminated. Finally, the clustering results under the modified MNIST dataset are reported in the second row of Table II. We make two observations from these results. First, SCLRSmC now outperforms all other methods. Second, performance of all subspace clustering methods significantly degrades in the presence of unaligned data. This points to the usefulness of the UoFS model for clustering of unaligned imaging data.

## V. CONCLUSION

In this paper, we presented a novel method for clustering of imaging data. The proposed method stacks images as lateral slices in a third-order tensor and then models them as lying near a union of low-dimensional free submodules. In order to segment images into their respective free submodules, it makes use of the self-expressiveness property of free submodules and uses an inexact augmented Lagrangian method to learn a representation tensor that has a low multi-rank and f-block diagonal structure. Numerical experiments performed on two real-world datasets demonstrate usefulness of the proposed method, especially for clustering of unaligned images. Possible future works include improving the method further for perfectly aligned images and developing its online variants.

## REFERENCES

[1] R. Vidal, Y. Ma, and S. Sastry, "Generalized principal component analysis (GPCA)," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 27, no. 12, pp. 1945–1959, Dec. 2005.

[2] S. Rao, R. Tron, R. Vidal, and Y. Ma, "Motion segmentation in the presence of outlying, incomplete, or corrupted trajectories," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 32, no. 10, pp. 1832–1845, Oct. 2010.

[3] C. Lu, H. Min, Z. Zhao, L. Zhu, D. Huang, and S. Yan, "Robust and efficient subspace segmentation via least squares regression," in *Proc. Eur. Conf. Computer Vision (ECCV)*, 2012, pp. 347–360.

[4] G. Liu, Z. Lin, S. Yan, J. Sun, Y. Yu, and Y. Ma, "Robust recovery of subspace structures by low-rank representation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 35, no. 1, pp. 171–184, Jan. 2013.

[5] E. Elhamifar and R. Vidal, "Sparse subspace clustering: Algorithm, theory, and applications," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 35, no. 11, pp. 2765–2781, Nov. 2013.

[6] K. Tang, R. Liu, Z. Su, and J. Zhang, "Structure-constrained low-rank representation," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 25, no. 12, pp. 2167–2179, Dec. 2014.

[7] R. Heckel and H. Bölcskei, "Robust subspace clustering via thresholding," *IEEE Trans. Inf. Theory*, vol. 61, no. 11, pp. 6320–6342, Nov. 2015.

[8] C. Zhang, H. Fu, S. Liu, G. Liu, and X. Cao, "Low-rank tensor constrained multiview subspace clustering," in *Proc. IEEE Int. Conf. Computer Vision (ICCV)*, 2015, pp. 1582–1590.

[9] Y. Xie, D. Tao, W. Zhang, Y. Liu, L. Zhang, and Y. Qu, "On unifying multi-view self-representations for clustering by tensor multi-rank minimization," *Int. J. Comput. Vis.*, 2018.

[10] L. De Lathauwer and B. De Moor, "From matrix to tensor: Multilinear algebra and signal processing," in *Proc. 4th IMA Int. Conf. Mathematics in Signal Processing*, 1996, pp. 1–15.

[11] K. Braman, "Third-order tensors as linear operators on a space of matrices," *Linear Algebra Appl.*, vol. 433, no. 7, pp. 1241–1253, Dec. 2010.

[12] M. E. Kilmer and C. D. Martin, "Factorization strategies for third-order tensors," *Linear Algebra Appl.*, vol. 435, no. 3, pp. 641–658, Aug. 2011.

[13] M. E. Kilmer, K. Braman, N. Hao, and R. C. Hoover, "Third-order tensors as operators on matrices: A theoretical and computational framework with applications in imaging," *SIAM J. Matrix Anal. Appl.*, vol. 34, no. 1, pp. 148–172, Feb. 2013.

[14] N. D. Sidiropoulos, L. De Lathauwer, X. Fu, K. Huang, E. E. Papalexakis, and C. Faloutsos, "Tensor decomposition for signal processing and machine learning," *IEEE Trans. Signal Process.*, vol. 65, no. 13, pp. 3551–3582, Jul. 2017.

[15] X. He, D. Cai, and P. Niyogi, "Tensor subspace analysis," in *Advances in Neural Information Processing Systems (NIPS)*, 2005, pp. 499–506.

[16] N. Hao, M. E. Kilmer, K. Braman, and R. C. Hoover, "Facial recognition using tensor-tensor decompositions," *SIAM J. Imaging Sci.*, vol. 6, no. 1, pp. 437–463, Feb. 2013.

[17] Z. Zhang, G. Ely, S. Aeron, N. Hao, and M. Kilmer, "Novel methods for multilinear data completion and de-noising based on tensor-SVD," in *Proc. IEEE Conf. Computer Vision and Pattern Recognition (CVPR)*, 2014, pp. 3842–3849.

[18] E. Kernfeld, S. Aeron, and M. Kilmer, "Clustering multi-way data: A novel algebraic approach," *arXiv preprint*, 2014. [Online]. Available: https://arxiv.org/abs/1412.7056

[19] S. Aeron and E. Kernfeld, "Group-invariant subspace clustering," in *Proc. Allerton Conf. Communication, Control, and Computing*, 2015, pp. 666–671.

[20] X. Piao, Y. Hu, J. Gao, Y. Sun, Z. Lin, and B. Yin, "Tensor sparse and low-rank based submodule clustering method for multi-way data," *arXiv preprint*, 2016. [Online]. Available: https://arxiv.org/abs/1601.00149

[21] C. Lu, J. Feng, Y. Chen, W. Liu, Z. Lin, and S. Yan, "Tensor robust principal component analysis: Exact recovery of corrupted low-rank tensors via convex optimization," in *Proc. IEEE Conf. Computer Vision and Pattern Recognition (CVPR)*, 2016, pp. 5249–5257.

[22] J. Zhang, X. Li, P. Jing, J. Liu, and Y. Su, "Low-rank regularized heterogeneous tensor decomposition for subspace clustering," *IEEE Signal Process. Lett.*, vol. 25, no. 3, pp. 333–337, Mar. 2018.

[23] S. Sun, "A survey of multi-view machine learning," *Neural Comput. Appl.*, vol. 23, no. 7-8, pp. 2031–2038, Dec. 2013.

[24] A. Y. Ng, M. I. Jordan, and Y. Weiss, "On spectral clustering: Analysis and an algorithm," in *Advances in Neural Information Processing Systems (NIPS)*, 2001, pp. 849–856.

[25] J. Nocedal and S. J. Wright, *Numerical Optimization*, 2nd ed. Springer, 2006.

[26] Z. Lin, M. Chen, and Y. Ma, "The augmented Lagrange multiplier method for exact recovery of corrupted low-rank matrices," University of Illinois Urbana-Champaign, Tech. Rep. UILU-ENG-09-2215, 2009.

[27] W. W. Hager and H. Zhang, "Inexact alternating direction multiplier methods for separable convex optimization," *arXiv preprint*, 2016. [Online]. Available: https://arxiv.org/abs/1604.02494

[28] Y. Lecun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-based learning applied to document recognition," *Proc. IEEE*, vol. 86, no. 11, pp. 2278–2324, Nov. 1998.

[29] T. Hastie and P. Y. Simard, "Metrics and models for handwritten character recognition," *Stat. Sci.*, vol. 13, no. 1, pp. 54–65, Feb. 1998.