IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received November 23, 2021, accepted December 13, 2021, date of publication December 23, 2021,
date of current version January 19, 2022.

Digital Object Identifier 10.1109/ACCESS.2021.3137869

A Comprehensive Tensor Framework for the Clustering of Hyperspectral Paper Data With an Application to Forensic Document Analysis

JOBIN FRANCIS[1], BABURAJ MADATHIL[2], SUDHISH N. GEORGE[1], (Member, IEEE), AND SONY GEORGE[3]

[1]Department of Electronics and Communication Engineering, National Institute of Technology Calicut, Calicut 673601, India
[2]Department of Electronics and Instrumentation Engineering, Government Engineering College Kozhikode, Calicut 673005, India
[3]Department of Computer Science, Norwegian University of Science and Technology, 2815 Gjøvik, Norway

Corresponding author: Sony George (sony.george@ntnu.no)

ABSTRACT In forensic document analysis, the authenticity of a document must be properly checked in the context of suspected forgery. Hyperspectral Imaging (HSI) is a non-invasive way of detecting fraudulent papers in a multipage document. The occurrence of a forged paper in a multi-page document may have a substantial difference from rest of the papers in its age, type, color, texture, and so on. Each pixel in an HSI data can be used as the material fingerprint for the spatial point it corresponds to. Hence, hyperspectral data of paper samples made of the same substance have similar characteristics and can be grouped into a single cluster. Similarly, paper samples made of different substances have different spectral properties. This paper relies on this heuristic and proposes a tensor based clustering framework for hyperspectral paper data, with an application to detect the forged papers in multi-page documents. Information embedded in the hyperspectral patches of the papers to be clustered is arranged into individual lateral slices of a third-order tensor in this framework. Further, this work employs the self-expressiveness property of submodules and an objective function is formulated to extract self-expressive representation tensor with low multirank and f-diagonal structure. Objective function of the proposed method incorporates $l_{\frac{1}{2}}$-induced Tensor Nuclear Norm (TNN) and $l_{\frac{1}{2}}$ regularization to impart better low rankness and f-diagonal structure to the representation tensor. Experimental results of the proposed method were compared to the state-of-the-art subspace clustering approaches. The results demonstrate improved performance of the proposed method over the existing clustering algorithms.

INDEX TERMS Forensic document analysis, hyperspectral imaging (HSI), clustering, self-expressiveness property.

I. INTRODUCTION

Hyperspectral sensors generate the hyperspectral image (HSI) of a spatial scene in hundreds of spectral bands [1]. Usually, an HSI data is characterized by a large number of images captured at different wavelengths [2]. In the HSI data of a given object/material, the information embedded in various spectral bands is varied; as the material absorbs or reflects differently at different wavelengths [3]. Since the HSI data contains a plethora of information both in its spectral and spatial bands, it has got a lot of interest in sectors such as geology, remote sensing [4], agriculture [5], forensic research [6], and so on. HSI techniques were previously used mainly in various fields of satellite imaging and remote sensing [7]. Later, HSI approaches have been widely employed in a number of applications such as food quality inspection [2], medical imaging [8], forensic analysis, and material science [9]. Furthermore, when compared to other invasive methods, HSI methodologies have considered to be a good candidate for non-invasive analysis in characterizing the material properties [6], [10].

An HSI data is typically represented by a three dimensional data cube with two spatial domains and one spectral domain [4], [5]. The spectral information are stacked along the third dimension [5]. Hence, a single HSI pixel can be viewed as an N dimensional vector, $\mathbf{p} \in \mathbb{R}^N$, where N denotes

The associate editor coordinating the review of this manuscript and approving it for publication was Jad Nasreddine.

the total number of spectral bands [5]. In general, each point or a pixel of an HSI data characterizes a given material with its spectrum at that point. As a result, HSI pixel information can be employed as a material fingerprint for the object/material under consideration, which can be used to differentiate the characteristics of different materials [6].

As a non-destructive contactless method and having enriched with the information from numerous hyperspectral bands, HSI methods have got a substantial recognition in the field of forensic document analysis [6]. In recent days, there have been a considerable number of cases reported on fraudulent manipulation of legal documents. In the context of a suspected forgery, the authenticity of the document need to be examined. This is normally done by examining both the ink and the paper that have been used to create the suspected document [6], [9]. Suppose, the inclusion of distinct paper type is identified in a multipage document, and which may be suspected of being forged. Hyperspectral analysis can be effectively used to detect the fraudulent paper/papers in the above described scenario. A hyperspectral paper data is the one that has been subjected to the HSI imaging technique [6]. Consider the hyperspectral data of a single paper, and assume that the entire portion that paper is composed of the same material. Then, the spectrum that corresponds to different sample areas of hyperspectral paper data would share similar spectral properties, unless or otherwise any portion of the paper under consideration is forged [6], [10]. To put it another way, the hyperspectral data of paper samples made of the same substance can have similar characteristics, and these samples may be grouped into a single cluster [4]. On the other hand, if paper samples considered are made of different material, the spectrum of these samples exhibit different spectral properties, and those samples may fall into different clusters [6]. The occurrence of a forged item, whether one or many, in a multipage document may have differed from the rest of the papers due to a variety of factors such as material, type, age, texture, color, and so on [6]. HSI methods, in comparison to three channel imaging techniques, can better detect and evaluate different objects by analyzing their spectral fingerprints through a wider spectrum [11]. As a result, the challenge of document analysis in a suspected forgery can be effectively addressed by the clustering analysis of hyperspectral paper data samples [6], [10].

II. RELATED WORKS

Because of the large dimensionality of hyperspectral data, supervised classification with prior labelling of the data would be difficult to implement [12]. Unsupervised clustering methods, which do not require prior labelling but partition datapoints based on their inherent similarity, are used to overcome this challenge [12]. In the literature on hyperspectral data clustering, many techniques have been identified, that adapt conventional unsupervised clustering methods. In remote sensing applications, methods such as fuzzy c-means [13], k-means [14], and spectral clustering [15] have been used to classify pixels for the clustering of hyperspectral images [16]. However, due to the large dimensionality of hyperspectral data, limited performance of them were observed. Methods such as PCA [17], ICA [18], and LDA [19] were used in certain works to reduce the dimensionality of hyperspectral data [6], [20]. However in some works, t-Stochastic Neighbouring Embedding (t-SNE) has been used in studies such as clustering of hyperspectral paper data and hyperspectral ink data, and has surpassed the aforementioned methods [6], [9].

In recent decades, principles of subspace modeling have been applied in the field of hyperspectral image clustering [12]. Due to the multidimensionality and highly correlated information contained in neighbouring spectral bands, hyperspectral data can be structured as multiple low dimensional subspaces embedded within a large dimensional space [12]. This assumption is based on the fact that, HSI data often contains large homogeneous regions, and the pixels inside those regions can have similar spectral properties [1]. With the advancement in Sparse Representation (SR) and Low Rank Representation (LRR) models, many works have been proposed to meet the challenge of HSI clustering, in recent decades [21]. Because of its robustness, Sparse Subspace Clustering (SSC) was used in HSI clustering [22]. Exploiting the abundance of spatial information and high spectral correlation, Zhang *et al.* proposed a spectral-spatial SSC (S^4C) for the effective grouping of HSI data [4]. Similarly. an l_2 norm regularized SSC proposed by Zhai *et al.* was employed in the field of hyperspectral remote sensing imagery [16], [23]. Then, Whang *et al.* introduced a Fast High Order SSC (FHoSSC) with Cumulative Markov Random Field (MRF) for subspace segmentation that took advantage of superpixels [16]. A number of Low Rank Subspace Clustering (LRSC) approaches have been proposed for the clustering of hyperspectral imagery, based on LRR techniques [12]. In addition, in works such as [24] and [25], hypergraphs have been used to obtain an accurate information about the manifold structure. Following that Xu *et al.* proposed a superpixel based LRSC that leverages hypergraphs for the classification of hyperspectral images [12].

A number of works have recently been published that exploit the properties of higher order tensors by utilizing the multilinear algebra and abstract algebra [26]. Since the invention of the t-product proposed by Kilmer *et al.* [27], the multiplication of third order tensors has become substantially simplified. Some recent works introduced the concept of free submodules by assuming that datapoints from a large dimensional space are lying near a union of free submodules [28]. In Sparse and Low Rank Submodule Clustering (SLRSmC), Kernfeld *et al.* employed Union of Free Submodules (UoFS) model in their work, which is based on the self expressiveness property of free submodules [29]. Relying on this self expressive representation, Wu *et al.* proposed a Structure Constrained Low Rank Submodule Clustering (SCLRSmC) framework for clustering of 2D images [26]. In SCLRSmC, images to be clustered are stacked into the lateral slices of a third order data tensor. Further, the data tensor has been

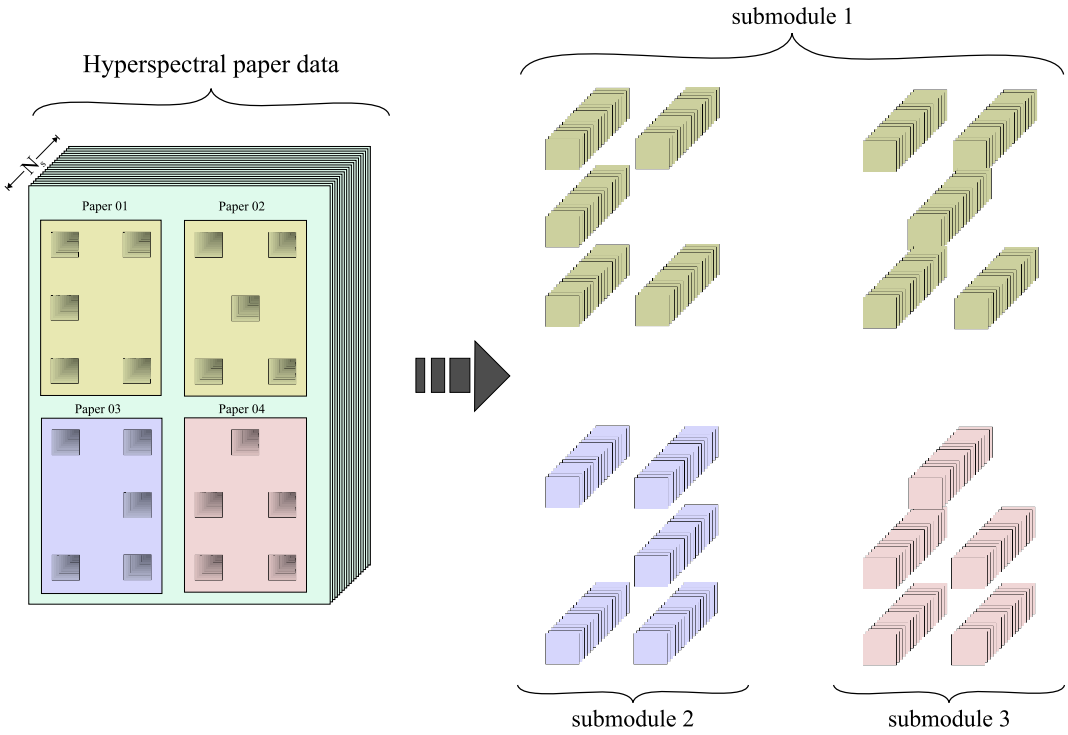


FIGURE 1. The complete overview of fetching 3D patches/samples from the hyperspectral paper data. The papers under consideration are arranged in a tiled format. For the ease of illustration, only four papers are considered in this figure. In this, three different types of papers are shown among which paper 1 and paper 2 are of same type. The FIGURE at the right side shows hyperspectral data samples fetched from different portions. It is assumed that samples with different spectra would lie in different submodules. N_s represents number of spectral bands.

modeled as a t-product of the data tensor and a structured low rank coefficient tensor, based on the self expressive representation [26].

According to the past literature on HSI data clustering, it is identified that the capabilities of tensor frameworks have not been properly exploited in majority of the existing works. Moreover, due to the multiple spectral bands present in hyperspectral data, most of the proposed works suffer from increased computational complexity [30]. Tensor based frameworks, on the other hand, are relatively easy to implement, and computations can be simplified using multilinear algebra tools [26]. In addition, depending on the problem and its needs, datapoints can be arranged into appropriate slices of a tensor in different orientations [31]. Since, the diverse information of a hyperspectral data is embedded in its hundreds of spectral bands, a tensor based framework is a good candidate for the execution of hyperspectral data clustering. Above all, a comprehensive framework for clustering hyperspectral paper data that fully explores the capabilities of a third order tensor space has not yet to been properly employed. Based on the aspects aforementioned, it is evident that there is still room for a tensor based clustering framework that could effectively address the challenge of clustering of hyperspectral paper data.

In this work, we assume that if every single paper in a multipage document is made of the same material, then attributes such as texture, color, and age would also be identical throughout the entire portion of that paper under consideration [6]. Based on the preceding assumption, it is obvious that the hyperspectral samples of a paper data fetched from different parts of a single paper will have similar spectral signatures and those samples will be lying in a single submodule [6], [26]. Hence, we further believe that if all of the pages in a multipage document are made of the same material, then hyperspectral data samples taken from various parts of all those papers in that document would almost undoubtedly be found in a single submodule. Therefore, a forged paper can be easily identified because the hyperspectral data samples of the forged item will be placed in a different submodule. We plotted the normalized reflectance spectrum of hyperspectral data of various papers presented in FIGURE 2 (a), to underline the above-mentioned aspect. It has been observed that the reflectance spectra of those papers differ significantly from one another and are easily distinguishable (FIGURE 2 (b)). Hence, there will be a high probability for those samples from various papers to be lying in different clusters/submodules. We also plotted the reflectance spectra of a number of specimens collected from the same paper. We used five hyperspectral samples with dimensions $\in \mathbb{R}^{S_1 \times S_3 \times N_s}$, ($S_1 = 10$, $S_3 = 10$, $N_s = 186$) from Paper 1 and Paper 12, shown in FIGURE 2 (a) for this. FIGURE 2 (c) shows that the spectra of the samples taken from same paper (say, Paper 1 or Paper 12) are overlaid on one another. The indistinguishable spectra of those samples from the same paper are very likely to belong to a single submodule.

On the basis of this heuristic described above, we propose a submodule clustering framework for hyperspectral paper data by embedding each hyperspectral data sample in a third order tensor space. An overview of the proposed framework is given in FIGURE 1 with an illustration of hyperspectral data of four papers arranged in a tiled format. Amongst, Paper 1 and Paper 2 are assumed to be composed of the same material. Furthermore, the fetched hyperspectral 3D patches/blocks $\in \mathbb{R}^{S_1 \times S_3 \times N_s}$ and its different submodules are also shown in FIGURE 1. Now, we list the major contributions of the proposed method,

1) We develop a comprehensive tensor based framework for the clustering hyperspectral paper data. First, the 3D patches $\in \mathbb{R}^{S_1 \times S_3 \times N_s}$ from hyperspectral paper data are arranged into lateral slices of a third order tensor, termed input data tensor in this framework. Employing t-product, the third order input data tensor is then self-expressively represented as the product of the input data tensor itself and a low rank structured coefficient tensor.
2) In a hyperspectral paper data, we use the heuristic that the three dimensional paper patches fetched from different locations of the same paper may have stronger correlations, whereas patches taken from distinct papers may exhibit lower correlations. In the proposed method, we incorporate this heuristic by employing a dissimilarity matrix which can clearly capture the different correlation exist in between the hyperspectral paper data samples/patches.
3) To the best of our knowledge, the aforementioned methodology is the primarily one that incorporate a complete tensor based framework for the clustering hyperspectral paper data.

III. PRELIMINARIES AND TECHNICAL BACKGROUND

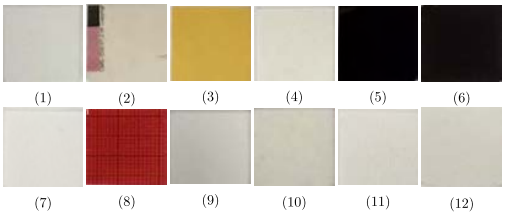
This section illustrates notations and mathematical concepts used in our paper. Further, important terms, its notations and corresponding descriptions are given in TABLE 1. Following that, the mathematical preliminaries and their expressions are presented.

A. TENSOR-PRODUCT (T-PRODUCT) [27]

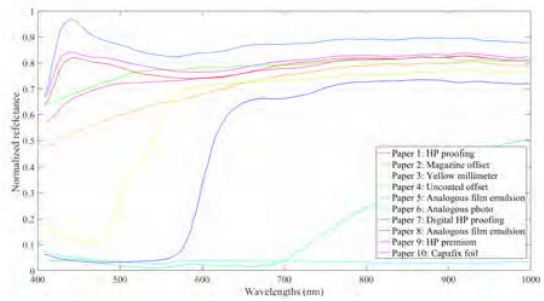
For tensors $\mathcal{X} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$ and $\mathcal{Y} \in \mathbb{R}^{n_2 \times n_4 \times n_3}$, the t-product $\mathcal{X} * \mathcal{Y}$ will be a tensor $\mathcal{Z}$ of order $n_1 \times n_4 \times n_3$ and '$*$' denotes the tensor product [26]. The $(l, m)^{th}$ tube of $\mathcal{Z}$ is given by $\mathcal{Z}(l, m, :) = \sum_{p=1}^{n_2} \mathcal{X}(l, p, :) \circ \mathcal{Y}(p, m, :)$, where $l = 1, 2, \ldots, n_1$ and $m = 1, 2, \ldots, n_4$, and '$\circ$' is the circular convolution operator [28].

B. TWIST OPERATOR [28]

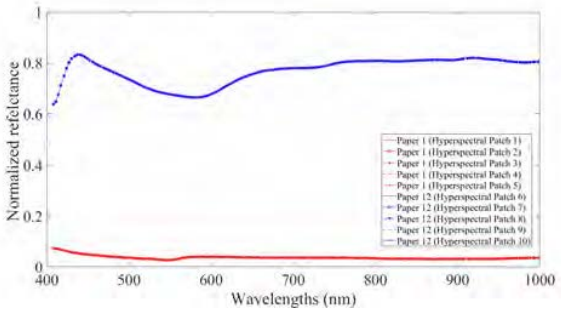
Let $\mathcal{X} \in \mathbb{R}^{n_1 \times N \times n_3}$, then $\overrightarrow{\mathcal{X}} \in \mathbb{R}^{n_1 \times n_3 \times N}$ is the twisted version of $\mathcal{X}$. Then, the lateral slices of $\mathcal{X}$ becomes the frontal slices of $\overrightarrow{\mathcal{X}}$. It is given by, $\overrightarrow{\mathcal{X}}(:, :, i) = twist(\mathcal{X}(:, i, :))$.



(a) (1) HP proofing, (2) magazine offset, (3) Yellow millimeter, (4) uncoated offset, (5) analog film emulsion, (6) analogous photo, (7) digital HP proofing, (8) analogous film emulsion, (9) HP premium, (10) capafix foil, (11) HP canvas, (12) coated offset.



(b) Normalized reflectance spectrum.



(c) Normalized reflectance spectrum.

FIGURE 2. (a) Types of papers. (b) Spectra that correspond to each hyperspectral patch/sample $\in \mathbb{R}^{S_1 \times S_3 \times N_S}$ fetched from Paper 1 to Paper 10. $S_1 \times S_3$: spatial dimension, N_S: Number of spectral bands. (c) Spectra that correspond to five samples each from Paper 1 and Paper 12.

C. HALF THRESHOLDING FUNCTION [32]

The half thresholding function for a vector, $\mathbf{z} = (z_1, z_2, \ldots z_N) \in \mathbb{R}^N$ is given by [33],

$$h_{\lambda, \frac{1}{2}}(z_i) = \begin{cases} \frac{2}{3} z_i \left(1 + cos\left(\frac{2\pi}{3} - \frac{2}{3}\Psi_\lambda(z_i)\right)\right), \\ \qquad |z_i| > \frac{\sqrt[3]{54}}{4}(\lambda)^{\frac{2}{3}} \\ 0, \quad \text{otherwise} \end{cases} \tag{1}$$

where, $\Psi_\lambda(z_i) = arccos(\frac{\lambda}{8}(\frac{|z_i|}{3})^{-\frac{3}{2}})$ and λ denotes the regularization parameter [33].

D. TENSOR MULTIRANK [26]

The multirank of a tensor, $\mathcal{X} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$ is a vector, $\mathbf{p} \in \mathbb{R}^{n_3}$ with the k^{th} element is equal to the rank of the k^{th} frontal slice of $\widehat{\mathcal{X}}$, where $\widehat{\mathcal{X}}$ represents the Fourier transform of $\mathcal{X}$. Then,

TABLE 1. TABLE represents different terms and its notations used in this paper.

Notation	Description	Notation	Description
a, b	Scalars	$\|\mathbf{A}\|_*$	Matrix Nuclear Norm
$\mathbf{x}, \mathbf{y}$	Vectors	$\|\mathbf{X}\|_1$	l_1 norm of matrix $\mathbf{X}$
$\mathbf{M}, \mathbf{P}$	Matrices	$\|\mathbf{X}\|_{\frac{1}{2}}^{\frac{1}{2}}$	$l_{\frac{1}{2}}$ norm of matrix $\mathbf{X}$
$\mathcal{X}, \mathcal{Q}, \mathcal{Z}$	Tensors	$\|\mathcal{Z}\|_{\circledast}$	Tensor Nuclear Norm (TNN)
$\mathcal{X}(:,l,:)$	l^{th} lateral slice of $\mathcal{X}$	$\|\mathcal{Z}\|_{\circledast_{\frac{1}{2}}}$	$l_{\frac{1}{2}}$-induced TNN
$\mathcal{X}(:,:,k)$ or $\mathcal{X}^{(k)}$	k^{th} frontal slice of $\mathcal{X}$	$\mathcal{U} * \Sigma * \mathcal{V}^T$	t-SVD representation
$\mathcal{X}(:,i,j)$, $\mathcal{X}(i,:,j)$	mode-1, mode-2 fibers	$*$	t-product
$\mathcal{X}(i,j,:)$	mode-3/tube fibers	$x_{i,j,k}$	$(i,j,k)^{th}$ element of $\mathcal{X}$

$\widehat{\mathcal{X}} = \text{fft}(\mathcal{X}, 3)$ denotes the DFT along the third dimension for the tensor, $\mathcal{X}$ [28].

E. TENSOR-SINGULAR VALUE DECOMPOSITION (T-SVD) [26]

The t-SVD of a third order tensor $\mathcal{X} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$ is given by, $\mathcal{X} = \mathcal{U} * \Sigma * \mathcal{V}^T$ 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 the orthogonal tensors. Then, $\Sigma \in \mathbb{R}^{n_1 \times n_2 \times n_3}$ is an f-diagonal tensor, where its frontal slices contain diagonal matrices [26]. The t-SVD of tensor, $\mathcal{X}$ can be found out using the SVDs of frontal slices of its Fourier tensor, $\widehat{\mathcal{X}}$ [26]. For the frontal slice $\widehat{\mathcal{X}}^{(k)}$ of the Fourier tensor $\widehat{\mathcal{X}}$, the SVD is given by, $\widehat{\mathcal{U}}(:,:,k) * \widehat{\Sigma}(:,:,k) * \widehat{\mathcal{V}}(:,:,k)$, for $k = 1, 2, \ldots, n_3$. Then, $\mathcal{U}$ =ifft($\widehat{\mathcal{U}}$, 3), Σ =ifft($\widehat{\Sigma}$, 3), $\mathcal{V}$ =ifft($\widehat{\mathcal{V}}$, 3), where ifft($\widehat{\mathcal{U}}$, 3) denotes the inverse DFT along mode-3 of $\widehat{\mathcal{U}}$ [26].

F. $l_{\frac{1}{2}}$-INDUCED TENSOR NUCLEAR NORM (TNN)

Consider a tensor, $\mathcal{X} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$ with t-SVD $\mathcal{X} = \mathcal{U} * \Sigma * \mathcal{V}^T$, then its $l_{\frac{1}{2}}$-induced TNN can be expressed as,

$$\|\mathcal{X}\|_{\circledast_{\frac{1}{2}}} = \sum_{k=1}^{n_3} \sum_{i=1}^{min(n_1,n_2)} \sqrt{|\widehat{\Sigma}(i,i,k)|} \tag{2}$$

The solution for $l_{\frac{1}{2}}$-induced TNN can be deduced from a number of steps. Each frontal slice, $\widehat{\Sigma}^{(k)}$ consists of $s_1 \geq s_2 \ldots \geq s_N \geq 0$ singular values at its diagonal positions and those values can be represented by a vector $\mathbf{s}$, where, $\mathbf{s} = (s_1, s_2, \ldots, s_N) \in \mathbb{R}^N$. First, apply half thresholding function stated in *Eq:* (1) onto each members of $\mathbf{s} \in \mathbb{R}^N$. This is accomplished with the half thresholding operator $H_{\lambda,\frac{1}{2}}(.)$, as proposed in [32]. The half thresholding operator $H_{\lambda,\frac{1}{2}}(.)$ is a non-linear mapping function and for any vector, $\mathbf{s} = (s_1, s_2, \ldots, s_N)$, it can be expressed as,

$$H_{\lambda,\frac{1}{2}}(\mathbf{s}) = (h_{\lambda,\frac{1}{2}}(s_1), h_{\lambda,\frac{1}{2}}(s_2), \ldots h_{\lambda,\frac{1}{2}}(s_N))^{th} \tag{3}$$

where, 'th' represents the threshold [32], [33]. Then, the entire procedure to find the $l_{\frac{1}{2}}$-induced TNN can be summarized in **Algorithm 1**.

Algorithm 1 Algorithm for $l_{\frac{1}{2}}$-Induced TNN

Require: $\mathcal{X} \in \mathbb{R}^{n_1 \times n_2 \times n_3}, \lambda > 0, \mu > 0, th > 0$
Ensure: $\mathcal{X}_{ht} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$
1: $\widehat{\mathcal{X}} = \text{fft}(\mathcal{X}, 3)$
2: **for** i= 1 to n_3 **do**
3: $[U, \Sigma, V] = \text{svd}(\widehat{\mathcal{X}}^{(i)})$
4: $\widehat{\mathcal{U}}^{(i)} = \mathbf{U}, \widehat{\Sigma}^{(i)} = \Sigma, \widehat{\mathcal{V}}^{(i)} = \mathbf{V}$
5: $\mathbf{s} = \text{diag}(\widehat{\Sigma}^{(i)})$
6: $H_{\lambda,\frac{1}{2}}(\mathbf{s}) = (h_{\lambda,\frac{1}{2}}(s_1), h_{\lambda,\frac{1}{2}}(s_2), \ldots h_{\lambda,\frac{1}{2}}(s_N))^{th}$
7: $\widehat{\Sigma}_{hf}(:,:,i) = \text{diag}(H_{\lambda,\frac{1}{2}}(\mathbf{s}))$
8: **end for**
9: $\mathcal{U} = \text{ifft}(\widehat{\mathcal{U}}, 3), \mathcal{H}_{\frac{1}{2}}(\Sigma_t) = \text{ifft}(\widehat{\Sigma}_{hf}, 3), \mathcal{V} = \text{ifft}(\widehat{\mathcal{V}}, 3)$
10: $\mathcal{X}_{ht} = \mathcal{U} * \mathcal{H}_{\frac{1}{2}}(\Sigma_t) * \mathcal{V}^T$

IV. PROPOSED METHOD

This session begins with an illustration of the self-expressiveness property of free submodules [26]. According to the self-expressive representation, a member or datapoint in a submodule can be represented as the t-linear combination of other members in the same submodule. To put it in another way, consider $\mathbb{K}_{n_3}$ to be a set of tube fibers belongs to $\mathbb{R}^{1 \times 1 \times n_3}$ that forms a commutative ring under regular addition and t-product [26]. The self-expressive representation of submodules can be expressed as the t-linear combination of oriented matrices and mode-3 fibers by making use of the multilinear algebra. An oriented matrix with dimensions of $n_1 \times 1 \times n_3$, can be formed from a matrix of size $n_1 \times n_3$ by twisting it perpendicular to a page [29], [34]. Let $\mathbb{K}_{n_3}^{n_1}$ denote the set of $n_1 \times 1 \times n_3$ oriented matrices. Further, a single oriented matrix can also be viewed as a one dimensional vector with dimension n_1 and the elements becomes $1 \times 1 \times n_3$ tube fibers [26]. The set of oriented matrices can then be considered as an n_1 dimensional free module over the ring $\mathbb{K}_{n_3}$ [26], [29]. Consider a generating set $\{\overrightarrow{\mathcal{D}}_i \in \mathbb{K}_{n_3}^{n_1}\}_{i=1}^{n_1}$, and any element $\overrightarrow{\mathcal{X}} \in \mathbb{K}_{n_3}^{n_1}$ can be uniquely represented as a t-linear combination of the $\overrightarrow{\mathcal{D}}_i$s [26]. Mathematically, it can be represented as,

$$\overrightarrow{\mathcal{X}} = \sum_{i=1}^{n_1} \overrightarrow{\mathcal{D}}_i * \overrightarrow{z}_i \tag{4}$$

where, $\overrightarrow{z}_i \in \mathbb{K}_{n_3}$ [26]. This is the same as generalizing vector space over a field [26]. Hence, with t-product, linear combination for submodules can be executed with the corresponding coefficients as mode-3 or tube fibers [29], [34]. Hence, consider $\mathbb{F}_{n_3}^{n_1}$, a free submodule which is a subset of the module $\mathbb{K}_{n_3}^{n_1}$. Then, consider a set of L free submodules, $\{^l\mathbb{F}_{n_3}^{n_1}\}_{l=1}^{L}$ and any element corresponding to a free submodule $\mathbb{F}$ can be represented as t-linear combination of elements in the union of L free submodules [26]. In such a representation, non-zero tube fibers represent the coefficients correspond to other elements belong to the same submodule and the zero tube fibers represent, coefficients correspond to elements from other free submodule [26]. Circular convolution is defined within t-product in the spatial domain. However, it can be replaced and simplified by multiplication in the Fourier domain using Discrete Fourier Transform (DFT) [26].

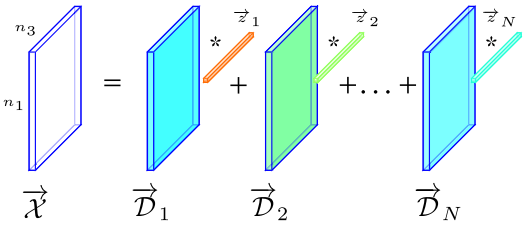


FIGURE 3. A member of a submodule is represented as t-linear combination of other elements in the same submodule [26].

A. REPRESENTATION OF HYPERSPECTRAL PAPER DATA INTO A THIRD ORDER TENSOR SPACE

Let $\mathcal{X}_i \in \mathbf{R}^{S_1 \times S_3 \times N_s}$, where $i = 1, 2, \ldots, N$ represent a set of hyperspectral pixels, that has been sliced from a hyperspectral paper data. In the above depiction, $S_1 \times S_3$ represents the spatial dimension of the hyperspectral patch, and N_s represents the number of spectral bands. Then, N denotes the total number of samples taken from different pages from a multipage document. Further, consider a matrix $\mathbf{B}_i \in \mathbb{R}^{S_1 S_3 \times N_s}$, where $i = 1, 2, \ldots, N$ which is to encapsulate the spectral information of each hyperspectral data sample $\mathcal{X}_i \in \mathbf{R}^{S_1 \times S_3 \times N_s}$. The information contained in each spectral sample of dimension $S_1 \times S_3$ of $\mathcal{X}_i \in \mathbf{R}^{S_1 \times S_3 \times N_s}$ is then encoded into the respective columns of the matrix $\mathbf{B}_i \in \mathbb{R}^{S_1 S_3 \times N_s}$. In other words, each spectral sample of dimension $\mathbb{R}^{S_1 \times S_3}$ will be reformed into a column vector with dimension $\mathbb{R}^{S_1 S_3}$, and there will be N_s such column vectors for a single hyperspectral patch, $\mathcal{X}_i \in \mathbf{R}^{S_1 \times S_3 \times N_s}$. Then, the matrix $\mathbf{B}_i$ consists of all the information contained in a hyperspectral patch or sliced from a hyperspectral paper data. The entire process described above is then repeated for each of the N hyperspectral data samples to be clustered.

Consider a third order tensor $\mathcal{X} \in \mathbf{R}^{S_1 S_3 \times N \times N_s}$, to integrate the information contained in each hyperspectral data sample, $\mathcal{X}_i \in \mathbf{R}^{S_1 \times S_3 \times N_s}$, where $i = 1, 2, \ldots N$, into a third order tensor space. Consequently, by the process of twisting, each matrix $\mathbf{B}_i$ can be converted into an oriented matrix, $\overrightarrow{\mathbf{B}}_i \in \mathbb{R}^{S_1 S_3 \times 1 \times N_s}$ as illustrated in FIGURE 4. The aforementioned process is repeated for N hyperspectral data samples, a collection of $\{\overrightarrow{\mathbf{B}}_i\}_{i=1}^{N}$ can be made. The set of oriented matrices is then organized into the lateral slices of a three-dimensional tensor $\mathcal{X} \in \mathbf{R}^{S_1 S_3 \times N \times N_s}$, and this process is illustrated in FIGURE 5. By this process, each lateral slice, $\mathcal{X}(:, i, :) \in \mathbb{R}^{S_1 S_3 \times 1 \times N_s}$ of the three dimensional data tensor, $\mathcal{X}$ holds the diverse information contained in respective hyperspectral data sample.

B. PROBLEM FORMULATION

To solve the clustering problem, the next step is to create a self-expressive representation for the hyperspectral paper data. For this task, we assume that the hyperspectral paper data placed in the lateral slices belong to a union of L free submodules [26]. In the self expressive representation, a datapoint belonging to a submodule can be expressed as a t-linear combination of other datapoints in the same submodule [26].

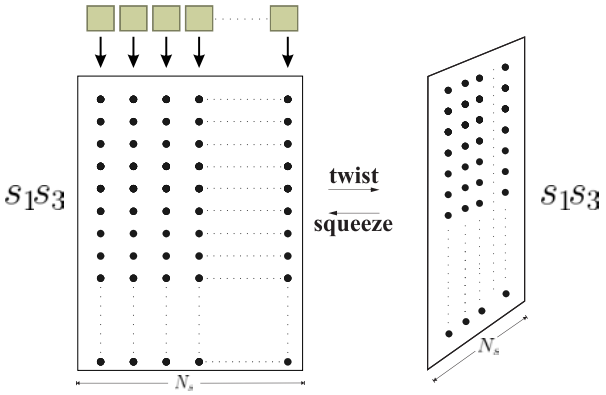


FIGURE 4. Formation of a lateral slice formed by the process of twist operation. FIGURE at the left: matrix $\mathbf{B}_i \in \mathbb{R}^{S_1 S_3 \times N_s}$, for $i = 1$ to N. FIGURE at the right: the twisted matrix. The big black dots represent intensity values.

On the other hand, a coefficient tensor $\mathcal{Z}$ will exist, where, the relationship between the data tensor $\mathcal{X}$ can be expressed in terms of t-product such that $\mathcal{X} = \mathcal{X} * \mathcal{Z}$ [26], [31], [35]. In the above representation, the data tensor, $\mathcal{X}$ itself acts as a dictionary. The coefficient tensor $\mathcal{Z}$ would have evolved with a low tensor multirank and an f-block diagonal structure for improved representation and reduced computational cost [26], [34]. Hence, the proposed method imposes low tensor multirank and a structure constraint on the coefficient tensor. Finding the tensor multirank, on the other side, is non-convex, thus its strong convex surrogate Tensor Nuclear Norm (TNN) is a good candidate. Then, TNN of $\mathcal{Z}$, denoted by $\|\mathcal{Z}\|_{\circledast}$ can be expressed as the sum of the singular values of all the frontal slices of $\widehat{\mathcal{Z}}$ [26]. Many recent methods such as [26], [31], [36] have employed TNN to impose the low tensor rank constraint on the representation tensor in their optimization problems.

In order to find the TNN of a particular tensor, l_1 norm is employed to compute the absolute sum of singular values of its frontal slices. The acceptance of l_1 minimization is due to the fact that it is convex and the sparser solution can be obtained with less computational bottleneck [32]. But in recent studies, it is observed that l_q $(0 < q < 1)$ regularization techniques provide more sparser solutions compared to l_1 norm [32], [33]. For a vector $\mathbf{x} \in \mathbb{R}^N$, l_q regularization problem from the observation $\mathbf{y} = \mathbf{A}\mathbf{x}$, can be represented as,

$$\min_{\mathbf{x} \in \mathbb{R}^N} \|\mathbf{A}\mathbf{x} - \mathbf{y}\|_2^2 + \lambda \|\mathbf{x}\|_q^q, \tag{5}$$

where, $\mathbf{y} \in \mathbb{R}^m$, $\mathbf{A} \in \mathbb{R}^{m \times N}$. Then, $\|\mathbf{x}\|_q$ represents the l_q quasi-norm and is defined by, $\|\mathbf{x}\|_q = (\sum_{i=1}^{N} |\mathbf{x}_i|^q)^{\frac{1}{q}}$. The unit ball representations of various norms are illustrated in FIGURE 6 in which l_2 norm has the spherical shape, whereas in l_1 norm, it is of diamond shaped. It is observed from FIGURE 6 (a) and FIGURE 6 (b) that l_1 regularization provides a sparser solution compared to l_2 norm. However, as the value of q is again reduced, the unit ball can assume the shape as shown in FIGURE 6 (c) in which there is higher probability for the $\mathbf{y} = \mathbf{A}\mathbf{x}$ line to coincide with axes. Hence,

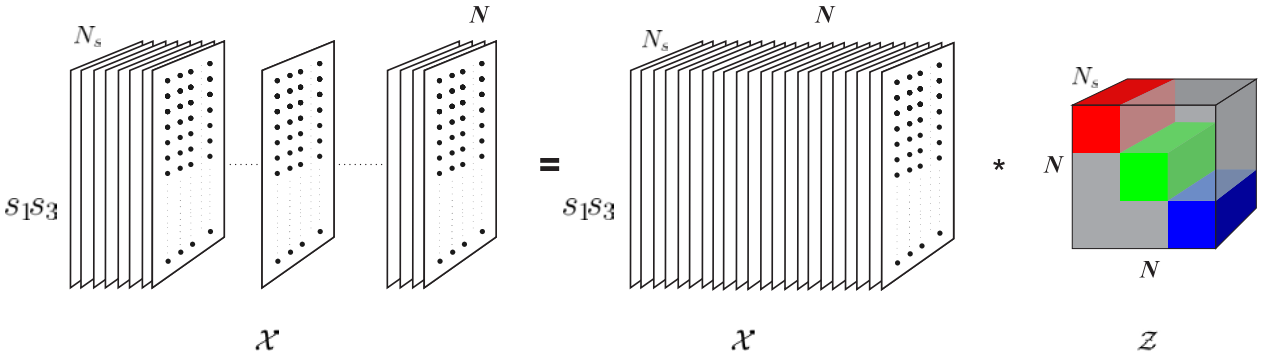


FIGURE 5. Self expressiveness representation ($\mathcal{X} = \mathcal{X}\mathcal{Z}$) of hyperspectral data using union of free submodules approach. Coefficient tensor $\mathcal{Z} \in \mathbb{R}^{N\times N\times N_s}$ is represented with low multirank and f-diagonal structure.

the probability of achieving sparser solution is higher as the value of q is changed from 0 to 1. For $q \in [\frac{1}{2}, 1)$, solution will be sparser for smaller value of q. But no significant change is observed in the performance for $q \in [0, \frac{1}{2})$ [32], [33], [37]. Hence, fixing the index $q = \frac{1}{2}$, $l_{\frac{1}{2}}$ regularization has been chosen as an improved regularization technique which yields more sparser solution than l_1 minimization [32], [33].

Moreover, iterative half thresholding algorithm proposed by Xu *et al.* provided a fast solution and convergence to the $l_{\frac{1}{2}}$ approach, despite the non-convex nature of $l_{\frac{1}{2}}$ norm [32]. Furthermore, within some constraints, Xu *et al.* had also verified the convergence of the half thresholding algorithm to a stationary point by a dynamic system methodology [33]. The strong sparsity inducing ability of $l_{\frac{1}{2}}$ regularization was successfully implemented in many sparsity problems [37]. Motivated from the aforementioned successful approaches and benefited from the strong theoretical background, to obtain more accurate tensor low rank representation, $l_{\frac{1}{2}}$-induced TNN is employed in the proposed method by replacing $l_{\frac{1}{2}}$ norm in place of l_1 norm in the expression of TNN. The formulation of $l_{\frac{1}{2}}$-induced TNN has already been detailed in Section III. Furthermore, an appropriate block diagonal structure for the representation tensor encourages clustering of multi-view data and improves clustering algorithm performance. Obviously, in the samples of the hyperspectral paper data, objects belonging to a single submodule have strong correlations, while objects belonging to distinct submodules have lower correlations [26]. Hence, the correlation between different datapoints in the hyperspectral data can be captured with a dissimilarity matrix, denoted by $\mathbf{P}_{DM} \in [0, 1]^{N\times N}$, where the entries of $\mathbf{P}_{DM}$ is given by [38],

$$\mathbf{P}_{DM\,i,j} = \frac{1 - \left(\frac{\sum_{i=1}^{N}(\mathbf{x}_i - \mu_{\mathbf{x}_i})(\mathbf{x}_j - \mu_{\mathbf{x}_j})}{\sqrt{(\mathbf{x}_i - \mu_{\mathbf{x}_i})^2} \times \sqrt{(\mathbf{x}_j - \mu_{\mathbf{x}_j})^2}}\right)}{2} \tag{6}$$

where, $i, j = \{1, 2, \ldots, N\}$ and $\left(\frac{\sum_{i=1}^{N}(\mathbf{x}_i - \mu_{\mathbf{x}_i})(\mathbf{x}_j - \mu_{\mathbf{x}_j})}{\sqrt{(\mathbf{x}_i - \mu_{\mathbf{x}_i})^2} \times \sqrt{(\mathbf{x}_j - \mu_{\mathbf{x}_j})^2}}\right)$ is the Pearson rank correlation coefficient represented by $r_{\mathbf{x}_i\mathbf{x}_j}$ [39]. Generally, Pearson rank correlation coefficient indicates a measure of linear relationship between the

(a) l_2 norm (b) l_1 norm (c) $l_{\frac{1}{2}}$ norm (d) l_0 norm

FIGURE 6. Unit ball representation of various norms in a three dimensional space $\mathbb{R}^3$.

two variables. Unlike Euclidean scores, the above metric shows how closely two variables are correlated. The value, $r_{\mathbf{x}_i\mathbf{x}_j} = +1$, indicates the positive correlation between the variables $\mathbf{x}_i$ and $\mathbf{x}_j$, whereas $r_{\mathbf{x}_i\mathbf{x}_j} = -1$ stands for a negative correlation. Further, $\mu_{\mathbf{x}_i}$ and $\mu_{\mathbf{x}_j}$ represent the sample mean of $\mathbf{x}_i$ and $\mathbf{x}_j$. In the above expression, $\mathbf{x}_i = (\text{vec}(\mathcal{X}(:, i, :))$ and $\mathbf{x}_j = (\text{vec}(\mathcal{X}(:, j, :))$, where, vec(.) indicates the vectorization of the lateral slices into a one dimensional vector [39]. The proposed method integrates the following aspects into its optimization problem.

1) The proposed method incorporates $l_{\frac{1}{2}}$-induced TNN to impart better low rankness on the representation tensor $\mathcal{Z}$.
2) Integrating the dissimilarity matrix within the proposed method depicts the higher correlations occur between members of the same submodule and lower correlations for those exist in distinct submodules. Furthermore, it aids in the better capturing of f-diagonal structure.
3) Since $l_{\frac{1}{2}}$ regularization can give a more sparse solution than $l_{\frac{1}{2}}$ norm, the submodule structure constraint is modified using $l_{\frac{1}{2}}$ norm. Furthermore, using the abilities of $l_{\frac{1}{2}}$-induced TNN and $l_{\frac{1}{2}}$ regularization, a single stage optimization problem is formulated to obtain a better self-expressive representation to retrieve the underlying clusters.

Combining all the above, the proposed optimization problem can be reformulated as,

$$\min_{\mathcal{Z}} \left\|\mathcal{Z}\right\|_{\circledast\frac{1}{2}} + \lambda_1 \sum_{k=1}^{N_s} \left\|\mathbf{P}_{DM} \odot \mathcal{Z}^{(k)}\right\|_{\frac{1}{2}}^{\frac{1}{2}} + \lambda_2 \left\|\mathcal{X} - \mathcal{X} * \mathcal{Z}\right\|_F^2 \tag{7}$$

where, $\|.\|_{\circledast \frac{1}{2}}$ represents the $l_{\frac{1}{2}}$-induced *TNN* and $\|.\|_{\frac{1}{2}}^{\frac{1}{2}}$ represents $l_{\frac{1}{2}}$ norm. Further, $\|.\|_F$ denotes the Frobenius norm. Finally, $\mathcal{X}$ represents the third order data tensor $\mathcal{X} \in \mathbb{R}^{S_1 S_3 \times N \times N_s}$. Further, we employ variable splitting for $\mathcal{Z}$, into the above equation such that $\mathcal{Z} = \mathcal{C}$ and $\mathcal{Z} = \mathcal{Q}$.

$$\begin{aligned} \min_{\mathcal{C},\mathcal{Q},\mathcal{Z}} \|\mathcal{C}\|_{\circledast \frac{1}{2}} + \lambda_1 \sum_{k=1}^{N_s} \|\mathbf{P}_{DM} \odot \mathcal{Q}^{(k)}\|_{\frac{1}{2}}^{\frac{1}{2}} \\ + \lambda_2 \|\mathcal{X} - \mathcal{X} * \mathcal{Z}\|_F^2 \\ \text{s. t. } \mathcal{Z} = \mathcal{C}, \quad \mathcal{Z} = \mathcal{Q} \end{aligned} \tag{8}$$

In the above expression, λ_1 and λ_2 denote the regularization parameters. The above constrained equation is transformed into a unconstrained one using Augmented Lagrangian (AL) Method [26], given by,

$$\begin{aligned} &\mathcal{L}(\mathcal{C}, \mathcal{Q}, \mathcal{Z}, \mathcal{G}_1, \mathcal{G}_2) \\ &= \|\mathcal{C}\|_{\circledast \frac{1}{2}} + \lambda_1 \sum_{k=1}^{N_s} \|\mathbf{P}_{DM} \odot \mathcal{Q}^{(k)}\|_{\frac{1}{2}}^{\frac{1}{2}} \\ &\quad + \lambda_2 \|\mathcal{X} - \mathcal{X} * \mathcal{Z}\|_F^2 + \langle \mathcal{G}_1, \mathcal{Z} - \mathcal{C} \rangle + \langle \mathcal{G}_2, \mathcal{Z} - \mathcal{Q} \rangle \\ &\quad + \frac{\mu}{2}\left(\|\mathcal{Z} - \mathcal{C}\|_F^2 + \|\mathcal{Z} - \mathcal{Q}\|_F^2\right) \end{aligned} \tag{9}$$

where, tensors $\mathcal{G}_1$ and $\mathcal{G}_2$ are the Lagrangian multipliers, $\mu \geq 0$ denotes the penalty parameter and $\langle ., . \rangle$ denotes the inner product. The above problem can be solved by iteratively minimizing the Lagrangian $\mathcal{L}$ over one tensor while keeping the others constant.

$\mathcal{C}$ Subproblem: The update expression for $\mathcal{C}$ is given by,

$$\mathcal{C}^{[j+1]} = \arg\min_{\mathcal{C}} \|\mathcal{C}\|_{\circledast \frac{1}{2}} + \langle \mathcal{G}_1, \mathcal{Z} - \mathcal{C} \rangle + \frac{\mu}{2}\|\mathcal{Z} - \mathcal{C}\|_F^2 \tag{10}$$

The subproblem of updating $\mathcal{C}$ can be transformed into the following form,

$$\mathcal{C}^{[j+1]} = \arg\min_{\mathcal{C}} \|\mathcal{C}\|_{\circledast \frac{1}{2}} + \frac{\mu^{[j]}}{2}\left\|\mathcal{C} - \left(\mathcal{Z}^{[j]} - \frac{\mathcal{G}_1^{[j]}}{\mu^{[j]}}\right)\right\|_F^2 \tag{11}$$

The solution to the above subproblem is obtained by,

$$\mathcal{C}^{[j+1]} = \mathcal{H}_\tau\left[\mathcal{Z}^{[j]} - \frac{\mathcal{G}_1^{[j]}}{\mu^{[j]}}\right] \tag{12}$$

where, $\mathcal{H}_\tau$ [.] is the singular value half thresholding operator and $\tau = \frac{1}{\mu}$ is the threshold value.

$\mathcal{Q}$ Subproblem: The update expression for $\mathcal{Q}$ is given by,

$$\begin{aligned} \mathcal{Q}^{[j+1]} = \arg\min_{\mathcal{Q}} \lambda_1 \sum_{k=1}^{N_s} \|\mathbf{P}_{DM} \odot \mathcal{Q}^{(k)}\|_{\frac{1}{2}}^{\frac{1}{2}} \\ + \langle \mathcal{G}_2, \mathcal{Z} - \mathcal{Q} \rangle + \frac{\mu}{2}\|\mathcal{Z} - \mathcal{Q}\|_F^2 \end{aligned} \tag{13}$$

The above equation can be decomposed into N_s expressions and the k^{th} frontal slice of $\mathcal{Q}$ can be updated by,

$$\begin{aligned} \mathcal{Q}^{(k)[j+1]} = \arg\min_{\mathcal{Q}} \lambda_1 \|\mathbf{P}_{DM} \odot \mathcal{Q}^{(k)}\|_{\frac{1}{2}}^{\frac{1}{2}} \\ + \frac{\mu^{[j]}}{2}\left\|\mathcal{Q} - \left(\mathcal{Z}^{(k)[j]} + \frac{\mathcal{G}_2^{(k)[j]}}{\mu^{[j]}}\right)\right\|_F^2 \end{aligned} \tag{14}$$

where, $\mathcal{Q}^{(k)[j+1]}$ is the k^{th} frontal slice/matrix of $\mathcal{Q}$. The solution to the above subproblem is given by half thresholding operator [37],

$$\mathcal{Q}^{(k)[j+1]} = H_{\frac{\lambda_2 \mathbf{P}_{DM}}{\mu^i}}\left[\mathcal{Z}^{(k)[j]} + \frac{\mathcal{G}_2^{(k)[j]}}{\mu^{[j]}}\right] \tag{15}$$

where, $H_{\frac{\lambda_2 \mathbf{P}_{DM}}{\mu^i}}$ is the halfthresholding operator [32]. Here, $\mathcal{Q}_{m,n}^{(k)}$ is the $(m, n)^{th}$ element of k^{th} frontal slice/matrix of $\mathcal{Q}$.

$\mathcal{Z}$ Subproblem: The subproblem for updating $\mathcal{Z}$ is given by,

$$\begin{aligned} \mathcal{Z}^{[j+1]} = \arg\min_{\mathcal{Z}} \lambda_2 \|\mathcal{X} - \mathcal{X} * \mathcal{Z}\|_F^2 + \langle \mathcal{G}_1^{[j]}, \mathcal{Z} - \mathcal{C}^{[j+1]} \rangle \\ + \frac{\mu^{[j]}}{2}\|\mathcal{Z} - \mathcal{C}^{[j+1]}\|_F^2 + \frac{\mu^{[j]}}{2}\|\mathcal{Z} - \mathcal{Q}^{[j+1]}\|_F^2 \\ + \langle \mathcal{G}_2^{[j]}, \mathcal{Z} - \mathcal{Q}^{[j+1]} \rangle \end{aligned} \tag{16}$$

It can be written as,

$$\begin{aligned} \mathcal{Z}^{[j+1]} = \arg\min_{\mathcal{Z}} \quad \lambda_2 \|\mathcal{X} - \mathcal{X} * \mathcal{Z}\|_F^2 \\ + \frac{\mu^{[j]}}{2}\left(\|\mathcal{Z} - \mathcal{C}^{[j+1]}\|_F^2 + \|\mathcal{Z} - \mathcal{Q}^{[j+1]}\|_F^2\right) \end{aligned} \tag{17}$$

Taking Fourier Transform on both sides, the above equation can be rewritten as,

$$\begin{aligned} \widehat{\mathcal{Z}}^{[j+1]} = \arg\min_{\widehat{\mathcal{Z}}} \lambda_2 \|\widehat{\mathcal{X}} - \widehat{\mathcal{X}} \otimes \widehat{\mathcal{Z}}\|_F^2 \\ + \frac{\mu^{[j]}}{2}\left(\|\widehat{\mathcal{Z}} - \widehat{\mathcal{P}}_1^{[j+1]}\|_F^2 + \|\widehat{\mathcal{Z}} - \widehat{\mathcal{P}}_2^{[j+1]}\|_F^2\right) \end{aligned} \tag{18}$$

where, $\widehat{\mathcal{Z}}$, $\widehat{\mathcal{P}}_1^{[j+1]}$ and $\widehat{\mathcal{P}}_2^{[j+1]}$ are the Fourier transforms of k^{th} frontal slice of $\mathcal{Z}$, $\mathcal{C}^{[j+1]} - \frac{\mathcal{G}_1^{[j]}}{\mu^{[j]}}$ and $\mathcal{Q}^{[j+1]} - \frac{\mathcal{G}_2^{[j]}}{\mu^{[j]}}$ respectively and '$\otimes$' indicates the slicewise multiplication. The analytic solution for the update of the k^{th} frontal slice is given by,

$$\begin{aligned} \widehat{\mathcal{Z}}^{(k)[j+1]} = \left(2\lambda_2 \widehat{\mathcal{X}}^{(k)T} \widehat{\mathcal{X}}^{(k)} + 2\mu^{[j]}\mathbf{I}\right)^{-1} \\ \times \left(2\lambda_2 \widehat{\mathcal{X}}^{(k)T} \widehat{\mathcal{X}}^{(k)} + \mu^{[j]}\left(\widehat{\mathcal{P}}_1^{(k)[j+1]} + \widehat{\mathcal{P}}_2^{(k)[j+1]}\right)\right) \end{aligned} \tag{19}$$

In the algorithm, stopping criterion is measured by the following condition as (20), shown at the bottom of the next page. The proposed method can be summarized in **Algorithm 2**.

Algorithm 2 Clustering Framework for Hyperspectral Data

Require: Data: $\mathcal{X} \in \mathbb{R}^{S_1 S_3 \times N \times N_s}$ and parameters λ_1, λ_2, μ_{max}, ρ

Ensure: $\mathcal{Z} \in \mathbb{R}^{N \times N \times N_s}$

$\mathcal{C}^{[0]}=\mathcal{Q}^{[0]}=\mathcal{Z}^{[0]}=\mathcal{G}_1^{[0]}=\mathcal{G}_2^{[0]} \leftarrow 0 \in \mathbb{R}^{N \times N \times N_s}$

$\lambda_1 > 0, \lambda_2 > 0, \mu^{[0]} > 0, \rho > 0$

while not converged **do**

  $\mathcal{C}^{[j+1]} \leftarrow$ Update using *Eq:* (12)

  $\mathcal{Q}^{[j+1]} \leftarrow$ Update using *Eq:* (15)

  $\mathcal{Z}^{[j+1]} \leftarrow$ Update using *Eq:* (19)

  $\mathcal{G}_1^{[j+1]} = \mathcal{G}_1^{[j]} + \mu^{[j]} (\mathcal{Z} - \mathcal{C})$

  $\mathcal{G}_2^{[j+1]} = \mathcal{G}_2^{[j]} + \mu^{[j]} (\mathcal{Z} - \mathcal{Q})$

  $\mu^{[j+1]} = \rho \mu^{[j]}$

  Check the convergence using *Eq:* (20)

  $[j] \leftarrow [j+1]$

end while

V. EXPERIMENTS AND RESULTS

A. HYPERSPECTRAL PAPER DATASET PREPARATION

The hyperspectral images of the papers to be clustered are captured using a push-broom hyperspectral camera, HySpex VNIR-1800 with wavelength ranges from 400 nm to 1000 nm with a spectral sampling of 3.18 nm. Further, a pre-processing software, HySpex RAD [9] is used to perform basic camera corrections including dark current subtraction, sensor correction and radiometric calibration. Then, the arrangement for creating hyperspectral paper data of those papers are illustrated in FIGURE 7 (a), where the papers are arranged in a tiled format. This dataset is prepared using papers of various colors, thicknesses, textures, ages, and manufacturers. The types of papers are given in TABLE 2. The obtained hyperspectral paper data has a spatial dimension of 7500×1800 and consists of 186 spectral bands. The hyperspectral paper data sample areas, $\mathbb{R}^{S_1 \times S_3 \times N_s}$ are selected from a wide range such as $\mathbb{R}^{10 \times 10 \times 186}$ to $\mathbb{R}^{50 \times 50 \times 186}$ for creating the input data tensor $\mathcal{X} \in \mathbb{R}^{S_1 S_3 \times N \times N_s}$. Please refer Section IV-A which describes the process of arranging the hyperspectral patches into the lateral slices of a third order input data tensor.

B. EXPERIMENTAL RESULTS

This section presents the Experimental results of the proposed method and the state-of-the-arts. Sparse Subspace Clustering (SSC) [40], Low Rank Subspace Clustering (LRSC) [41], Least Square Regression (LSR) [42], Structure Constrained-Low Rank Representation (SCLRR) [38], Structure Constrained Low Rank Submodule Clustering (SCLRSmC) [26] and l_0-LRSC [43] are the state-of-the-art clustering algorithms chosen for comparison. Similarly, Accuracy (ACC), Normalized Mutual Information (NMI), Purity, Adjusted Rand Index (ARI), F-score, Precision, and Recall are the quality metrics that have been employed for evaluating the algorithms. All of the metrics described above have already been defined and presented in various papers. Accuracy, NMI definitions and equations can be found in [44] and in [45]. Then, in [31] and [46], expressions for Purity and ARI are given. Similarly, F-score, Precision and Recall measures are expressed in terms of True Positives (TP), True Negatives (TN), False Positives (FP) and False Negatives (FN), where the expressions of the aforementioned can be found in [47], [48]. Most of these metrics have widely been used in the clustering methods described in the literature as well as in the state-of-the arts [34], [49]. The values of these metrics are normalized in the range [0, 1], with 1 indicating perfect clustering. However, in practise, higher values of these measures close to 1 imply good clustering results. In this work, all of the algorithms have been subjected to at least 20 trials, and the evaluation metrics are presented in terms of mean and standard deviations ($m \pm \sigma$).

We compare the performance of our proposed method on different hyperspectral sample sizes $\in \mathbb{R}^{S_1 \times S_3 \times N_s}$, where the spatial dimensions $S_1 \times S_3$ have been varied from 10×10 to 50×50. Furthermore, the number of papers selected for clustering have been divided into four cases. In Case I, the first ten papers (Paper 1 to Paper 10) listed in TABLE 2 are considered, whereas for case II, the first twenty papers, i.e. HP proofing paper to 65 gr/m^2 white paper, and so on. All of

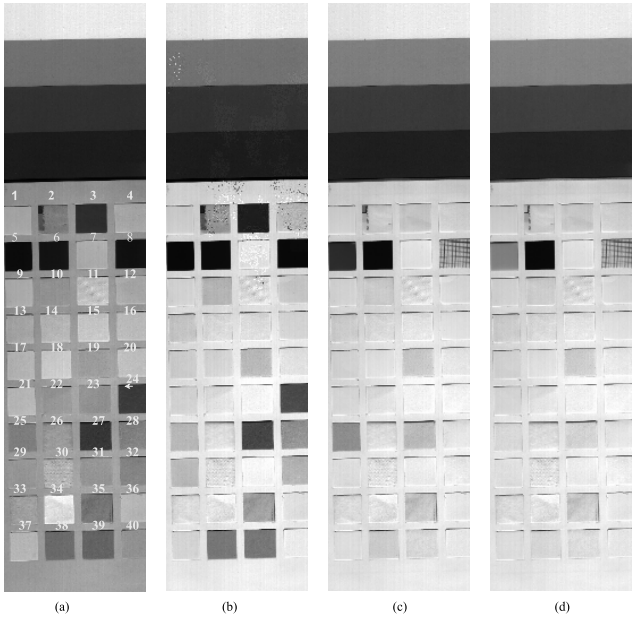


FIGURE 7. FIGURE illustrates the hyperspectral paper data arranged in tiled format. The order of the hyperspectral paper data used in this work are shown in the leftmost image. Images correspond to various bands captured at wavelengths of (a) 408 nm. (b) 475 nm. (c) 756 nm. (d) 985 nm.

$$\max \left\{ \begin{matrix} \|\mathcal{Z}^{[j+1]} - \mathcal{C}^{[j]}\|_\infty, & \|\mathcal{Z}^{[j+1]} - \mathcal{Q}^{[j]}\|_\infty, & \|\mathcal{Z}^{[j+1]} - \mathcal{Z}^{[j]}\|_\infty \\ \|\mathcal{C}^{[j+1]} - \mathcal{C}^{[j]}\|_\infty, & \|\mathcal{Q}^{[j+1]} - \mathcal{Q}^{[j]}\|_\infty & \end{matrix} \right\} < \epsilon \tag{20}$$

TABLE 2. TABLE shows different types and materials of the papers used to prepare hyperspectral paper data. In this, types of the papers are listed in the order that the papers given in FIGURE 7 (a).

Paper	Type	Paper	Type	Paper	Type	Paper	Type
Paper 1	HP proofing	Paper 11	HP canvas substrate	Paper 21	90 gr/m^2 White paper	Paper 31	Standard Copy 80 gr/m^2
Paper 2	Magazine offset	Paper 12	Coated offset paper	Paper 22	70 gr/m^2 White paper	Paper 32	Flip cover paper
Paper 3	Yellow millimeter	Paper 13	60 gr/m^2 White paper	Paper 23	Business Card	Paper 33	Butter paper
Paper 4	Uncoated offset	Paper 14	68 gr/m^2 White paper	Paper 24	Sticky note	Paper 34	Thick envelope
Paper 5	Analogous film emulsion	Paper 15	80 gr/m^2 White paper	Paper 25	Sticky note	Paper 35	News paper from Aften Posten
Paper 6	Analogous photo	Paper 16	100 gr/m^2 White paper	Paper 26	Baking	Paper 36	Paper bag
Paper 7	Digital HP proofing	Paper 17	128 gr/m^2 White paper	Paper 27	Envelop from China	Paper 37	Standard offset
Paper 8	Analog film emulsion	Paper 18	170 gr/m^2 White paper	Paper 28	Billing	Paper 38	Sticky note
Paper 9	HP premium	Paper 19	54 gr/m^2 White paper	Paper 29	Unknown white	Paper 39	News paper
Paper 10	Capafix foil	Paper 20	65 gr/m^2 White paper	Paper 30	Hnad tissue	Paper 40	Drawing paper

TABLE 3. Quantitative comparison of all the methods for the data tensor $\mathcal{X} \in \mathbb{R}^{S_1 S_3 \times N \times N_S}$ composed by hyperspectral patches of dimension $\mathbb{R}^{10\times10\times186}$. Evaluation measures are given as mean and standard deviations ($m \pm \sigma$). Best: **Bold**.

Dataset	Data Tensor $\mathcal{X} \in \mathbb{R}^{S_1 S_3 \times N \times N_s}$	Methods	ACC	NMI	ARI	F-score	Precision	Recall
Case I Hyperspectral data of 10 papers [Paper 1 to Paper 10]	$\mathcal{X} \in \mathbb{R}^{100\times100\times186}$	SSC	0.7790±0.0033	0.7633±0.0005	0.7334±0.0007	0.7063±0.0007	0.7157±0.0004	0.7266±0.0002
		LRSC	0.8000±0.0000	0.7889±0.0000	0.7723±0.0000	0.7347±0.0000	0.7355±0.0000	0.7566±0.0000
		LSR	0.8505±0.0023	0.8356±0.0005	0.8240±0.0052	0.8225±0.0005	0.8246±0.0015	0.8256±0.0125
		SCLRR	0.9400±0.0050	0.9400±0.0023	0.9350±0.0052	0.9200±0.0550	0.9255±0.0030	0.9300±0.0050
		SCLRSmC	0.9000±0.0002	0.9020±0.0000	0.8990±0.0015	0.8600±0.0005	0.8895±0.0003	0.8750±0.0022
		l_0-LRSSC	0.8400±0.0000	0.7800±0.0000	0.7098±0.0002	0.7606±0.0000	0.7600±0.0020	0.7780±0.0002
		Proposed	**0.9650±0.0020**	**0.9760±0.0020**	**0.9594±0.0605**	**0.9525±0.0546**	**0.9462±0.0826**	**0.9633±0.0233**
Case II Hyperspectral data of 20 papers [Paper 1 to Paper 20]	$\mathcal{X} \in \mathbb{R}^{100\times200\times186}$	SSC	0.5705±0.0030	0.5632±0.0013	0.5547±0.0023	0.5538±0.0003	0.5276±0.0005	0.5350±0.0005
		LRSC	0.5000±0.0002	0.5299±0.0000	0.4754±0.0000	0.5099±0.0022	0.4509±0.0005	0.4990±0.0000
		LSR	0.6900±0.0520	0.6756±0.0002	0.6750±0.0012	0.6745±0.0052	0.6775±0.0055	0.6807±0.0025
		SCLRR	0.8098±0.0000	0.7896±0.0002	0.7738±0.0042	0.7786±0.0022	0.7647±0.0073	0.7845±0.0012
		SCLRSmC	0.8858±0.0059	0.8660±0.0003	0.8428±0.0020	0.8500±0.0022	0.8425±0.0017	0.8450±0.0006
		l_0-LRSSC	0.8500±0.0000	0.8600±0.0000	0.8268±0.0020	0.8357±0.0000	0.8012±0.0000	0.8578±0.0000
		Proposed	**0.9365±0.0625**	**0.9334±0.0230**	**0.9250±0.0789**	**0.9214±0.0750**	**0.9198±0.1067**	**0.9273±0.0391**
Case III Hyperspectral data of 30 papers [Paper 1 to Paper 30]	$\mathcal{X} \in \mathbb{R}^{100\times300\times186}$	SSC	0.5577±0.0052	0.5333±0.0060	0.5290±0.0018	0.5232±0.0030	0.5310±0.0004	0.5325±0.0004
		LRSC	0.3333±0.0005	0.3745±0.0052	0.3849±0.0000	0.3518±0.0002	0.3164±0.0020	0.3552±0.0002
		LSR	0.4099±0.0005	0.3895±0.0020	0.3750±0.0050	0.3805±0.0022	0.3802±0.0022	0.3799±0.0015
		SCLRR	0.6438±0.0059	0.6255±0.0016	0.6004±0.0113	0.6039±0.0113	0.6055±0.0070	0.6110±0.0004
		SCLRSmC	0.8025±0.0021	0.8007±0.0030	0.7998±0.0022	0.7950±0.0005	0.8029±0.0054	0.7988±0.0005
		l_0-LRSSC	0.6667±0.0022	0.6956±0.0030	0.4864±0.0110	0.5094±0.0000	0.5418±0.0022	0.5560±0.0030
		Proposed	**0.9220±0.0962**	**0.9205±0.0310**	**0.8697±0.1192**	**0.8924±0.1104**	**0.8853±0.0025**	**0.9046±0.0474**
Case IV Hyperspectral data of 40 papers [Paper 1 Paper 40]	$\mathcal{X} \in \mathbb{R}^{100\times400\times186}$	SSC	0.4286±0.0038	0.3955±0.0012	0.4083±0.0025	0.3958±0.0002	0.4005±0.0022	0.4155±0.0005
		LRSC	0.3150±0.0000	0.3527±0.0000	0.3100±0.0000	0.2900±0.0025	0.3002±0.0050	0.3250±0.0022
		LSR	0.3225±0.0022	0.3105±0.0050	0.2995±0.0052	0.3020±0.0005	0.2900±0.0050	0.3005±0.0052
		SCLRR	0.5159±0.0011	0.4998±0.0053	0.5128±0.0025	0.5005±0.0011	0.4875±0.0205	0.5011±0.0220
		SCLRSmC	0.8005±0.0005	0.8004±0.0042	0.7755±0.0016	0.7846±0.0018	0.7725±0.0052	0.7625±0.0005
		l_0-LRSSC	0.4625±0.0005	0.5263±0.0020	0.4089±0.0000	0.4303±0.0002	0.4568±0.0002	0.4581±0.0033
		Proposed	**0.9202±0.1052**	**0.9071±0.1036**	**0.9016±0.1385**	**0.8890±0.1348**	**0.8816±0.1832**	**0.8990±0.0636**

TABLE 4. Quantitative comparison of all the methods for the data tensor $\mathcal{X} \in \mathbb{R}^{S_1 S_3 \times N \times N_S}$ composed by hyperspectral patches of dimension $\mathbb{R}^{20\times20\times186}$. Evaluation measures are given as mean and standard deviations ($m \pm \sigma$). Best: **Bold**.

Dataset	Data Tensor $\mathcal{X} \in \mathbb{R}^{S_1 S_3 \times N \times N_s}$	Methods	ACC	NMI	ARI	F-score	Precision	Recall
Case I Hyperspectral data of 10 papers [Paper 1 to Paper 10]	$\mathcal{X} \in \mathbb{R}^{400\times100\times186}$	SSC	0.8720±0.0042	0.8548±0.0006	0.8599±0.0016	0.8336±0.0015	0.8301±0.0005	0.8402±0.0028
		LRSC	0.9200±0.0000	0.9360±0.0000	0.9106±0.0005	0.9189±0.0025	0.9247±0.0210	0.9244±0.0100
		LSR	0.7500±0.0021	0.7495±0.0050	0.7490±0.0021	0.7450±0.0082	0.7350±0.0120	0.7400±0.0020
		SCLRR	0.7830±0.0052	0.7352±0.0025	0.7390±0.0025	0.7295±0.0070	0.7199±0.0050	0.7200±0.0050
		SCLRSmC	0.9020±0.0008	0.8971±0.0058	0.8906±0.0012	0.8828±0.0052	0.8802±0.0030	0.8850±0.0025
		l_0-LRSSC	0.7300±0.0000	0.7345±0.0022	0.7342±0.0000	0.7524±0.0000	0.7500±0.0015	0.7644±0.0000
		Proposed	**0.9375±0.1200**	**0.9496±0.0453**	**0.9223±0.1464**	**0.9248±0.0302**	**0.9248±0.0084**	**0.9253±0.0004**
Case II Hyperspectral data of 20 papers [Paper 1 to Paper 20]	$\mathcal{X} \in \mathbb{R}^{400\times200\times186}$	SSC	0.7750±0.0158	0.7624±0.0046	0.7822±0.0136	0.7661±0.0128	0.7597±0.0188	0.7614±0.0011
		LRSC	0.5700±0.0023	0.5905±0.0000	0.4917±0.0056	0.4609±0.0023	0.4975±0.0012	0.5000±0.0015
		LSR	0.5505±0.0042	0.5412±0.0950	0.5395±0.0050	0.5402±0.0302	0.5355±0.0052	0.5400±0.0020
		SCLRR	0.7100±0.0062	0.6850±0.0012	0.6556±0.0021	0.6600±0.0022	0.6725±0.0022	0.6650±0.0050
		SCLRSmC	0.9100±0.0253	0.9056±0.0025	0.9089±0.0054	0.9006±0.0023	0.8952±0.0025	0.9005±0.0022
		l_0-LRSSC	0.7853±0.0001	0.7700±0.0540	0.7693±0.0053	0.7595±0.0020	0.7500±0.0021	0.7536±0.0021
		Proposed	**0.9303±0.0733**	**0.9290±0.0240**	**0.9263±0.0086**	**0.9224±0.0010**	**0.9225±0.0040**	**0.9224±0.0383**
Case III Hyperspectral data of 30 papers [Paper 1 to Paper 30]	$\mathcal{X} \in \mathbb{R}^{400\times300\times186}$	SSC	0.7267±0.0236	0.7290±0.0030	0.7333±0.0312	0.7146±0.0297	0.7266±0.0006	0.7285±0.0036
		LRSC	0.3833±0.0000	0.3866±0.0018	0.3701±0.0090	0.3570±0.0000	0.3669±0.0003	0.3825±0.0120
		LSR	0.5206±0.0110	0.5006±0.0210	0.4952±0.0210	0.5002±0.0040	0.4955±0.0020	0.5015±0.0025
		SCLRR	0.6210±0.0015	0.6105±0.2300	0.6095±0.0480	0.6090±0.0050	0.6000±0.0050	0.6090±0.0800
		SCLRSmC	0.9100±0.0001	0.8872±0.0002	0.8571±0.0001	0.8711±0.0002	0.8812±0.0043	0.8600±0.0052
		l_0-LRSSC	0.7700±0.0062	0.7650±0.0012	0.7550±0.0000	0.7495±0.0220	0.7450±0.0060	0.7458±0.0022
		Proposed	**0.9189±0.0050**	**0.9171±0.0191**	**0.9163±0.0529**	**0.9114±0.0051**	**0.9116±0.0010**	**0.9123±0.0210**
Case IV Hyperspectral data of 40 papers [Paper 1 Paper 40]	$\mathcal{X} \in \mathbb{R}^{400\times400\times186}$	SSC	0.5475±0.0312	0.5250±0.0007	0.5214±0.0289	0.5169±0.0272	0.5172±0.0028	0.5195±0.0012
		LRSC	0.2500±0.0002	0.3200±0.0006	0.2949±0.0061	0.2176±0.0000	0.2275±0.0000	0.2405±0.0006
		LSR	0.3850±0.0520	0.3305±0.0210	0.3550±0.0020	0.3360±0.0004	0.3205±0.0021	0.3305±0.0120
		SCLRR	0.6535±0.0022	0.6450±0.0050	0.6395±0.0020	0.6250±0.0050	0.6290±0.0012	0.6300±0.0050
		SCLRSmC	0.7033±0.0002	0.7278±0.0042	0.6905±0.0012	0.6603±0.0120	0.6692±0.0135	0.6549±0.0090
		l_0-LRSSC	0.7200±0.0000	0.7290±0.0050	0.6998±0.0002	0.6716±0.0025	0.6615±0.0020	0.6825±0.0023
		Proposed	**0.9075±0.0027**	**0.9074±0.0031**	**0.8963±0.0048**	**0.9008±0.0025**	**0.8995±0.0360**	**0.9026±0.0515**

TABLE 5. Quantitative comparison of all the methods for data tensor $\mathcal{X} \in \mathbb{R}^{S_1 S_3 \times N \times N_S}$ composed by hyperspectral patches of dimension $\mathbb{R}^{30\times30\times186}$, $\mathbb{R}^{40\times40\times186}$ and $\mathbb{R}^{50\times50\times186}$. Evaluation measures are given as mean and standard deviations ($m \pm \sigma$).

Datas Tensor $\mathcal{X} \in \mathbb{R}^{S_1 S_3 \times N \times N_s}$	ACC	NMI	ARI	F-score	Precision	Recall
$\mathcal{X} \in \mathbb{R}^{900\times400\times186}$	0.8893±0.0522	0.8809±0.0306	0.8848±0.0696	0.8871±0.0674	0.8859±0.0120	0.8893±0.0691
$\mathcal{X} \in \mathbb{R}^{1600\times400\times186}$	0.8672±0.0005	0.8608±0.0042	0.8644±0.0125	0.8579±0.0122	0.8595±0.0135	0.8559±0.0090
$\mathcal{X} \in \mathbb{R}^{2500\times400\times186}$	0.8604±0.0041	0.8590±0.0038	0.8547±0.0910	0.8496±0.0028	0.8444±0.0029	0.8508±0.0520

the cases have been clearly illustrated in TABLE 3, TABLE 4 and TABLE 5 and in which, detailed comparison of the clustering results of the proposed method with conventional subspace clustering algorithms are also presented. The proposed method produces good clustering results under different scenarios, as illustrated in the aforementioned TABLEs. As given in Case I of TABLE 3, proposed method attains the ACC, NMI and ARI values as (0.9650 ± 0.0020), (0.9760 ± 0.0020) and (0.9594 ± 0.0605) respectively. Our method outperforms SSC, LSR and LRSC in all the cases we have considered. For lesser dimensions of hyperspectral data samples, mentioned in Case I of TABLE 3, methods such as SSC, LSR, and LRSC have produced comparable results. However, for increased dimensions, (Case III and Case IV of TABLE 3), their clustering results show a decline. Other methods, such as SCLRR, l_0-LRSSC and SCLRSmC also have produced good clustering results in general. However, proposed method exhibits superior performance over these two methods in all the scenarios. In Case IV of TABLE 3, where the hyperspectral paper data samples of 40 papers, proposed method obtains comparatively better ACC (0.9202 ± 0.1052), NMI (0.9071 ± 0.1036), ARI (0.9016 ± 0.1385) and F-score (0.8890 ± 0.1348) values. At the same time, other methods produced low evaluation scores in the same scenario.

Similarly, for the data tensor $\mathcal{X} \in \mathbb{R}^{S_1 \times S_3 \times N \times N_s}$, composed of several hyperspectral paper data patches of dimensions $\in \mathbb{R}^{20 \times 20 \times 186}$, proposed method have shown better performance over the other methods. The comparison results are tabulated in TABLE 4. For the data tensor $\mathcal{X} \in \mathbb{R}^{400 \times 400 \times 186}$, in Case IV, proposed method have obtained the ACC value of (0.9075 ± 0.0027) and F-score value of (0.9005 ± 0.0025) respectively. Among the methods we compared, SCLRSmC produces comparatively good results for all cases mentioned in TABLE 4. Methods such as LRSC, LSR, shows severe decline in their performance for increased number of papers taken for clustering. Other methods, SCLRR and l_0-LRSC produces good results for Case I and Case II of TABLE 4, but fails for increasing dimensions of the data tensor $\mathcal{X}$.

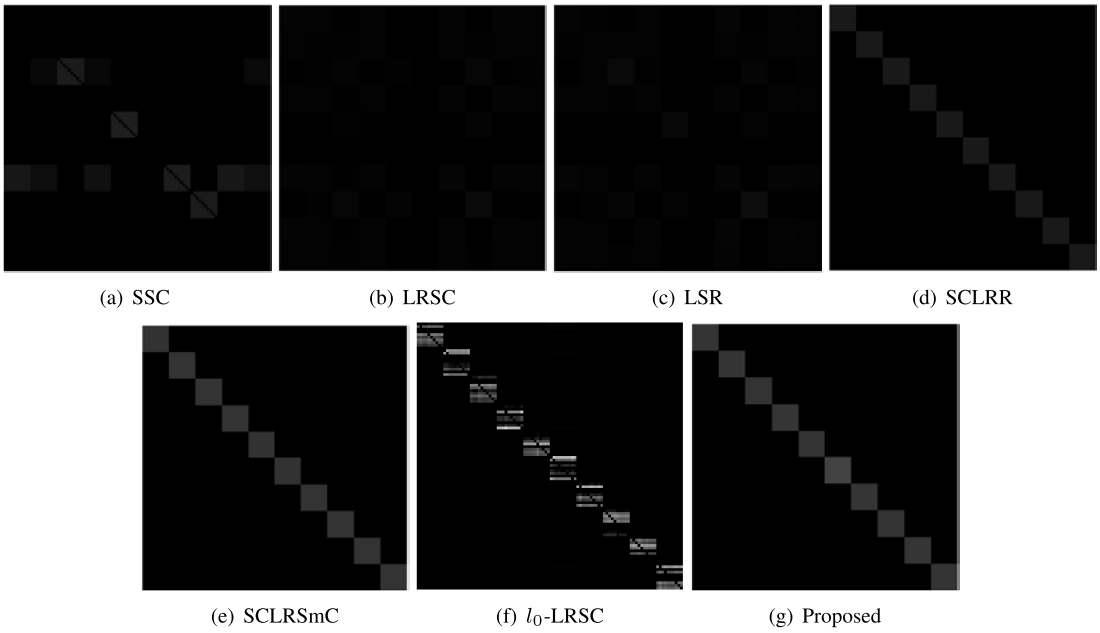
We further tested our method by changing the spatial dimensions of the hyperspectral patch $S_1 \times S_3$ in varying dimensions as 30×30, 40×40 and 50×50 respectively. The obtained results are reported in TABLE 5. Overall, the proposed method performs well under the conditions described in TABLE 3, TABLE 4 and TABLE 5. Moreover, it could outperform the existing methods and maintains consistent performance throughout the different scenarios considered in this work. The proposed method surpasses existing methods due to a number of the following factors. The tensorial arrangement of the hyperspectral samples could help in stacking them into distinct lateral slices. Moreover, each lateral slice of the data tensor $\mathcal{X}$ accommodates the information from all the spectral bands. Also, the dissimilarity matrix employed in the proposed method aids to get a proper f-diagonal structure for the representation tensor $\mathcal{Z}$ which can clearly showcase the high correlation that exists in intra-cluster datapoints and less correlation that exists between datapoints that belong to inter clusters. In addition, a comparison study of the affinity matrices generated by the proposed method and other state-of-the-art methods was also conducted and the learned affinity matrices are given in FIGURE 8. The proposed method could generate an affinity matrix with an accurate block diagonal structure, as can be seen in FIGURE 8 (g). This is also one of the reasons for the proposed method's consistent behaviour in producing good clustering outcomes. Among the compared methods, SCLRSmC and l_0-LRSC produces comparatively better affinity matrices. At the same time, as seen in FIGUREs 8(a), (b) and (c), methods such as SSC [40], LRSC [41] and LSR [42] were failed to maintain the required block diagonal structure for their corresponding affinity matrices.

Hence, from the evaluation results, it is clear that proposed method can be used in the context of detecting forged one in a multi-page document. The method we proposed could effectively cluster the hyperspectral samples selected from different papers. As proposed in this work, it is observed that hyperspectral samples with similar spectral properties have been grouped into a single cluster, while those with a different spectral properties have been grouped into respective clusters. On the other hand, this results precisely demonstrates the truthfulness of the heuristic we proposed in our study. The forged paper/papers in a multipage document will be mapped into some other clusters and the original papers of the documents as a whole will be mapped into a single cluster and thereby the forged papers can be detected easily.

C. COMPARISON OF EXECUTION TIME

Even-though the proposed method consistently producing good clustering results, it has been also realized that, execution time of the proposed method increases with increasing dimensions of the data tensor $\mathcal{X}$. The time required for the proposed method for various dimensions have been shown in FIGURE 9 (a). For data tensor, $\mathcal{X} \in \mathbb{R}^{1600 \times 186 \times 400}$, composed by hyperspectral patches of $40 \times 40 \times 186$, the computational time of the proposed method is nearly 3000 Seconds. Moreover, for $\mathcal{X} \in \mathbb{R}^{2500 \times 186 \times 400}$, the execution time reaches to 5069 seconds. Hence, the computational time of the proposed method increases in a linear fashion with increasing dimensions of hyperspectral slices as represented in FIGURE 9 (a). Similarly, we made an analysis of the computational time required for all the algorithms. FIGURE 9 (b) shows the computational time comparison of all the algorithms with the data tensor $\mathcal{X} \in \mathbb{R}^{S_1 S_3 \times N \times N_s}$, where $S_1 S_3 = 10$, $N_s = 186$ and N varies from 100 to 400. Similarly, in FIGURE 9 (c), we selected hyperspectral patches $\in \mathbb{R}^{20 \times 20 \times 186}$ for all cases reported in TABLE 4.

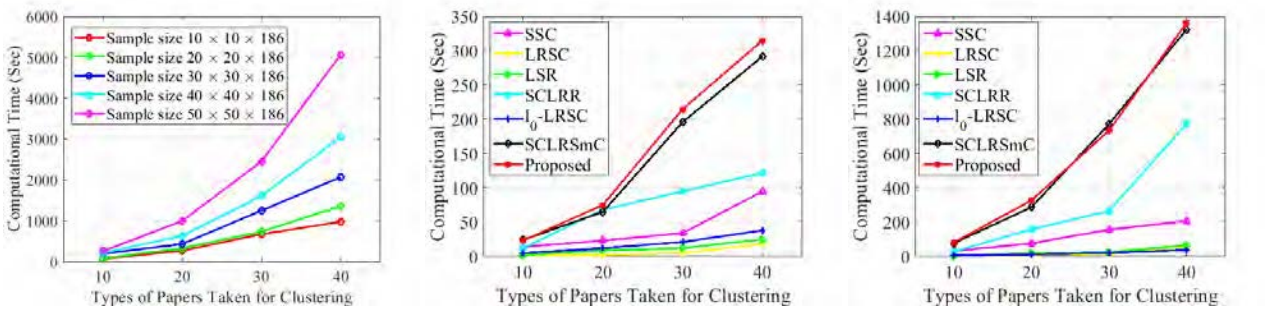
Among the compared methods, SCLRSmC, takes much computational time as shown in FIGURE 9 (b). Methods, such as SSC, LRSC and l_0-LRSC consume less execution time but the results are considerably reduced at varying scenarios as aforementioned. Hence, a disadvantage of the proposed method is huge computational time required for its execution. But, since the proposed method incorporates



(a) SSC (b) LRSC (c) LSR (d) SCLRR

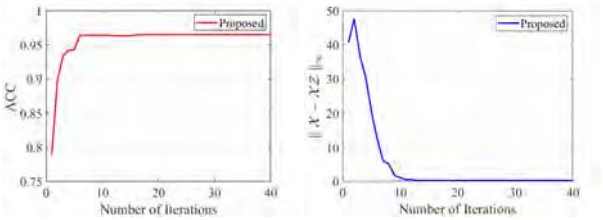
(e) SCLRSmC (f) l_0-LRSC (g) Proposed

FIGURE 8. **Affinity matrices produced by all the algorithms for the data tensor $\mathcal{X} \in \mathbb{R}^{100\times100\times186}$ (Case I of TABLE 3).**



(a) Computational time of the proposed method for varying dimensions of the hyperspectral data samples. (b) Comparison of computational time of all the algorithms for hyperspectral data patches of dimension $\in \mathbb{R}^{10\times10\times186}$. (c) Comparison of computational time of all the algorithms for hyperspectral data patches of dimension $\in \mathbb{R}^{20\times20\times186}$.

FIGURE 9. **Computational time of the proposed method and state-o-the-art algorithms. Vertical axes represent computation time given in seconds and horizontal axes represent the types (number) of hyperspectral paper data taken for clustering.**



(a) ACC V/s Number of iterations. (b) $\|\mathcal{X} - \mathcal{X}\mathcal{Z}\|_\infty$ V/s Number of iterations.

FIGURE 10. **Figure illustrates the convergence analysis of the proposed method with ACC metric and representation error term, $\|\mathcal{X} - \mathcal{X}\mathcal{Z}\|_\infty$.**

the information embedded in all spectral bands, this may be acceptable to some extent. The problem of redundancy in spectral bands can be reduced by incorporating a simultaneous band selection to the proposed method. Then, the hyperspectral samples can be represented by minimum number of spectral bands and thereby the computational time of the proposed method can be reduced to a greater extent. However, we address this challenge in our future work.

D. PARAMETER TUNING, CONVERGENCE AND COMPUTATIONAL COMPLEXITY

The optimum values of the regularization parameters λ_1 and λ_2 have been determined by a grid search to achieve the best clustering results. For the proposed method, we fixed $\lambda_1 = 0.0085$ and $\lambda_2 = 0.0045$ for all the experiments. We evaluated the convergence behaviour of the proposed method against the evaluation metrics ACC as well as the representation error term, $\|\mathcal{X} - \mathcal{X}\mathcal{Z}\|_\infty$ which are presented in FIGUREs 10 (a) and (b) respectively. It has been observed that the proposed algorithm shows a good convergence rate and converges quickly within 10-20 iterations. The computational complexity of the proposed method lies in $\mathcal{C} \in \mathbb{R}^{N\times N\times N_s}$ and $\mathcal{Q} \in \mathbb{R}^{N\times N\times N_s}$ updates. Then, $\mathcal{C}$ update involves $l_{\frac{1}{2}}$-induced TNN which requires,

$\mathcal{O}(N^2N_s\log_2 N_s + \frac{1}{2}N_sN^2 + N^3)$ and $\mathcal{Q}$ update with $l_{\frac{1}{2}}$ regularization requires $\mathcal{O}(N^2N_s)$ operations per iterations. Overall the proposed method bears moderate computational complexity.

VI. CONCLUSION

A tensor framework for clustering of hyperspectral paper data with an application to forensic document analysis has been proposed. In the proposed framework, spectral information from the hyperspectral patches fetched from papers to be clustered are stacked into lateral slices of a third order tensor. Objective function of the proposed method incorporates $l_{\frac{1}{2}}$-induced TNN which improves the low rankness of the representation tensor. Similarly, the structural constraint employed by means of $l_{\frac{1}{2}}$ regularization and the dissimilarity matrix facilitates to achieve f-diagonal structure of the representation tensor. The optimization problem formulated has been solved using Inexact Augmented Lagrangian Method. The proposed method has been evaluated and further compared with state-of-the-art clustering techniques. The results show that proposed method produces consistent clustering results and outperforms the other methods.

ACKNOWLEDGMENT

The authors would like to thank Norwegian Colour and Visual Computing Laboratory, Department of Computer Science, Norwegian University of Science and Technology-NTNU for providing access to the hyperspectral paper dataset used in this work.

REFERENCES

[1] Y. Chen, N. M. Nasrabadi, and T. D. Tran, "Hyperspectral image classification using dictionary-based sparse representation," *IEEE Trans. Geosci. Remote Sens.*, vol. 49, no. 10, pp. 3973–3985, Oct. 2011.

[2] C. Cao, J. Yu, C. Zhou, K. Hu, F. Xiao, and X. Gao, "Hyperspectral image denoising via subspace-based nonlocal low-rank and sparse factorization," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 12, no. 3, pp. 973–988, Mar. 2019.

[3] L. Bilius and S. Pentiuc, "Unsupervised clustering for hyperspectral images," *Symmetry*, vol. 12, no. 2, p. 277, Feb. 2020.

[4] H. Zhang, H. Zhai, L. Zhan, and P. Li, "Spectral–spatial sparse subspace clustering for hyperspectral remote sensing images," *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 6, pp. 3672–3684, Mar. 2016.

[5] A. W. Bitar, L.-F. Cheong, and J.-P. Ovarlez, "Sparse and low-rank matrix decomposition for automatic target detection in hyperspectral imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 8, pp. 5239–5251, Aug. 2019.

[6] B. M. Devassy, S. George, and P. Nussbaum, "Unsupervised clustering of hyperspectral paper data using t-SNE," *J. Imag.*, vol. 6, no. 5, p. 29, May 2020.

[7] G. A. Shaw and H.-H. K. Burke, "Spectral imaging for remote sensing," *Lincoln Lab. J.*, vol. 14, no. 1, pp. 3–28, 2003.

[8] B. Fei, "Hyperspectral imaging in medical applications," in *Data Handling in Science and Technology*, vol. 32. Amsterdam, The Netherlands: Elsevier, 2020, pp. 523–565.

[9] B. Melit Devassy and S. George, "Dimensionality reduction and visualisation of hyperspectral ink data using t-SNE," *Forensic Sci. Int.*, vol. 311, Jun. 2020, Art. no. 110194.

[10] B. M. Devassy and S. George, "Contactless classification of strawberry using hyperspectral imaging," in *Proc. CEUR Workshop*, 2020, pp. 1–12.

[11] B. M. Devassy, S. George, and J. Y. Hardeberg, "Comparison of ink classification capabilities of classic hyperspectral similarity features," in *Proc. Int. Conf. Document Anal. Recognit. Workshops (ICDARW)*, Sep. 2019, pp. 25–30.

[12] J. Xu, J. E. Fowler, and L. Xiao, "Hypergraph-regularized low-rank subspace clustering using superpixels for unsupervised spatial–spectral hyperspectral classification," *IEEE Geosci. Remote Sens. Lett.*, vol. 18, no. 5, pp. 871–875, May 2021.

[13] J. C. Bezdek, W. Full, and R. Ehrlich, "FCM: The fuzzy c-means clustering algorithm," *Comput. Geosci.*, vol. 10, nos. 2–3, pp. 191–203, 1984.

[14] A. Likas, N. Vlassis, and J. J. Verbeek, "The global k-means clustering algorithm," *Pattern Recognit.*, vol. 36, no. 2, pp. 451–461, Feb. 2003.

[15] A. Sellami, A. B. Abbes, V. Barra, and I. R. Farah, "Fused 3–D spectral–spatial deep neural networks and spectral clustering for hyperspectral image classification," *Pattern Recognit. Lett.*, vol. 138, pp. 594–600, Oct. 2020.

[16] L. Wang, S. Niu, X. Gao, K. Liu, F. Lu, Q. Diao, and J. Dong, "Fast high-order sparse subspace clustering with cumulative MRF for hyperspectral images," *IEEE Geosci. Remote Sens. Lett.*, vol. 18, no. 1, pp. 152–156, Jan. 2021.

[17] W. Sun and Q. Du, "Graph-regularized fast and robust principal component analysis for hyperspectral band selection," *IEEE Trans. Geosci. Remote Sens.*, vol. 56, no. 6, pp. 3185–3195, Jun. 2018.

[18] J. Wang and C.-I. Chang, "Independent component analysis-based dimensionality reduction with applications in hyperspectral image analysis," *IEEE Trans. Geosci. Remote Sens.*, vol. 44, no. 6, pp. 1586–1600, Jun. 2006.

[19] T. V. Bandos, L. Bruzzone, and G. Camps-Valls, "Classification of hyperspectral images with regularized linear discriminant analysis," *IEEE Trans. Geosci. Remote Sens.*, vol. 47, no. 3, pp. 862–873, Mar. 2009.

[20] W. Li, S. Prasad, J. E. Fowler, and L. M. Bruce, "Locality-preserving dimensionality reduction and classification for hyperspectral image analysis," *IEEE Trans. Geosci. Remote Sens.*, vol. 50, no. 4, pp. 1185–1198, Apr. 2012.

[21] H. Zhai, H. Zhang, L. Zhang, and P. Li, "Laplacian-Regularized low-rank subspace clustering for hyperspectral image band selection," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 3, pp. 1723–1740, Mar. 2019.

[22] Y. Kong, Y. Cheng, C. L. P. Chen, and X. Wang, "Hyperspectral image clustering based on unsupervised broad learning," *IEEE Geosci. Remote Sens. Lett.*, vol. 16, no. 11, pp. 1741–1745, Nov. 2019.

[23] H. Zhai, H. Zhang, L. Zhang, P. Li, and A. Plaza, "A new sparse subspace clustering algorithm for hyperspectral remote sensing imagery," *IEEE Geosci. Remote Sens. Lett.*, vol. 14, no. 1, pp. 43–47, Jan. 2017.

[24] Q. Liu, Y. Sun, R. Hang, and H. Song, "Spatial–spectral locality-constrained low-rank representation with semi-supervised hypergraph learning for hyperspectral image classification," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 10, no. 9, pp. 4171–4182, Sep. 2017.

[25] F. Luo, B. Du, L. Zhang, L. Zhang, and D. Tao, "Feature learning using spatial–spectral Hypergraph discriminant analysis for hyperspectral image," *IEEE Trans. Cybern.*, vol. 49, no. 7, pp. 2406–2419, Jul. 2019.

[26] T. Wu and W. U. Bajwa, "A low tensor-rank representation approach for clustering of imaging data," *IEEE Signal Process. Lett.*, vol. 25, no. 8, pp. 1196–1200, Aug. 2018.

[27] 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, 2013.

[28] 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," 2016, *arXiv:1601.00149*.

[29] E. Kernfeld, S. Aeron, and M. Kilmer, "Clustering multi-way data: A novel algebraic approach," 2014, *arXiv:1412.7056*.

[30] M. Imani and H. Ghassemian, "Band clustering-based feature extraction for classification of hyperspectral images using limited training samples," *IEEE Geosci. Remote Sens. Lett.*, vol. 11, no. 8, pp. 1325–1329, Aug. 2014.

[31] J. Francis, B. M, and S. N. George, "A unified tensor framework for clustering and simultaneous reconstruction of incomplete imaging data," *ACM Trans. Multimedia Comput., Commun., Appl.*, vol. 16, no. 3, pp. 1–24, Sep. 2020.

[32] Z. Xu, X. Chang, F. Xu, and H. Zhang, "$L_{1/2}$ regularization: A thresholding representation theory and a fast solver," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 23, no. 7, pp. 1013–1027, Jul. 2012.

[33] J. Zeng, S. Lin, Y. Wang, and Z. Xu, "$L_{1/2}$ regularization: Convergence of iterative half thresholding algorithm," *IEEE Trans. Signal Process.*, vol. 62, no. 9, pp. 2317–2329, May 2014.
[34] M. Yin, J. Gao, S. Xie, and Y. Guo, "Multiview subspace clustering via tensorial t-product representation," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 30, no. 3, pp. 851–864, Mar. 2018.
[35] T. Wu, "Graph regularized low-rank representation for submodule clustering," *Pattern Recognit.*, vol. 100, Apr. 2020, Art. no. 107145.
[36] W. Hu, D. Tao, W. Zhang, Y. Xie, and Y. Yang, "The twist tensor nuclear norm for video completion," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 28, no. 12, pp. 2961–2973, Dec. 2016.
[37] A. J. Tom and S. N. George, "A three-way optimization technique for noise robust moving object detection using tensor low-rank approximation, l1/2, and TTV regularizations," *IEEE Trans. Cybern.*, vol. 51, no. 2, pp. 1004–1014, Feb. 2021.
[38] 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.
[39] A. Kumar and S. Abirami, "Aspect-based opinion ranking framework for product reviews using a Spearman's rank correlation coefficient method," *Inf. Sci.*, vols. 460–461, pp. 23–41, Sep. 2018.
[40] 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.
[41] P. Favaro, R. Vidal, and A. Ravichandran, "A closed form solution to robust subspace estimation and clustering," in *Proc. CVPR*, Jun. 2011, pp. 1801–1807.
[42] C.-Y. Lu, H. Min, Z.-Q. Zhao, L. Zhu, D.-S. Huang, and S. Yan, "Robust and efficient subspace segmentation via least squares regression," in *Proc. Eur. Conf. Comput. Vis.* Berlin, Germany: Springer, 2012, pp. 347–360.
[43] M. Brbić and I. Kopriva, "ℓ_0-motivated low-rank sparse subspace clustering," *IEEE Trans. Cybern.*, vol. 50, no. 4, pp. 1711–1725, Apr. 2020.
[44] Q. Wang, J. Cheng, Q. Gao, G. Zhao, and L. Jiao, "Deep multi-view subspace clustering with unified and discriminative learning," *IEEE Trans. Multimedia*, vol. 23, pp. 3483–3493, 2021.
[45] M. Abavisani and V. M. Patel, "Deep multimodal subspace clustering networks," *IEEE J. Sel. Topics Signal Process.*, vol. 12, no. 6, pp. 1601–1614, Dec. 2018.
[46] B. Madathil and S. N. George, "Noise robust image clustering based on reweighted low rank tensor approximation and $\ell_{\frac{1}{2}}$ regularization," *Signal, Image Video Process.*, pp. 1–9, 2020.
[47] P. Das, A. K. Das, J. Nayak, D. Pelusi, and W. Ding, "A graph based clustering approach for relation extraction from crime data," *IEEE Access*, vol. 7, pp. 101269–101282, 2019.
[48] A. Saxena, M. Prasad, A. Gupta, N. Bharill, O. P. Patel, A. Tiwari, M. J. Er, W. Ding, and C.-T. Lin, "A review of clustering techniques and developments," *Neurocomputing*, vol. 267, pp. 664–681, Dec. 2017.
[49] J. Wu, Z. Lin, and H. Zha, "Essential tensor learning for multi-view spectral clustering," *IEEE Trans. Image Process.*, vol. 28, no. 12, pp. 5910–5922, Dec. 2019.

JOBIN FRANCIS received the B.Tech. degree in electronics and communication engineering and the M.Tech. degree in communication engineering and signal processing from the University of Kannur, Kerala, India, in 2009 and 2015, respectively. He is currently pursuing the Ph.D. degree with the National Institute of Technology Calicut, India. His research interests include computer vision, machine learning, and subspace clustering of high dimensional data.

BABURAJ MADATHIL received the B.Tech. degree in electronics and communication engineering from the University of Calicut, India, in 1999, and the M.Tech. degree in signal processing and the Ph.D. degree from the National Institute of Technology Calicut, India, in 2015 and 2019, respectively. Since 2009, he has been working as an Assistant Professor with the Department of Applied Electronics and Instrumentation Engineering, Government Engineering College, Calicut, India. His research interests include low rank signal modeling and sparse signal processing.

SUDHISH N. GEORGE (Member, IEEE) received the B.Tech. degree in electronics and communication engineering from the M. A. College of Engineering (M. G. University), Kothamangalam, India, in 2004, the M.Tech. degree in signal processing from the College of Engineering Trivandrum (Kerala University), India, in 2007, and the Ph.D. degree in multimedia security from the National Institute of Technology Calicut, India, in 2014. Since 2010, he has been an Assistant Professor with the Department of Electronics and Communication Engineering, National Institute of Technology Calicut. His current research interests include sparse signal processing and computer vision.

SONY GEORGE received the Ph.D. degree in photonics from the Cochin University of Science and Technology, Cochin, India, in 2012. He has been an Associate Professor with The Norwegian Colour and Visual Computing Laboratory, Norwegian University of Science and Technology (NTNU), since 2017. Before joining NTNU, he worked as a Researcher with Gjøvik University College Norway. He has been involved in imaging research in various leading research centers in India and has authored several journal and conference publications. His research interests include color imaging, hyper and multi-spectral imaging, imaging applications in forensics, and cultural heritage. He has been involved in several national and EU projects in multiple roles, including EU MSCA-ITN projects HiPerNav and CHANGE.