

A Novel Spectral-Unmixing-Based Green Algae Area Estimation Method for GOCI Data

Bin Pan, Zhenwei Shi, Zhenyu An, Zhiguo Jiang and Yi Ma

Abstract

Geostationary Ocean Color Imager (GOCI) data have been widely used in the detection and area estimation of green algae blooms. However, due to the low spatial resolution of GOCI data, pixels in an image are usually “mixed”, which means that the region a pixel covers may include many different materials. Traditional index-based methods can detect whether there are green algal blooms in each pixel, whereas it is still challenging to determine the proportion that green algae blooms occupy in a pixel. In this paper, we propose a novel sub-pixel-level area estimation method for green algae blooms based on spectral unmixing, which can not only detect the existence of green algae but also determine their proportion in each pixel. A fast endmember extraction method is proposed to automatically calculate the endmember spectral matrix, and the abundance map of green algae which could be regarded as the area estimation is obtained by nonnegatively constrained least squares. This new fast endmembers extraction technique outperforms the classical N-FINDR method by applying two models: candidates location and distance-based vertices determination. In the first model, we propose a medium-distance-based candidates location strategy, which could reduce the searching space during vertices selection. In the second model, we replace the simplex volume measure with a more simple distance measure, thus complex matrix determinant calculation is avoided. We have theoretically proven the equivalency of volume and distance measure. Experiments on GOCI data and synthetic data demonstrate the superiority of the proposed method compared with some state-of-art approaches.

The work was supported by the National Natural Science Foundation of China under the Grants 61273245, the Beijing Natural Science Foundation under the Grant 4152031, the funding project of State Key Laboratory of Virtual Reality Technology and Systems, Beihang University under the Grant BUAA-VR-16ZZ-03, and the Fundamental Research Funds for the Central Universities under the Grant YWF-14-YHXY-028 and YWF-15-YHXY-003. (*Corresponding author: Zhenwei Shi.*)

Bin Pan is with State Key Laboratory of Virtual Reality Technology and Systems, School of Astronautics, Beihang University, Beijing 100191, China and with Beijing Key Laboratory of Digital Media, Beihang University, Beijing 100191, China and also with Image Processing Center, School of Astronautics, Beihang University, Beijing 100191, China, (e-mail: panbin@buaa.edu.cn).

Zhenwei Shi (Corresponding Author) is with State Key Laboratory of Virtual Reality Technology and Systems, School of Astronautics, Beihang University, Beijing 100191, China and with Beijing Key Laboratory of Digital Media, Beihang University, Beijing 100191, China and also with Image Processing Center, School of Astronautics, Beihang University, Beijing 100191, China, (e-mail: shizhenwei@buaa.edu.cn).

Zhenyu An is with State Key Laboratory of Virtual Reality Technology and Systems, School of Astronautics, Beihang University, Beijing 100191, China and with Beijing Key Laboratory of Digital Media, Beihang University, Beijing 100191, China and also with Image Processing Center, School of Astronautics, Beihang University, Beijing 100191, China, (e-mail: anzhenyu@sa.buaa.edu.cn).

Zhiguo Jiang is with Beijing Key Laboratory of Digital Media, Beihang University, Beijing 100191, China and also with Image Processing Center, School of Astronautics, Beihang University, Beijing 100191, China, (e-mail: jiangzg@buaa.edu.cn).

Yi Ma is with First Institute of Oceanography, State Oceanic Administration, Qingdao 266061, China (e-mail: mayimail@fio.org.cn).

Index Terms

Green Algae Blooms, Area Estimation, Fast Endmember Extraction, Spectral Unmixing

I. INTRODUCTION

Over the last decade, the green algae blooms (mainly *Ulva prolifera*) were frequently observed during April to August in the Yellow Sea and East Sea (China) [1]. Massive *Ulva prolifera* blooms (called green tides) may cause severe harm to tourism, coastal fishery and marine ecosystem. For example, in the summer of 2008, serious green tides disaster occurred around the coastal oceans in Qingdao, China, which have posed direct influence to the Olympic Sailing Games. Consequently, it is of great significance to accurately detect green algae blooms and estimate their area. In 2010, Korea launched the world's first ocean color satellite. It carried Geostationary Ocean Color Imager (GOCI) sensor which could provide multispectral images with 8 bands and 500m spatial resolution [2]. Nowadays GOCI data have been widely used in the detection and monitoring of green algae blooms [3]–[5].

Traditional green algae detection methods are mainly based on certain indexes obtained by bands ratio [1], [3], [6]–[9]. Normalized Difference Vegetation Index (NDVI) [10] and Enhanced Vegetation Index (EVI) [11] are classical vegetation detection algorithms. Based on NDVI and EVI, some modified methods were developed [1], [7], [8]. Hu observed that approaches based on modified NDVI were sensitive to environment and conditions variation [12]. To reduce the potential error, Hu proposed the Floating Algae Index (FAI) to detect floating algae [12]. The Rayleigh-corrected reflectance of Short-Wave Infrared (SWIR) band using Moderate Resolution Imaging Spectroradiometer (MODIS) was utilized by FAI to enhance the detection results. Based on the work of FAI, Zhang *et.al.* proposed the algae pixel-growing algorithm (APA) [13], to estimate the algal bloom coverage in Lake Taihu. In [3], a novel index called index of floating green algae for GOCI (IGAG) was developed, which could discriminate green algae from complex water conditions. To further classify different growth stages of algal blooms, ocean surface algal bloom technique (OSABT) was proposed in [9] (also denoted as OSABI).

The idea of the methods described above is to calculate a certain index for each pixel in the images, and a hard-threshold-based labelling process is followed to determine whether the algae blooms exist. However, the area estimation and the detection of green algae are similar problems, but not identical. The former must estimate the accurate covering area of green algae blooms, whereas the latter only need to confirm whether green algae exists in a pixel. Because of the low resolution of GOCI data, a single pixel would represent as much as $\sim 0.25km^2$ area of ocean. Under this circumstances, only a small proportion of all pixels could completely be covered by green algae blooms. Thus the green algae spectra obtained directly from the images are usually "mixed". Traditional index-based algorithms, which provide binary images as output results, are faced with difficulties in accurately calculating the green algae proportion in a single pixel. In [9] this problem was termed as "sub-pixel problem". Although some detection and classification methods can give a confidence coefficient or probability, the physical meaning is ambiguous. In summary, area estimation of green algae blooms should not be considered as a "detection" or "classification" problem, because detection or classification methods could only determine whether the green algae exist in a pixel, whereas the accurate area is hard to obtain.

To solve this problem, in this paper we proposed a novel green algae area estimation algorithm based on spectral unmixing and fast endmember extraction. Spectral unmixing is a process that decomposes the pixel spectra matrix into two parts: spectral signatures (called *endmembers*) and corresponding fractional *abundances* [14]. Endmembers are generally denoted by the spectra of *pure* materials in an image and the abundances, usually determined by the endmembers, denote the percentages of different endmembers in a pixel. Spectral-unmixing-based methods are able to give an estimation for the abundance of green algae which can be approximately considered as the proportion it covers in each pixel. This intrinsic property makes it possible for spectral unmixing to contribute to the problem of green algae area estimation. Compared with traditional methods, spectral unmixing requires less about the radiometric calibration and atmosphere correction. As is analyzed above, spectral unmixing contains two stages: endmember extraction and abundance calculation. In this paper, when estimating the area of green algae blooms, we first extract the pure materials spectra, such as green algae and clear water, and then the abundance of each endmember is calculated via inversion. Different from detection-based methods, our method does not demand a threshold. We use the green algae abundance as the area estimation so as to achieve more precise results in sub-pixel level.

How to extract the endmembers spectra and how to calculate abundance maps are two key problems in spectral unmixing. Here, we mainly focus on endmember extraction. Researchers have proposed many unsupervised or semisupervised algorithms to solve this problem. Bioucas-Dias *et al.* divided these algorithms into three classes: geometric-based, statistic-based and sparse-regression-based approaches [14]. Statistical approaches formulate spectral unmixing as a parameter estimation problem [15], [16]. Though perform well when the spectral signatures are highly mixed, statistic-based methods encounter high computational complexity [14]. Sparse-regression-based methods usually assume that a mixed spectrum can be approximated by a linear mixture of endmembers spectra. Then the endmember extraction problem is equal to how to select appropriate spectra from a complete library under certain constraints [17], [18]. However, for GOCI data, there is no spectral library available. Geometric-based methods mainly utilize the property that hyperspectral/multispectral data can constitute a high-dimensional convex set. Because of the clear physical meaning and relatively lower computational complexity, geometric-based methods have been widely used in spectral unmixing. Typical methods include N-FINDR [19], pixel purity index (PPI) [20], vertex component analysis (VCA) [21], simplex growing algorithm (SGA) [22], minimum volume simplex analysis (MVSA) [23], [24], simplex identification by variable splitting and augmented lagrangian (SISAL) [25]. Among all these geometric-based methods, N-FINDR is a popular one. N-FINDR is based on the pure pixel assumption, which assumes the existence of pure pixels in the images. The idea of N-FINDR is to find a maximum-volume simplex with all the pixel spectra enclosed, and vertices drawn from the simplex could serve as reliable estimates of endmembers. However, though the computational complexity of N-FINDR is much lower than that of PPI, VCA and most statistic-based algorithms, it still costs a long time when tackling large-size remote sensing images. During each vertex selection process, N-FINDR needs to iterate through all the pixels in the image, and has to conduct complex determinant calculation for each pixel. Researchers have devised many approaches to further improve the efficiency of N-FINDR [22], [26]–[28].

In this paper, we propose a two-step strategy to improve the efficiency of N-FINDR: Medium-distance-based candidates location and geometric-distance-based vertices determination. In the first step, we remove most non-candidate pixels based on the medium distance between pixel vectors and the data center in the whitened data space. In the second step, we replace the simplex volume measure with an equivalent distance measure and thus complex determinant calculation could be avoided.

The major contributions of this paper can be summarized as follows:

- We introduce a novel spectral-unmixing-based method to accurately estimate the area of green algae blooms in sub-pixel level.
- We propose a medium-distance-based candidates location strategy to compress the searching space in N-FINDR algorithm.
- We propose an equivalent distance-measure-based method to improve the efficiency of N-FINDR when determining the simplex vertices, and theoretical proof is provided.

The rest of this paper is organized as follows. In section II we describe some related works for green algae detection and endmember extraction. In section III, we give detailed description of our spectral-unmixing-based green algae area estimation method. The experimental results on GOCI image and synthetic data are shown in Section IV. Section V concludes this paper.

II. RELATED WORKS

Here, we first give a brief introduction to some classical green algae detection methods, and then we present the theoretical foundation about spectral unmixing and N-FINDR.

A. Classical Green Algae Detection Methods

Existing green algae detection methods are mainly based on certain indices, including NDVI [7], EVI [11], IGAG [3], OSABI [9], *etc.* Usually, these methods could be used only in images with atmospheric correction.

NDVI, first proposed in the 1970s, is a traditional vegetation mapping method [10]. In [7], NDVI was used to analyze the offshore extent of floating algae. The basic assumption of NDVI is that all forms of vegetation have a sharp increase in the reflectance spectra near 700nm (also called red edge) [12]. Based on this property, an index can be defined by the following equation:

$$NDVI = \frac{R_{NIR} - R_{RED}}{R_{NIR} + R_{RED}} \quad (1)$$

where R_{NIR} and R_{RED} are Rayleigh-corrected reflectance in near-infrared (NIR) and red bands. For GOCI, R_{NIR} and R_{RED} use reflectance at 865nm and 660nm, respectively. NDVI is a conventional vegetation detection method which has been widely used in practical applications. Based on similar assumption, many improved methods are developed. Consequently, we adopt NDVI as one of the comparison methods in the experiments.

Huete and Justice modified NDVI by proposing EVI, which is defined as:

$$EVI = \frac{G \times (R_{NIR} - R_{RED})}{R_{NIR} + C_1 \times R_{RED} - C_2 \times R_{BLUE} + C_3} \quad (2)$$

where R_{BLUE} for GOCI is 443nm [3]. G , C_1 , C_2 and C_3 are compensation coefficients which are set as $G = 2.5$, $C_1 = 6$, $C_2 = 7.5$ and $C_3 = 1$ for MODIS data [11], and the same coefficients are also used for GOCI data [3]. EVI is a classical variation of NDVI, and it has been used in green algae detection. In this paper, EVI is also discussed in the experiments.

In [3], based on the analysis for the spectral characteristics of green algae from *in situ* optical measurements and satellite data, Son *et al.* developed a green algae detection method specially for GOCI data, named IGAG. The major advantage of the IGAG method is that it can separate muted or subtle signals of floating green algae from surrounding complex water signals [3]. IGAG is defined by:

$$IGAG = \frac{R_{555} + R_{660}}{R_{745} - R_{660}} + \frac{R_{745}}{R_{660}} \quad (3)$$

where R_{555} , R_{660} and R_{745} are reflectance wavelengths of Rayleigh-corrected GOCI data.

To further characterize ocean surface algae blooms, Shanmugam *et al.* proposed a novel index, OSABI, which could not only detect the existence of algae blooms, but determine their growth stages [9]. In OSABI method, the significant increase in aerosol reflectance from NIR to red wavelengths is taken into consideration, and targeted aerosol correction is conducted to remove the aerosol scattering effects. OSABI is defined by the following equation:

$$OSABI = \alpha \times (L_w(\lambda_1) - (L_w(\lambda_2) - L_w(\lambda_1))). \quad (4)$$

For MODIS data, $L_w(\lambda_1)$ and $L_w(\lambda_2)$ are the water-leaving radiance at 748nm and 667nm, and α is a constant set as 10 [9]. For GOCI data, we set λ_1 and λ_2 as 745nm and 660nm, respectively.

NDVI and EVI are two representative vegetation detection methods, and many studies are based on the variation of them. IGAG and OSABI are both recently developed methods for green algae detection. The promising performance shown in [3] and [9] demonstrates that they are one of the state-of-art methods. In this paper, all the above four methods are discussed. However, we note that when detecting algae blooms, the methods described above depend on certain thresholds. Obviously, the selection of thresholds would impact the results to a great extent.

B. Spectral Unmixing

Spectral unmixing refers to the process of decomposing a hyperspectral/multispectral image which is denoted by a mixing model [14]. The most popular model is linear mixing model (LMM) [14]. In LMM, the hyperspectral/multispectral image is considered as a linear combination of endmember signatures. LMM is defined as follows:

$$\mathbf{y} = \mathbf{M}\boldsymbol{\alpha} + \mathbf{n} \quad (5)$$

where $\mathbf{y} \in \mathbb{R}^{L \times 1}$ denotes an observed pixel spectrum, $\mathbf{M} = [\mathbf{m}_1, \mathbf{m}_2, \dots, \mathbf{m}_p] \in \mathbb{R}^{L \times p}$ is the endmember matrix in which each column stands for an endmember spectrum, $\boldsymbol{\alpha} = [\alpha_1, \alpha_2, \dots, \alpha_p]^T \in \mathbb{R}^{p \times 1}$ is the abundance vector, and $\mathbf{n} \in \mathbb{R}^{L \times 1}$ denotes an additive noise. L is the number of bands, and p is the number of endmembers. To be physically meaningful, the fractional abundances should meet the constraint of nonnegative (ANC) and sum-to-one (ASC), which are defined by,

$$\text{ANC} : \alpha_i \geq 0, i = 1, 2, \dots, p \quad (6)$$

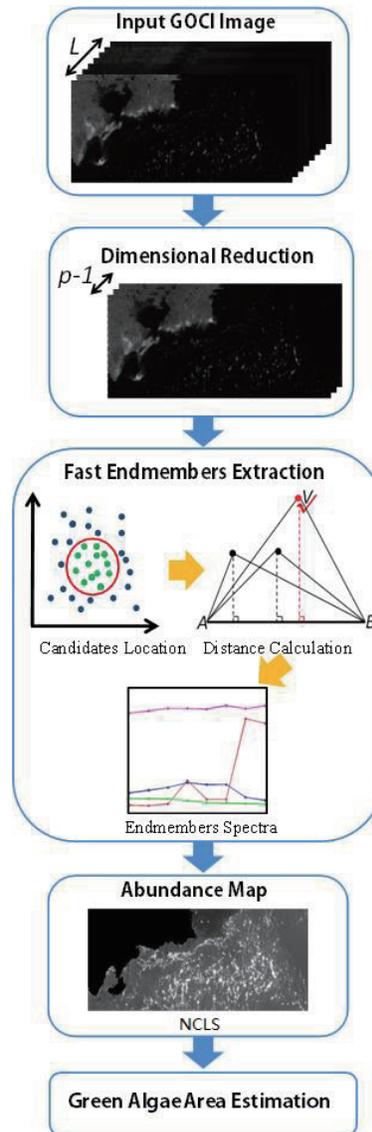


Fig. 1. Flowchart of the proposed method.

$$\text{ASC} : \sum_{i=1}^p \alpha_i = 1 \quad (7)$$

Spectral unmixing contains two steps: endmember extraction and abundance inversion. In this paper, we propose a fast endmember extraction strategy, and then classical nonnegatively constrained least squares (NCLS) [29] is used to obtain the abundance maps. Note that the above two steps can also be implemented simultaneously [14], [30].

C. N-FINDR

In the 1990s, Winter proposed a popular endmember extraction method called N-FINDR [19]. The idea of this algorithm is to find a maximum-volume simplex which could enclose all the observed pixel vectors, and vertices drawn from the simplex are considered as endmembers. The volume of simplex is calculated by

$$\text{vol}(\mathbf{V}) = \frac{1}{(p-1)!} |\det(\begin{bmatrix} \mathbf{1}_p^T \\ \mathbf{V} \end{bmatrix})| \quad (8)$$

where $\mathbf{V} = [\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_p]$ is the vertices of simplex and $\mathbf{1}_p$ is a $p \times 1$ all-one vector. Note that $\mathbf{v}_i \in \mathbf{V}$ is a vector with $p-1$ dimensions. Usually, \mathbf{v}_i is obtained by running a dimension-reduction algorithm on original image data. In this paper, maximum noise fraction (MNF) [31] is used for dimension reduction. MNF is a transformation that would select the bands with highest signal noise ratio (SNR). That means the image noise could be suppressed. Assume the coordinate origin is a known vertex, then we can get the equivalent form of Eq. (8):

$$\text{vol}(\mathbf{V}') = \frac{1}{(p-1)!} \sqrt{\det(\mathbf{V}'^T \mathbf{V}')} \quad (9)$$

where $\mathbf{V}' = [\mathbf{v}_p - \mathbf{v}_1, \mathbf{v}_p - \mathbf{v}_2, \dots, \mathbf{v}_p - \mathbf{v}_{p-1}]$. Vertices that maximizing the volume are regarded as endmembers. The principle of N-FINDR is to exhaustively seek a collection of pixel vectors which could maximize the volume of the simplex. First, simplex vertices are randomly initialized by pixel vectors. And then alternately replace each vertex with all the pixels. If one selected pixel could enlarge the simplex volume, this pixel are considered as a new vertex. Repeat this process, until the volume no longer changes by any replacement. The obtained vertices are the final endmembers. There are two factors that affect the computation cost of N-FINDR. First, when determining each vertex, the whole image data are taken into account. Second, for each pixel vector, complex determinant calculation have to be done. The above processes are time consuming, especially when the image data are quite large.

III. SPECTRAL-UNMIXING-BASED METHOD FOR GREEN ALGAE AREA ESTIMATION

In this section we describe the proposed green algae area estimation method in detail. The flowchart of the whole method is shown in Figure 1. When extracting endmembers, to overcome the difficulties which aggravate the computational complexity of N-FINDR, we propose two efficient algorithms: medium-distance-based candidates location and geometric-distance-based equivalent simplex volume calculation. Once endmember signatures are selected, NCLS is utilized to obtain their abundance maps. Based on the reference green algae spectrum provided by Chinese State Oceanic Administration (CSOA), we use spectral angle mapping (SAM) to choose the green algae spectrum from the extracted endmembers, and the corresponding abundance map could be used to estimate the covering area.

A. Candidates Location

We use the medium distance between all pixels and the data center as the criterion of candidates location. Data centering and whitening [32], [33] are used to preprocess the original image. Data centering is a process of removing the mean value of the data, and data whitening could make data covariance matrix be an identity matrix. Then, in

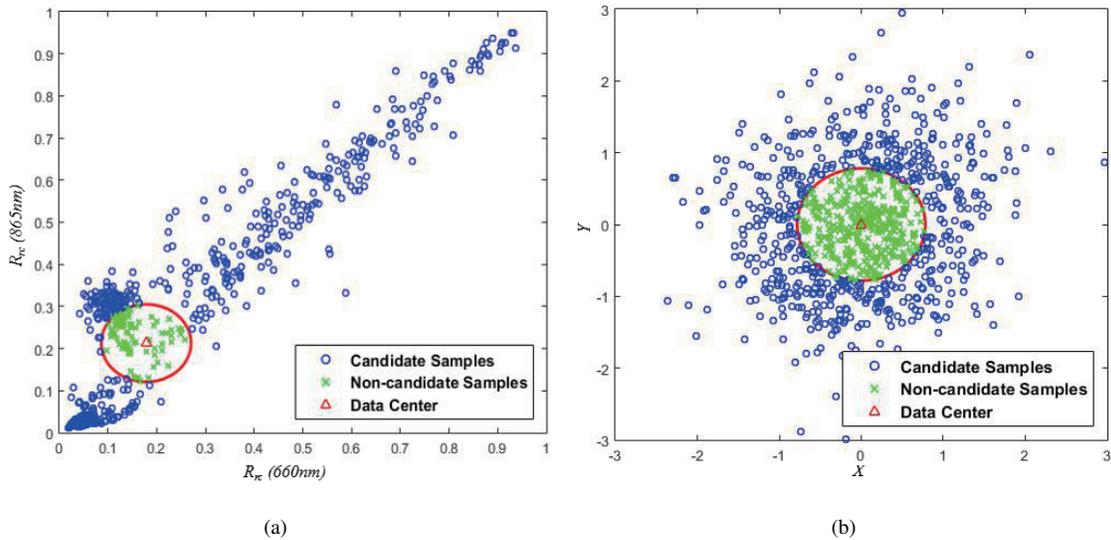


Fig. 2. Candidates selection for (a) data without whitening and (b) data with whitening.

the whitened data space, we calculate the distance of each pixel to the center and, if smaller than their median, those pixels are regarded as non-candidates. In other words, only pixels with larger distance to the center than their median may be candidate vertices. By this means pixels that are likely to be non-candidates are removed, and thus the computational complexity declines. Illustration of candidates location is shown in Figure 2. Data in Figure 2 are randomly selected from GOCI data (collected on 29 June, 2013), totally containing 420 samples. X-axis and Y-axis are the Rayleigh-corrected reflectance value (denoted by R_{rc}) of 660nm and 865nm, respectively. Reflectance is a ratio, and thus the unit is not necessary. As is shown in Figure 2, data whitening is necessary. The data are sphered after whitening, with weaker correlation and isotropic distribution. The process of candidates location could be summarized as follows:

step 1: Input GOCI data $\mathbf{Y} = [\mathbf{y}_1, \mathbf{y}_2, \dots, \mathbf{y}_N] \in \mathbb{R}^{L \times N}$. N is the total number of pixels.

step 2: Data centering. Let $\hat{\mathbf{Y}} = [\hat{\mathbf{y}}_1, \hat{\mathbf{y}}_2, \dots, \hat{\mathbf{y}}_N] \in \mathbb{R}^{L \times N}$ denote the centered data. Then the i^{th} point $\hat{\mathbf{y}}_i$ is calculated by

$$\hat{\mathbf{y}}_i = \mathbf{y}_i - \frac{1}{N} \sum_{k=1}^N \mathbf{y}_k \quad (10)$$

step 3: Data whitening. For $\hat{\mathbf{y}}_i \in \mathbb{R}^{L \times 1}$, its covariance matrix $\mathbf{C} = E\{\hat{\mathbf{y}}_i \hat{\mathbf{y}}_i^T\}$. Let $\mathbf{C} = \mathbf{E} \mathbf{D} \mathbf{E}^T$ denotes the eigenvalue decomposition for \mathbf{C} , then define

$$\mathbf{W} = \mathbf{E} \mathbf{D}^{-\frac{1}{2}} \mathbf{E}^T = \mathbf{C}^{-\frac{1}{2}} \quad (11)$$

where \mathbf{D} is the eigenvalue matrix, \mathbf{E} is the eigenvectors, and \mathbf{W} is the whitening matrices. The whitened vector $\hat{\mathbf{y}}'_i$ can be obtained by

$$\hat{\mathbf{y}}'_i = \mathbf{W} \hat{\mathbf{y}}_i. \quad (12)$$

step 4: Medium-distance calculation and candidates location. In the whitened space, we calculate the medium distance of the data center to each pixel. The medium distance d_m is defined by

$$d_m = \sqrt{\sum_{j=1}^L (\text{medium}(\hat{\mathbf{Y}}'j))^2}, \quad (13)$$

where $\hat{\mathbf{Y}}'$ is the whitened data, j denotes the value in the j^{th} band. Pixels with larger distance than d_m are chosen as candidates.

Since data whitening is a linear transformation of the original data ($\hat{\mathbf{Y}}' = \mathbf{W}\mathbf{Y}$), the simplex vertices is not changed by data whitening. That is, data whitening would never affect the final results of endmember extraction. After candidates location, about half of all pixels are selected as candidates. Note that median distance is not the only criteria that is acceptable to locate candidates. Other distance measure, such as 0.9 times medium distance, is also available.

B. Geometric-Distance-Based Vertices Extraction

In N-FINDR, determinant calculations with size $p \times p$ have to be conducted N times in each vertex selection process. To improve the computation efficiency, we implement a novel geometric-distance-based fast vertices extraction method, and the relationship between the new approach and traditional volume-based method is discussed.

The illustration of our method is shown in Figure 3. Assume A , B and C are vertices having been selected, we intend to find another vertex D which would maximize the volume of the simplex. Obviously, D_1 would generate larger simplex than D_2 , since the distance between D_1 and plane ABC is larger than that of D_2 . Motivated by this theorem, we utilize the distance between each candidate and the hyperplane which is composed by other vertices to substitute the volume calculation of the simplex. Candidate with the largest distance is selected as the vertex in the current searching. The mathematical model of this idea can be expressed as follows

$$\begin{aligned} \arg \max_{\mathbf{v}_i \in \mathbb{R}^{p-1}} \text{dis}\{\mathbf{v}_i, \text{span}(\mathbf{v}_1, \dots, \mathbf{v}_{i-1}, \mathbf{v}_{i+1}, \dots, \mathbf{v}_p)\} \\ \text{s.t. } \mathbf{v}_i \in \mathbf{Y}_C \end{aligned} \quad (14)$$

where \mathbf{v}_i is an estimated vertex of the simplex, span denotes the subspace spanned by vertices, $\text{dis}\{\cdot\}$ is a function calculating the distance between vertex and hyperplane, and \mathbf{Y}_C are the candidate data set obtained in Section III-A. When extracting \mathbf{v}_i , we should first determine the hyperplane \mathbf{w} , which is expressed by

$$\mathbf{w}^T \mathbf{v}' + 1 = 0, \quad (15)$$

$$\mathbf{v}' \in \{\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_{i-1}, \mathbf{v}_{i+1}, \dots, \mathbf{v}_p\}$$

and

$$\mathbf{w} = -(\mathbf{V}'^T)^{-1} \mathbf{1}_{p-1}^T, \quad (16)$$

$$\mathbf{V}' = [\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_{i-1}, \mathbf{v}_{i+1}, \dots, \mathbf{v}_p]$$

where Eq. (16) is the solution of Eq. (15). Dimension reduction process can make sure that \mathbf{V}' is a phalanx. Generally, \mathbf{V}' is reversible because the bands number of GOCI data is 8, which is much higher than the number

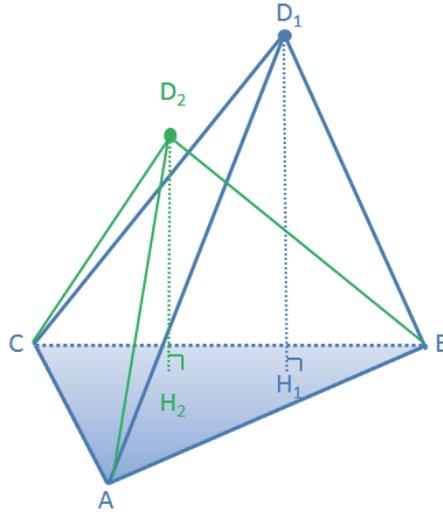


Fig. 3. Illustration for distance-based vertices extraction.

of endmembers. And even if \mathbf{V}' is irreversible, pseudo-inverse is also available. Here the final solution is the least squares minimum norm solution. Thus the distance function in Eq. (14) can be written by

$$dis\{\mathbf{v}_i, span(\mathbf{v}_1, \dots, \mathbf{v}_{i-1}, \mathbf{v}_{i+1}, \dots, \mathbf{v}_p)\} = \frac{\mathbf{w}^T \mathbf{v}_i + 1}{\|\mathbf{w}^T\|} \quad (17)$$

The pixel which maximizes the distance in Eq. (17) is chosen as a vertex. In our algorithm, initialization is conducted by random sampling from the image data.

In our method, the most heavy computational burden is $(\mathbf{V}'^T)^{-1}$. However, during each searching process for an endmember, $(\mathbf{V}'^T)^{-1}$ only needs to compute once. N-FINDR, by comparison, has to calculate the matrix determinant for each candidates, totally N times. In this way the computation complexity of our method significantly declines.

Next, we give theoretical proof for the following theorem:

THEOREM 1. *Under ANC, ASC, pure pixel assumption and the same initialization, distance-based and volume-based measure lead to the same vertices extraction results.*

Proof for theorem 1: Let $\mathbf{V} = [\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_p]$ be vertices obtained by N-FINDR, we prove the following conclusion: In the process of extracting \mathbf{v}_i , the distance-based and volume-based measure would achieve the same result. In order to simplify the computation, we assume the coordinates origin is also a vertex. Then Eq. (15) could be rewritten as

$$\mathbf{w}^T \mathbf{v}' = 0, \mathbf{v}' \in \{\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_{i-1}, \mathbf{v}_{i+1}, \dots, \mathbf{v}_p\} \quad (18)$$

This simplification strategy will not change the final obtained vertices, since it is equal to data space translation. For a pixel vector $\mathbf{u} \in \mathbf{V}$, the distance between \mathbf{u} and hyperplane \mathbf{w} is defined by

$$dis(\mathbf{u}) = \frac{|\mathbf{w}^T \mathbf{u}|}{\|\mathbf{w}\|} \quad (19)$$

According to Eq. (5), \mathbf{u} can be expressed by

$$\mathbf{u} = \sum_{k=1}^p \alpha_k \mathbf{v}_k \quad (20)$$

Put Eq. (20) into Eq. (19), we get

$$\begin{aligned} dis(\mathbf{u}) &= \frac{|\mathbf{w}^T \mathbf{u}|}{\|\mathbf{w}\|} = \frac{|\mathbf{w}^T \sum_{k=1}^p \alpha_k \mathbf{v}_k|}{\|\mathbf{w}\|} = \frac{|\sum_{k=1}^p \alpha_k \mathbf{w}^T \mathbf{v}_k|}{\|\mathbf{w}\|} \\ &= \frac{|\alpha_1 \mathbf{w}^T \mathbf{v}_1| + |\alpha_2 \mathbf{w}^T \mathbf{v}_2| + \dots + |\alpha_p \mathbf{w}^T \mathbf{v}_p|}{\|\mathbf{w}\|} \\ &= \frac{|\alpha_i \mathbf{w}^T \mathbf{v}_i|}{\|\mathbf{w}\|} \end{aligned} \quad (21)$$

When ANC, ASC and pure pixel assumption are satisfied, to maximize Eq. (21), the abundance α_i should be 1. In this condition, the selected vertex is exactly \mathbf{v}_i which is also the one obtained by N-FINDR. That is, distance-based measure achieves the same result as volume-based measure. The proof of Theorem 1 is thus completed. \square

In real scenarios, the density and growth stages of green algae blooms also affect the observed spectra [9]. Obviously, green algae blooms of maturity stage are more dense than that of initial stage. A region completely covered by maturity or initial stage green algae blooms may present various spectra. This is mainly because the influence of nonlinear mixture [14], [34]. However, this spectral variation is little, when compared with the spectral difference between an endmember and another. In the data simplex, dense and less dense green algae pixels would both appear around one of the vertices. In this paper we ignore this difference to simplify the unmixing process.

C. Green Algae Determination and Abundance-Based Area Estimation

When endmember signatures are determined, the abundance maps can be obtained by many classical algorithms [14]. In this paper, NCLS [29] is used. Specially, SAM is utilized to identify which is the green algae spectrum among all the endmember spectra. The spectral angle between two spectra is defined by

$$\theta = \arccos \frac{\sum_{i=1}^L t_i r_i}{\left(\sum_{i=1}^L t_i\right)^{\frac{1}{2}} \left(\sum_{i=1}^L r_i\right)^{\frac{1}{2}}} \quad (22)$$

where $\mathbf{t} = [t_1, t_2, \dots, t_L]^T$ is one of an endmember spectrum, and $\mathbf{r} = [r_1, r_2, \dots, r_L]^T$ is the reference spectrum. We use the green algae spectra provided by CSOA as the reference spectrum. For each extracted endmember, we calculate its spectral angles with the reference spectrum, and the one with the minimum value is considered as the green algae endmember. The SAM algorithm may perform poor when the number of matching targets is huge. However, since GOCI data have only 8 bands, the amount of extracted endmembers is not more than 7 (usually 3 to 5). In this case the accuracy of SAM is guaranteed and, even the reference spectrum and the extracted green algae spectrum are not exactly the same, we can also achieve satisfying results because we only need to select endmember most similar to the reference.

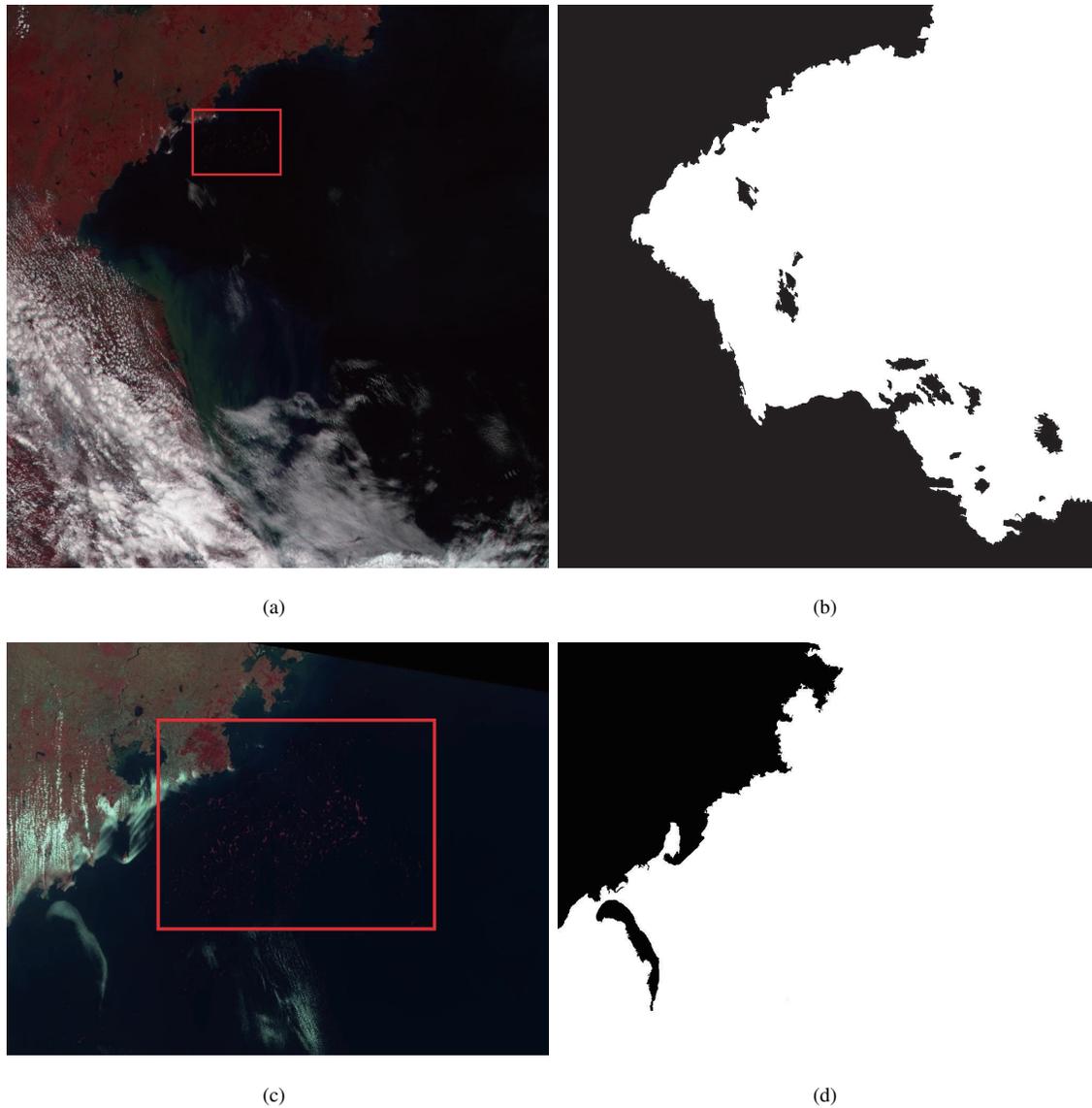


Fig. 4. GOCI and HJ-1B data used in this paper. (a) A false composite GOCI image (R-G-B=865-660-555nm). (b) Labeled GOCI data without land and cloud. (c) A false composite HJ-1B image (R-G-B=830-660-560nm). (d) Labeled HJ-1B data without land and cloud.

When green algae endmember is determined, we use the corresponding abundance map to estimate the real covering area of green algae blooms in the ocean. The area is calculated as follows

$$Area = Res^2 \sum_{i=1}^N a_i \quad (23)$$

where a_i denotes the green algae abundance of the i th pixel in the image, and Res is the resolution of GOCI data (~ 500 m).

IV. EXPERIMENTS AND DISCUSSION

Here, we use both real-world GOCI data and synthetic data to test the performance of the proposed algorithm. Our method was compared with some state-of-art algorithms, including NDVI [7], EVI [11], [12], IGAG [3] and OSABT [9]. In related literatures, the thresholds for NDVI, EVI and IGAG are all set as 0 [3], [7], [11], [12]. For OSABT, the set of threshold is concerned with the growth stages of green algae, usually 0~4 [9]. In our experiments, since we aim to estimate the accurate area of green algae blooms rather than simply detected their existence, besides reporting the estimation results based on thresholds given by literatures, we also slightly increase the thresholds, as shown in Table II and Table IV. In addition, we compared our endmember extraction process with traditional N-FINDR [19] and an improved method, SGA [22], in extraction results and computation complexity. For GOCI data, we mask out land and cloud regions manually. The number of endmembers in test GOCI data, which is set as 4, is determined by HySime [35]. Since determining the number of endmembers is a prior step before spectral unmixing, N-FINDR, SGA and our method share the same setting. After extracting endmembers, the abundance of each material is calculated by NCLS. The algorithms were implemented using Matlab 2013a on a desktop PC equipped with an Intel i7-6700K CPU (at 4.0 GHz) and 32 GB of RAM memory.

TABLE I
GENERAL CHARACTERISTICS OF GOCI AND HJ-1B DATA.

| | GOCI | HJ-1B |
|------------------------|---------|---------|
| Spatial Resolution (m) | 500 | 30 |
| Revisiting Period (h) | 1 | 96 |
| Digitization (bit) | 12 | 8 |
| Band 1 (nm) | 402-422 | 430-520 |
| Band 2 (nm) | 433-453 | 520-600 |
| Band 3 (nm) | 480-500 | 630-690 |
| Band 4 (nm) | 545-565 | 760-900 |
| Band 5 (nm) | 650-670 | - |
| Band 6 (nm) | 675-685 | - |
| Band 7 (nm) | 735-755 | - |
| Band 8 (nm) | 845-885 | - |

TABLE II
AREA ESTIMATION OF GREEN ALGAE BLOOMS BY DIFFERENT METHODS (GOCI DATA).

| | Groundtruth | | Our Method | NDVI | | EVI | | IGAG | | OSABT | | |
|------------------------|-------------|------|------------|------|------|------|------|-------|-----|-------|-----|-----|
| Threshold | >0 | >0.1 | - | >0 | >0.1 | >0 | >0.1 | >0 | >2 | >0 | >1 | >2 |
| Area(km ²) | 317 | 279 | 299 | 2527 | 1331 | 2527 | 427 | 38752 | 739 | 38563 | 553 | 192 |

A. Experiments on GOCI Data

The GOCI data used in this paper were collected on 12:16, June 29, 2013, Beijing time, which can be downloaded from the Korea Ocean Satellite Center (<http://kosc.kordi.re.kr/>). In theory, Rayleigh scattering and aerosol effects should be removed by atmospheric correction beforehand. However, Son *et al.* [3] observed that for GOCI data, if standard atmospheric correction was used to remove the aerosol reflectance, some green algae may be mistaken as aerosol signal. In addition, researches have shown that Rayleigh signals were dominant in general atmospheric conditions [3], [8], [12]. In our experiments, we use the Rayleigh-corrected reflectance data to derive green algae bloom indices and verify our spectral-unmixing-based algorithm. The same strategy was also used in [3]. We used the GOCI Data Processing System (GDPS) [2] which was specially designed for GOCI data to perform Rayleigh correction. A false color composite image was shown in Figure 4(a) (R-G-B=865-660-555nm). Green algae blooms approximately distributed in the box region according to visual observation. Note that visual observation is not involved in the quantitative evaluation for the experimental results. It could only provide an intuitive feeling about the distribution of green algal blooms. In our experiments, we mainly focus on the area estimation of green algae blooms in sub-pixel level. To highlight this point, we manually removed land and cloud regions. The manually segmentation result was presented in Figure 4(b), where only the white region was used in green algae area estimation. Note that the image shown in Figure 4(a), which is bound by 118.03°E to 124.54°E and 31.71°N to 37.20°N with 1175×1219 pixels size, is part of the original GOCI data.

Due to the extensive distribution of green algae blooms in the ocean, determining the real green algae covering area via field observations is almost an impossible work. In this paper, to obtain a relatively accurate groundtruth, we utilized the HJ-1B data which was collected in the same date, close time and similar region with GOCI. HJ-1B data used in this paper (it could be downloaded from China Center For Resources Satellite Data and Application, <http://218.247.138.121/DSSPlatform/productSearch.html>) were collected on 11:13, June 29, 2013, covering 119.53°E to 122.32°E and 34.98°N to 36.63°N with 8256×6287 pixels size. General characteristics of GOCI and HJ-1B data are exhibited in Table I. HJ-1B data only have 4 bands, but their spatial resolution (30m) is much higher than that of GOCI. Figure 4(c) shows the false color composite image (R-G-B=830-660-560nm), and green algae blooms mainly distribute in the box region. This image is a 2-level product, with atmospheric and geometric correction completed. Similarly, in Figure 4(d) we have removed land and cloud regions. Because HJ-1B image covered less region than GOCI data, we first implement registration for the two image based on geographic information. Only the region both covered by GOCI and HJ-1B data is used in our experiments to estimate the green algae area. To refine the registration result, slight manual modification was followed. The registration process did not need to quite exact, since the green algae blooms were mainly observed in the middle of the images. Because the HJ-1B and GOCI images were collected on close time, we could assume that green algae blooms covered the same area in the above two images. We run NDVI algorithm in HJ-1B data to determine the area of green algae blooms (using 0 as the threshold), and the result was considered as the groundtruth for GOCI data. We adopted this strategy mainly based on the following reasons:

- Labeling the green algae pixels via visual interpretation in GOCI data is influenced by many subjective factors. Most of all, it is still challenging for visual interpretation to determine the green algae covering area in sub-pixel level.
- As is discussed above, in 500m-resolution GOCI data, the area estimation results by NDVI may be larger than the real value owing to the existence of mixed pixels. However, HJ-1B data have 30m resolution. Under this resolution, the spectra mixing problem is highly suppressed. NDVI algorithm could achieve satisfying results.
- Some novel green algae detection methods, such as IGAG and OSABT, have to use the 745nm band which is not included in HJ-1B data. By comparison, NDVI is a more popular and commonly used method.

Figure 5(a-f) illustrate the green algae distribution maps of different methods. All these maps cover the same region. Figure 5(a) shows the detection result for HJ-1B data based on NDVI, which is regarded as the real green algae distribution. Figure 5(b) shows the abundance map obtained by our proposed method, and Figure 5(c-f) present distribution maps by NDVI, EVI, IGAG and OSABT, respectively. Table II gives the quantitative area estimation results for green algae blooms. Figure 5 demonstrates that all the algorithms could approximately obtain the distribution of green algae blooms. However, compared with the groundtruth, the area estimation by NDVI (Figure 5(c)) is obviously larger. This result indicates that for 500m-resolution GOCI data, NDVI could only detect whether green algae exists rather than accurately estimate their covering area. EVI performs better when estimating the distribution situation of green algae blooms (Figure 5(d)), however, as is displayed in Table II, if 0 is set as the threshold, EVI obtains the same result as NDVI. IGAG algorithm could exactly detect muted and subtle green algae signals [3], however, due to the sub-pixel problem, IGAG performs poor when estimating the covering area (Figure 5(e) and Table II). OSABT could not only detect the existence of green algae blooms, but also classify different growth stages (Figure 5(f)). Therefore, the final results are strongly impacted by the setting of threshold. Here, we assume that for green algae blooms in each pixel, growth stages could be represented by the covering area. However, Table II shows that if we use the thresholds given in reference literatures, area estimation results obtained by all the traditional green algae detection methods are much larger than the groundtruth. Meanwhile, traditional methods are sensitive to thresholds changing in GOCI data. Note that thresholds changing has little impact on HJ-1B data. This is mainly because HJ-1B data have higher resolution, and in this case materials mixing is much weaker. Figure 5 and Table II demonstrate that NDVI, EVI and IGAG may lead to larger area estimation results than the groundtruth in GOCI data. OSABT can achieve satisfying performance by setting suitable threshold. However, the threshold choosing requires abundant experiential and experimental support. Moreover, as is discussed above, even a little disturbance of threshold may cause remarkable results variation. In summary, for GOCI data, traditional methods could approximately obtain the distribution region of green algae blooms, but perform poor in estimating the accurate area. This is because green algae blooms could hardly completely cover a pixel region in GOCI data, and pixel-level area estimation is likely to cause larger results.

The proposed method aims at accurately estimating the green algae area in 500m-resolution GOCI data. Figure 5(b) shows the distribution of green algae obtained by the proposed method, which is quite similar to the groundtruth.

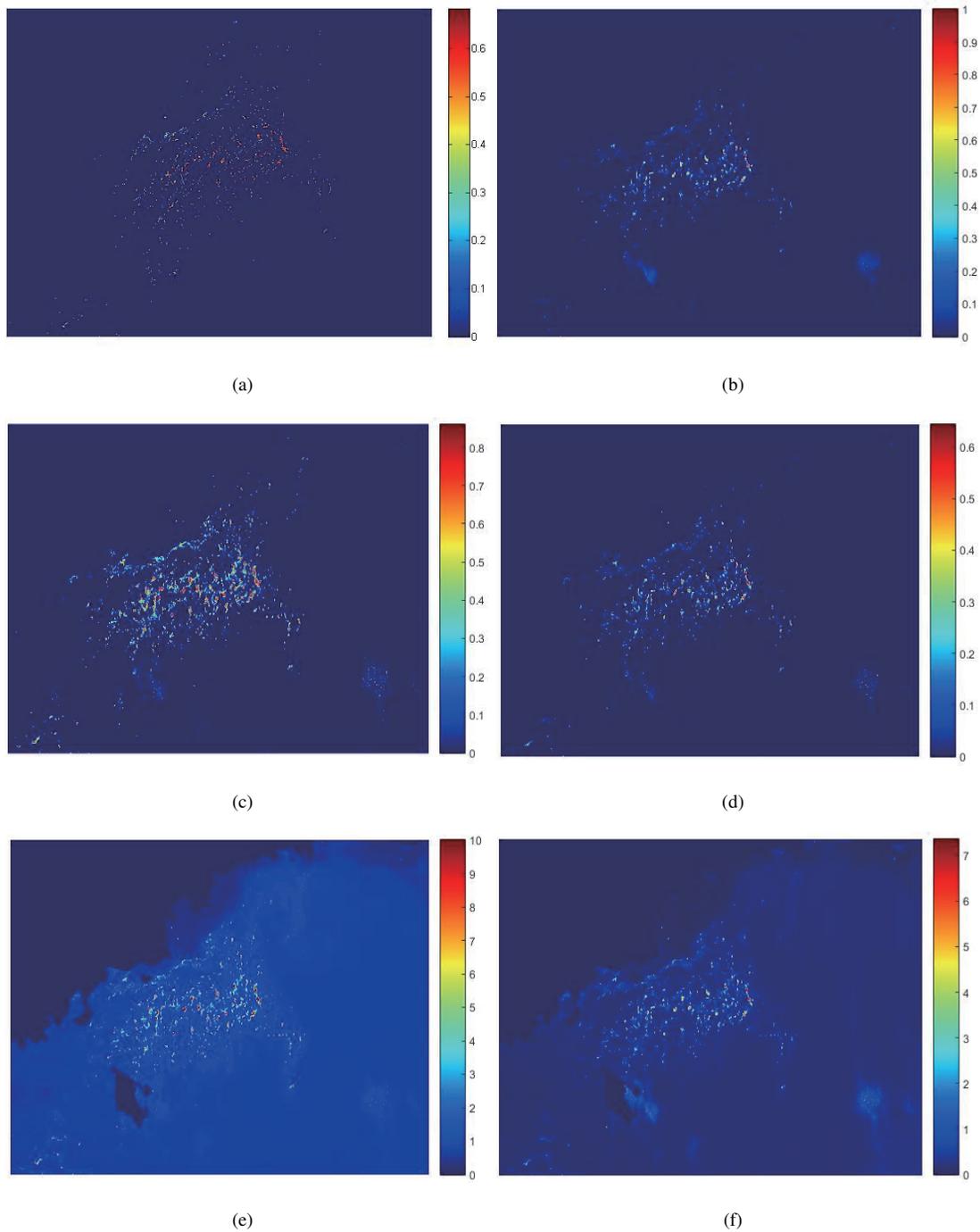


Fig. 5. Evaluation for green algae blooms distribution. (a) NDVI on HJ-1B data (Regarded as the groundtruth). (b) Green algae abundance map by the proposed method on GOCI data.(c) NDVI on GOCI data. (d) EVI on GOCI data. (d) IGAG on GOCI data. (f) OSABT on GOCI data.

In a single pixel, the covering area of green algae is described by the corresponding abundance. Moreover, in the broad ocean field that do not contain green algae, the abundance map tends to 0. Table II quantifies the area estimation results, which is the average result by 5 different initializations. We could see that our method is the

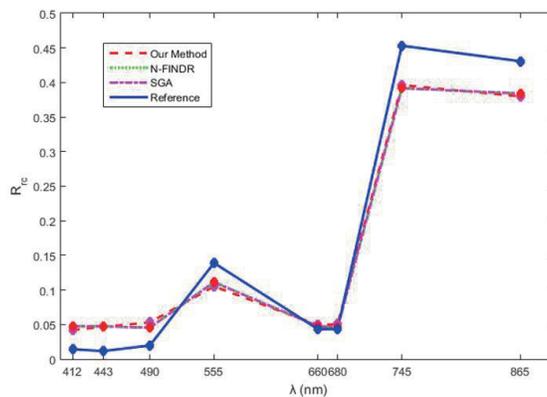


Fig. 6. Green algae spectral signatures obtained by N-FINDR, SGA, our method and the reference

closest to the groundtruth. In our result, we totally find 10296 pixels with green algae abundance greater than 0.01, and 19 of them present higher abundance than 0.9. This result may confirm the assumption that in most cases green algae blooms cannot cover the whole region a pixel occupies. Another advantage of our method is that few parameters are required. Especially, our approach is immune to threshold, and this property makes it a more robust algorithm. The comparison experiments indicate for GOCI data, our method could accurately estimate the area of green algae blooms in sub-pixel level.

Figure 6 shows the green algae spectrum extracted by the proposed method, N-FINDR, SGA and the reference. The spectrum extracted by our method is almost overlapped with that by N-FINDR, and this result also verifies the equivalence of the two methods. The reference spectrum is not extracted directly from our test GOCI data, therefore slight distinction is observed. Though not completely the same, we could still determine the real green algae endmember by SAM. Table III shows the processing time of endmember extraction by N-FINDR, SGA and our method. All the sample pixels are randomly selected from the test image (with endmembers number set as 4). Compared with N-FINDR and SGA, our method only takes about 1/10 and 1/2 processing time, respectively.

TABLE III
COMPUTATION TIME OF ENDMEMBERS EXTRACTION BY N-FINDR, SGA AND OUR METHOD (GOCI DATA).

| Pixel Number | N-FINDR(s) | SGA(s) | Our Method(s) |
|--------------|------------|--------|---------------|
| 20301 | 1.88 | 0.33 | 0.29 |
| 54876 | 6.84 | 1.31 | 0.61 |
| 111341 | 17.19 | 3.35 | 1.59 |
| 152153 | 25.63 | 4.87 | 2.23 |
| 300238 | 52.31 | 10.02 | 4.77 |

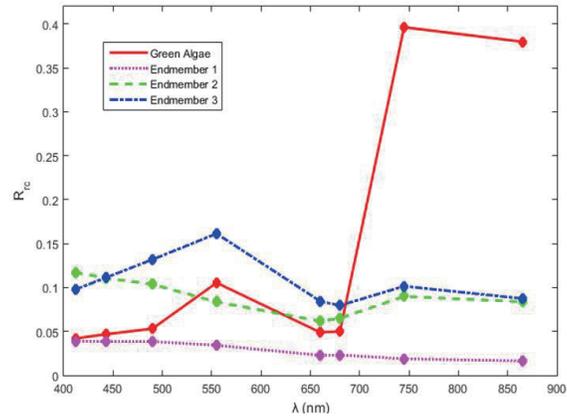


Fig. 7. Endmembers spectra extracted from GOCI data.

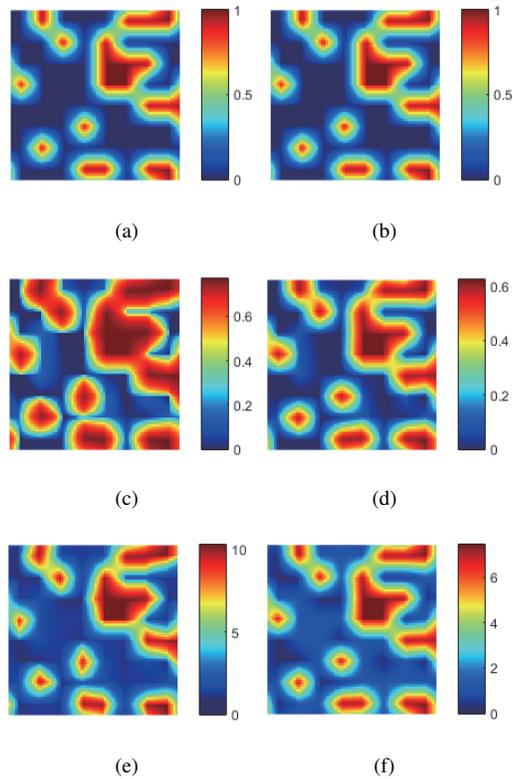


Fig. 8. Evaluation for green algae blooms distribution on synthetic data ($z = 10$). (a) The groundtruth. (b) Abundance map by the proposed method. (c) NDVI (d) EVI (e) IGAG (f) OSABT

B. Experiments on Synthetic Data

In this section, to further verify that our spectral-unmixing-based method could give more accurate green algae area estimation in sub-pixel level, we designed a series of synthetic data experiments. The synthetic data are created as follows:

- We use the endmembers extracted from the previous real scene experiment as the synthetic experiments endmembers matrix (size 8×4). Spectral signatures used in synthetic experiments are shown in Figure 7.
- We create a scene with size $z^2 \times z^2$, and divide it into $z \times z$ patches. Note that the parameter z determines the size of synthetic images. Initialize each patch with a single endmember (abundance equal to 1). Similar processes are also used in [36]–[39].
- Run a $(z+1) \times (z+1)$ spatial low-pass filter to generate mixed pixels, i.e., the abundance maps of endmembers.
- Obtain the final synthetic images based on LMM, and add Gaussian white noise (SNR=30dB) simultaneously.

Figure 8 displays the true abundance map as well as the distribution maps estimated by different models (z is set as 10). Though under different dimensions, all the green algae coverage maps are close to the groundtruth. That is to say both the traditional and our methods perform well in determining the distribution of green algae blooms. NDVI and EVI achieved similar results (Figure 8(c)(d)), while for IGAG and OSABT, many regions with values slightly exceeding 0 are observed. As a result the area estimation results by IGAG and OSABT are much larger than the groundtruth (Figure 8(e)(f)). By contrast, our method performs better. The green algae abundance map obtained by our method is nearly the same as the groundtruth. In Table IV, quantitative statistics results are displayed, where parameter z corresponds to image size. It is observed that in synthetic data, the results of NDVI, EVI, IGAG and OSABT are less sensitive to thresholds than that in real-world data. The reason may be that the environmental conditions in synthetic data are ideal, and moreover, the materials spectral characteristic are not complex. However, despite all that, traditional green algae detection methods still present larger estimation results. When mixed pixels exist, traditional methods cannot accurately estimate the area of green algae. We note that in real-world GOCI data experiments, similar conclusion could also be inferred. This result verifies the rationality of our synthetic data experiments. According to Figure 8(b) and Table IV, our method works well. Slight errors are still observed in our results, however, the index-based methods' intrinsic problem of higher estimation for the green algae blooms is overcome. The green algae area is denoted alternatively by the abundance map and, even existing spectral mixing, we are able to achieve closest results to the groundtruth.

TABLE IV
AREA EVALUATION OF GREEN ALGAE BLOOMS BY DIFFERENT METHODS (SYNTHETIC DATA).

| Groundtruth | Our Method | NDVI | | EVI | | IGAG | | OSABT | | |
|-------------------|------------|--------|--------|--------|--------|--------|--------|--------|--------|--------|
| | | >0 | >0.1 | >0 | >0.1 | >0 | >2 | >0 | >1 | >2 |
| – | – | | | | | | | | | |
| 2499($z = 10$) | 2523 | 8646 | 6557 | 8646 | 4558 | 10000 | 4585 | 10000 | 7268 | 3519 |
| 8401($z = 14$) | 8516 | 33673 | 25129 | 33673 | 18270 | 38416 | 18196 | 38416 | 28165 | 14868 |
| 26770($z = 18$) | 26204 | 93205 | 71538 | 93205 | 55175 | 104976 | 54580 | 104976 | 77537 | 45517 |
| 17971($z = 30$) | 180103 | 678125 | 510254 | 678125 | 401235 | 810000 | 391587 | 810000 | 568513 | 321510 |

V. CONCLUSION

In this paper, we propose a novel green algae area estimation method based on spectral unmixing. The idea of spectral unmixing is utilized to estimate the green algae area in sub-pixel level. To reduce the computational complexity of N-FINDR, we propose two efficient strategies, namely medium-distance-based candidates location and geometric-distance-based simplex vertices extraction methods. The former reduces the searching space by removing massive pixels that are considered as non-vertices, and the latter transfers the simplex volume computation to an equivalent distance measure. Compared with traditional green algae detection approaches, our algorithm could not only detect the existence of green algae but also accurately estimate their covering area. The effectiveness of our method has been verified in real-world GOCI data and synthetic data. In general, spectral unmixing methods are more useful when endmember materials are widely distributed in an entire image scene. However, in GOCI data, the ocean occupies most of the image, which may lead to uncertain results. In addition, spectral variability in real-world data also affects unmixing accuracy. Further studies will be devoted to design a more robust and efficient abundance inversion algorithm under such a circumstance.

VI. ACKNOWLEDGEMENT

The authors would like to thank the Editor-in-Chief and the Associate Editor who handled our paper for their outstanding. We would also like to thank the three anonymous reviewers for providing many insightful comments and suggestions that significantly helped us improve the technical quality and presentation of our paper.

REFERENCES

- [1] C. Hu, D. Li, C. Chen, J. Ge, F. E. Muller-Karger, J. Liu, F. Yu, and M.-X. He, "On the recurrent *Ulva prolifera* blooms in the Yellow Sea and East China Sea," *Journal of Geophysical Research: Oceans (1978–2012)*, vol. 115, no. C5, 2010.
- [2] J.-H. Ryu, H.-J. Han, S. Cho, Y.-J. Park, and Y.-H. Ahn, "Overview of geostationary ocean color imager (GOCI) and goci data processing system (GDPS)," *Ocean Science Journal*, vol. 47, no. 3, pp. 223–233, 2012.
- [3] Y. B. Son, J.-E. Min, and J.-H. Ryu, "Detecting massive green algae (*Ulva prolifera*) blooms in the Yellow Sea and East China sea using geostationary ocean color imager (GOCI) data," *Ocean Science Journal*, vol. 47, no. 3, pp. 359–375, 2012.
- [4] X. Lou and C. Hu, "Diurnal changes of a harmful algal bloom in the East China Sea: Observations from GOCI," *Remote Sensing of Environment*, vol. 140, pp. 562–572, 2014.
- [5] Y. B. Son, B.-J. Choi, Y. H. Kim, and Y.-G. Park, "Tracing floating green algae blooms in the Yellow Sea and the East China Sea using GOCI satellite data and lagrangian transport simulations," *Remote Sensing of Environment*, vol. 156, pp. 21–33, 2015.
- [6] P. Shanmugam, Y.-H. Ahn, and S. Sanjeevi, "A comparison of the classification of wetland characteristics by linear spectral mixture modelling and traditional hard classifiers on multispectral remotely sensed imagery in southern India," *Ecological Modelling*, vol. 194, no. 4, pp. 379–394, 2006.
- [7] C. Hu and M.-X. He, "Origin and offshore extent of floating algae in Olympic sailing area," *Eos, Transactions American Geophysical Union*, vol. 89, no. 33, pp. 302–303, 2008.
- [8] W. Shi and M. Wang, "Green macroalgae blooms in the Yellow Sea during the spring and summer of 2008," *Journal of Geophysical Research: Oceans (1978–2012)*, vol. 114, no. C12, 2009.
- [9] P. Shanmugam, M. Suresh, and B. Sundarabalan, "OSABT: An innovative algorithm to detect and characterize ocean surface algal blooms," *Selected Topics in Applied Earth Observations and Remote Sensing, IEEE Journal of*, vol. 6, no. 4, pp. 1879–1892, 2013.
- [10] J. Rouse Jr, R. Haas, J. Schell, and D. Deering, "Monitoring vegetation systems in the Great Plains with ERTS," *NASA special publication*, vol. 351, pp. 309–317, 1973.

- [11] A. Huete, C. Justice, and W. Van Leeuwen, "MODIS vegetation index (MOD13) algorithm theoretical basis document. Greenbelt: NASA Goddard Space Flight Centre," 1999.
- [12] C. Hu, "A novel ocean color index to detect floating algae in the global oceans," *Remote Sensing of Environment*, vol. 113, no. 10, pp. 2118–2129, 2009.
- [13] Y. Zhang, R. Ma, H. Duan, S. A. Loiselle, J. Xu, and M. Ma, "A novel algorithm to estimate algal bloom coverage to subpixel resolution in lake taihu," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 7, no. 7, pp. 3060–3068, 2014.
- [14] J. M. Bioucas-Dias, A. Plaza, N. Dobigeon, M. Parente, Q. Du, P. Gader, and J. Chanussot, "Hyperspectral unmixing overview: Geometrical, statistical, and sparse regression-based approaches," *Selected Topics in Applied Earth Observations and Remote Sensing, IEEE Journal of*, vol. 5, no. 2, pp. 354–379, 2012.
- [15] S. Moussaoui, C. Carteret, D. Brie, and A. Mohammad-Djafari, "Bayesian analysis of spectral mixture data using Markov chain Monte Carlo methods," *Chemometrics and Intelligent Laboratory Systems*, vol. 81, no. 2, pp. 137–148, 2006.
- [16] M. Arngren, M. N. Schmidt, and J. Larsen, "Unmixing of hyperspectral images using Bayesian non-negative matrix factorization with volume prior," *Journal of Signal Processing Systems*, vol. 65, no. 3, pp. 479–496, 2011.
- [17] W. Tang, Z. Shi, Y. Wu, and C. Zhang, "Sparse unmixing of hyperspectral data using spectral a priori information," *Geoscience and Remote Sensing, IEEE Transactions on*, vol. 53, no. 2, pp. 770–783, 2015.
- [18] Z. Shi, W. Tang, Z. Duren, and Z. Jiang, "Subspace matching pursuit for sparse unmixing of hyperspectral data," *Geoscience and Remote Sensing, IEEE Transactions on*, vol. 52, no. 6, pp. 3256–3274, 2014.
- [19] M. E. Winter, "N-FINDR: an algorithm for fast autonomous spectral end-member determination in hyperspectral data," in *SPIE's International Symposium on Optical Science, Engineering, and Instrumentation*, pp. 266–275, International Society for Optics and Photonics, 1999.
- [20] J. W. Boardman, F. A. Kruse, and R. O. Green, "Mapping target signatures via partial unmixing of AVIRIS data," in *Proc. JPL airborne earth sci. workshop*, vol. 1, pp. 23–26, 1995.
- [21] J. M. Nascimento and J. M. B. Dias, "Vertex component analysis: A fast algorithm to unmix hyperspectral data," *Geoscience and Remote Sensing, IEEE Transactions on*, vol. 43, no. 4, pp. 898–910, 2005.
- [22] C.-I. Chang, C.-C. Wu, W.-m. Liu, and Y.-C. Ouyang, "A new growing method for simplex-based endmember extraction algorithm," *Geoscience and Remote Sensing, IEEE Transactions on*, vol. 44, no. 10, pp. 2804–2819, 2006.
- [23] J. Li and J. M. Bioucas-Dias, "Minimum volume simplex analysis: A fast algorithm to unmix hyperspectral data," in *Geoscience and Remote Sensing Symposium, 2008. IGARSS 2008. IEEE International*, vol. 3, pp. 250–253, IEEE, 2008.
- [24] J. Li, A. Agathos, D. Zaharie, J. M. Bioucas-Dias, A. Plaza, and X. Li, "Minimum volume simplex analysis: A fast algorithm for linear hyperspectral unmixing," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 53, no. 9, pp. 5067–5082, 2015.
- [25] J. M. Bioucas-Dias, "A variable splitting augmented lagrangian approach to linear spectral unmixing," in *Hyperspectral Image and Signal Processing: Evolution in Remote Sensing, 2009. WHISPERS'09. First Workshop on*, pp. 1–4, IEEE, 2009.
- [26] T.-H. Chan, W.-K. Ma, A. Ambikapathi, and C.-Y. Chi, "Robust endmember extraction using worst-case simplex volume maximization," in *Hyperspectral Image and Signal Processing: Evolution in Remote Sensing (WHISPERS), 2011 3rd Workshop on*, pp. 1–4, IEEE, 2011.
- [27] A. Ambikapathi, T.-H. Chan, C.-Y. Chi, and K. Keizer, "Two effective and computationally efficient pure-pixel based algorithms for hyperspectral endmember extraction," in *Acoustics, Speech and Signal Processing (ICASSP), 2011 IEEE International Conference on*, pp. 1369–1372, IEEE, 2011.
- [28] W. Xiong, C.-I. Chang, C.-C. Wu, K. Kalpakakis, and H. M. Chen, "Fast algorithms to implement N-FINDR for hyperspectral endmember extraction," *Selected Topics in Applied Earth Observations and Remote Sensing, IEEE Journal of*, vol. 4, no. 3, pp. 545–564, 2011.
- [29] C.-I. Chang and D. C. Heinz, "Constrained subpixel target detection for remotely sensed imagery," *Geoscience and Remote Sensing, IEEE Transactions on*, vol. 38, no. 3, pp. 1144–1159, 2000.
- [30] J. Li, J. M. Bioucas-Dias, A. Plaza, and L. Liu, "Robust collaborative nonnegative matrix factorization for hyperspectral unmixing (R-CoNMF)," *arXiv preprint arXiv:1506.04870*, 2015.
- [31] A. A. Green, M. Berman, P. Switzer, and M. D. Craig, "A transformation for ordering multispectral data in terms of image quality with implications for noise removal," *Geoscience and Remote Sensing, IEEE Transactions on*, vol. 26, no. 1, pp. 65–74, 1988.
- [32] A. Hyvärinen and E. Oja, "Independent component analysis: algorithms and applications," *Neural networks*, vol. 13, no. 4, pp. 411–430, 2000.

- [33] X. Zhang, *Matrix analysis and applications*. Tsinghua and Springer Publishing house, 2013.
- [34] R. Heylen, M. Parente, and P. Gader, "A review of nonlinear hyperspectral unmixing methods," *Selected Topics in Applied Earth Observations and Remote Sensing, IEEE Journal of*, vol. 7, no. 6, pp. 1844–1868, 2014.
- [35] J. M. Bioucas-Dias and J. M. Nascimento, "Hyperspectral subspace identification," *Geoscience and Remote Sensing, IEEE Transactions on*, vol. 46, no. 8, pp. 2435–2445, 2008.
- [36] L. Miao and H. Qi, "Endmember extraction from highly mixed data using minimum volume constrained nonnegative matrix factorization," *Geoscience and Remote Sensing, IEEE Transactions on*, vol. 45, no. 3, pp. 765–777, 2007.
- [37] X. Lu, H. Wu, Y. Yuan, P.-g. Yan, and X. Li, "Manifold regularized sparse NMF for hyperspectral unmixing," *Geoscience and Remote Sensing, IEEE Transactions on*, vol. 51, no. 5, pp. 2815–2826, 2013.
- [38] W. Tang, Z. Shi, and Y. Wu, "Regularized simultaneous forward–backward greedy algorithm for sparse unmixing of hyperspectral data," *Geoscience and Remote Sensing, IEEE Transactions on*, vol. 52, no. 9, pp. 5271–5288, 2014.
- [39] W. Tang, Z. Shi, and Z. Duren, "Sparse hyperspectral unmixing using an approximate L_0 norm," *Optik-International Journal for Light and Electron Optics*, vol. 125, no. 1, pp. 31–38, 2014.