RESEARCH ARTICLE OPEN ACCESS

Multi clustering approach for fast hyperspectral endmember extraction using k-means

Shobana G , Muthuvel A2
1Shobana G. Author is currently pursuing M.E (Applied Electronics) in Mohamed Sathak Engineering College, Kilakarai. email:mailshoba@gmail.com,
2Mr.A. Muthuvel M.E Currently working as Asst.prof.of Dept of ECE Mohamed Sathak Engineering College.

Abstract:
Hyperspectral image processing represents a valuable tool for remote sensing of the Earth. Linear spectral unmixing is a very important tool for analyzing the content of remotely sensed hyper spectral images. This fact has led to the inclusion of hyper spectral sensors in different airborne and satellite missions for Earth observation. However, one of them a in drawbacks encountered when dealing with hyperspectral images is the huge amount of data to be processed, in particular, when advanced analysis techniques such as spectral unmixing are used. The proposed paper focus end member extraction on hyper spectral image. The proposed approach can be combined with existing algorithms for end member extraction, reducing the computational complexity of those algorithms while providing similar figures of accuracy. The main contribution of this project integration of spatial and spectral information and K-mean cluster means algorithm. Hyper spectral image has multi band data. Each band can be extracted by k-means and spatial spectral information can be finds the end region depend on image band. The spatial process deals pixel location based in mixing the spectral image. Compared to previous approaches based on similar spatial and spatial–spectral PP strategies, SE2PP clearly outperforms their results in terms of accuracy and computation speed, as it is demonstrated with artificial and real hyper spectral images.
Keywords— spatial–spectral PP (SSPP), vertex component analysis (VCA).

I. INTRODUCTION

Linear spectral unmixing is a very important tool for analyzing the content of remotely sensed hyperspectral images. It is based on the idea that each pixel vector in a hyperspectral image composed by spectral bands can be represented as a linear combination of spectrally pure constituent spectra or endmembers weighted by their corresponding abundance fractions ai that quantify. This work was supported in part by the Spanish Government algorithms can be classified into two groups. The first one is

A Endmember Extraction

The extraction process generally demands a formidable computational effort which becomes prohibitive for applications under real-time constraints. In order to improve the performance of the aforementioned endmember extraction step, different algorithms that not only take into account the spectral characteristics of the hyperspectral image to be unmixed but also benefit from the inherent spatial correlation between pixels in spatially adjacent regions have been recently uncovered in the scientific literature. These algorithms can be classified into two groups.

The first one is constituted by those approaches which incorporate the spatial information of the targeted image to the extraction process itself, such as the automatic morphological endmember extraction and the spatial–spectral endmember extraction approaches. The second group is composed by the algorithms in which the spatial information is exploited at a preprocessing (PP) stage that modifies the hyperspectral image prior to the extraction, such as the spatial PP (SPP), the region-based SPP (RBSPP), and the spatial–spectral PP (SSPP) algorithms.

As it can be noticed, the main difference between both groups is that, while the approaches categorized in the first one are endmember extraction algorithms themselves, the algorithms in the second group can be simply combined with existing extraction algorithms in order to compute the endmembers of an image.

This introduces several advantages, including the fact that there is no need to modify the extraction algorithm in order to include spatial information. More concretely, the RBSPP and the SSPP algorithms have demonstrated to offer not only proper levels of extraction accuracy but also a reduced computational cost when combined with classical endmember extraction algorithms. This is due to the fact that, prior to the extraction, both PP stages select a subset of pixels from the input image that are more likely to be the sought-after endmembers. However, these algorithms suffer, at least, from the following three drawbacks.

B Homogenously Mixed Areas

First, they are based on guiding the endmember extraction process to spatially large homogeneous areas expected to contain the purest signatures available in the scene. Although this strategy leads to obtaining a successful set of endmembers for many hyperspectral images, it also tends to obviate small targets or anomalous areas that may be present in the image. In addition, they may also guide the subsequent extraction algorithm to homogenously mixed areas, which do not contain pure pixels.

Second, when combined with computationally competitive endmember extraction approaches such as the vertex component analysis (VCA) algorithm, it results that the joint action of PP the image by means of the RBSPP or the SSPP algorithms and then extracting the endmembers from the resulting image takes more time than directly applying the VCA algorithm to the original non preprocessed image, which makes no sense from a computational point of view. This is mainly due to the computationally complex nature of the operations involved in the RBSPP and the SSPP algorithms, as well as to the fact that both PP stages tend to retain a significant amount of pixels from the original image.

Third, both algorithms show execution patterns that leave little room for introducing parallelism at the programming level, which is crucial for applications under real-time constraints.

Moreover, the complex operations required by the RBSPP and the SSPP algorithms also demand a large amount of hardware resources when mapped onto dedicated hardware computing platforms, as it is the case of an hypothetical scenario of a satellite equipped with a hyperspectral imaging sensor, where linear unmixing could take place on board in order to reduce the delays in the delivery of Earth observation payload data to ground processing facilities. In order to overcome the aforementioned limitations, this letter introduces a novel spatial–spectral PP module, which is based on selecting the pixels that are in the spatial edges (SEs) and in the spectral extremes (SEs) of the hyperspectral image under analysis. The proposed approach, called SE2PP, solves the identified drawbacks while maintaining the same degree of accuracy than RBSPP and SSPP when coupled with traditional endmember extraction algorithms.

With very high spectral resolution, hyper spectral sensors can now uncover many unknown signal sources which cannot be identified by visual inspection or a priori. In order to account for such unknown signal sources, a new definition, referred to as virtual dimensionality (VD) in this paper.

VD is defined as the minimum number of spectrally distinct signal sources that characterize the hyper spectral data from the perspective view of target detection and classification. It is different from the commonly used intrinsic dimensionality (ID) in the sense that the signal sources are determined by the proposed VD based only on their distinct spectral properties.

These signal sources may include unknown interfering sources, which cannot be identified by prior knowledge. With this new definition, three Neyman–Pearson detection theory-based thresholding methods are developed to determine the VD of hyper spectral imagery, where Eigen values are used to measure signal energies in a detection model. In order to evaluate the performance of the proposed methods, two information criteria, an information criterion (AIC) and minimum description length (MDL), and the factor analysis-based method proposed by Malinowski, are considered for comparative analysis. As demonstrated in computer simulations, all the methods and criteria studied in this paper may work effectively when noise is independent identically distributed. This is, unfortunately, not true when some of them are applied to real image data.

Harsanyi–Farrand–Chang (HFC), the noise-whitened HFC (NWHFC), and the noise subspace projection (NSP) methods) produce more reliable estimates of VDcompared to the AIC, MDL, and Malinowski's empirical indicator function, which generally overestimate VD significantly. In summary, three contributions are made in this paper, 1) an introduction of the new definition of VD, 2) three Neyman–Pearson detection theory-based thresholding methods, HFC, NWHFC, and NSP derived for VD estimation, and 3) experiments that show the AIC and MDL commonly used in passive array processing and the second-order statistic-based Malinowski's method are not effective measures in VD estimation.

II. RELATED WORK

Signal subspace identification is a crucial first step in many hyperspectral processing algorithms such as target detection, change detection, classification, and unmixing. The identification of this subspace enables a correct dimensionality reduction, yielding gains in algorithm performance and complexity and in data storage. This paper introduces a new minimum mean square error-based approach to infer the signal subspace in hyperspectral imagery. The method, which is termed hyperspectral signal identification by minimum error, is eigen decomposition based, unsupervised, and fully automatic (i.e., it does not depend on any tuning parameters). It first estimates the signal and noise correlation matrices and then selects the subset of eigenvalues that best represents the signal subspace in the least squared error sense.

A Spatial Domain Analysis

For many segmentation tasks, the image can contain objects with completely different shapes or an object that exhibits shape variability, such as the side view of a walking person. In such situations, the prior shape energy must make use of a set of prior templates or the multiple instances of a single object. The latter case is normally addressed by formulating the shape energy based on a statistical shape space.

B Remote Sensing

Signal subspace identification is a crucial first step in many hyperspectral processing algorithms such as target detection, change detection, classification, and unmixing. The identification of this subspace enables a correct dimensionality reduction, yielding gains in algorithm performance and complexity and in data storage. This paper introduces a new minimum mean square error-based approach to infer the signal subspace in hyperspectral imagery. The method, which is termed hyperspectral signal identification by minimum error, is eigen decomposition based, unsupervised, and fully automatic (i.e., it does not depend on any tuning parameters). It first estimates the signal and noise correlation matrices and then selects the subset of eigenvalues that best represents the signal subspace in the least squared error sense.

Hyperspectral sensors sample the reflected solar radiation from the Earth's surface in the portion of the spectrum extending from the visible region through the near infrared and mid infrared in hundreds of narrow contiguous bands. Such huge data volumes put stringent requirements on communications, storage, and processing.

C Higher spectral resolution

Imaging spectrometers measure electromagnetic energy scattered in their instantaneous field view in hundreds or thousands of spectral channels with higher spectral resolution than multispectral cameras. Imaging spectrometers are therefore often referred to as hyper spectral cameras (HSCs). Higher spectral resolution enables material identification via spectroscopic analysis, which facilitates countless applications that require identifying materials in scenarios unsuitable for classical spectroscopic analysis.

III.PROPOSED METHOD

This approaches for SPP prior to endmember extraction, which will be used for quantitative validation. The SPP estimates, for each input pixel vector, a scalar factor that is intimately related to the spatial similarity between the pixel and its spatial neighbors and then uses this scalar factor to spatially weigh the spectral information associated to the pixel.

The correction is performed so that pixels located in spatially homogeneous areas are expected to have a smaller displacement with regard to their original location in the data cloud than pure pixels surrounded by spectrally distinct substances. Resulting from the aforementioned operation, a modified simplex is formed, using not only spectral but also spatial information. The approaches are based on promoting the spatially homogeneous areas as the ones in which it is most likely to find pure pixels where, for the scope of this work, a spatially homogeneous area can be coarsely defined as a region of the image in which the dissimilarities between all its pixels are below a certain and typically small threshold.

One of the assumptions on which the SE2PP algorithm is based is that, theoretically, and if no more information is provided, all the pixels enclosed in a region identified as a spatially homogeneous area have the same probability of becoming an end member of the image. Moreover, in the surroundings of the frontier between two of these regions coexist pixels from both of them. Hence, if, rather than selecting a small amount of spatially large homogenous areas, a large amount of spatially small heterogeneous areas are

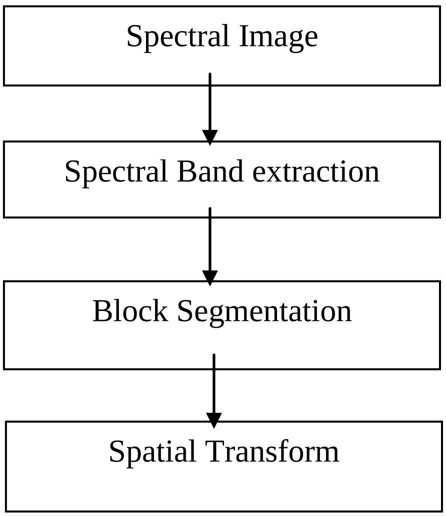


Fig.1 Proposed system

Where represents an image that has the same pixels as R and only one spectral value per pixel obtained as the average of all the spectral bands of R (i.e., if R has P pixels of N spectral bands each as indicated in, then will have P pixels of one spectral value per pixel, which is obtained for each pixel as R_{avg} = ((r1 + r2 + · · · + rN)/N)); (m1, n1) and (m2, n2) denote the absolute spatial coordinates of the region of the image being processed; and finally, μ represents the average value of all the pixels enclosed in the targeted region of the R_{avg} image. From the definition mentioned earlier, it can be easily concluded that spatially homogenous areas will have low SA values, while on the contrary; heterogeneous areas will produce large values of the SA parameter. The salient pixels and regions locally. A center-surround template is usually employed to evaluate the distinctiveness of a local area by measuring the local contrast.

Difference of Gaussian (DoG) and Gabor filters are commonly used to measure the local contrast. However, recently, researchers have also used Independent Component Analysis (ICA) bases as the filters. It has been shown that by training on tens of thousands of natural image patches, the resulting filters turn out to be quite similar to receptive fields found in the visual cortex. There are 192 color features from as the filters and obtain192 response maps.

There are two models to unmixed the hyperspectral sensor data, named linear and nonlinear. The linear mixture model identifies a collection of spectrally pure constituent spectra (endmembers) and expresses the measured spectrum of each mixed pixel as a linear combination of endmembers weighted by fractional abundances that indicate the proportion of each endmember present in the pixel. It assumes minimal secondary reflections and multiple scattering effects in the data collection procedure, and hence the measured spectra can be expressed as a linear combination of the spectral signatures of materials present in the mixed pixel. Endmember extraction of hyperspectral data aims at obtaining a good estimation of the mixing matrix. Several methods have been used to perform endmember extraction, including geometrical, statistical, and sparse regression-based approaches. A successful and widely used algorithm in the first category has been the N-FINDR. It relies on the assumption that, when the noise vector is negligible, all the spectrum vectors of hyperspectral pixels are contained in a convex set (named simplex) of high-dimensional space, and the endmembers are vertices of the simplex. Thus, the problem of endmember extraction is transformed to solving the vertices of the simplex.

A Fast Endmember Extraction Based on GPUs

It could be noted that there are three factors which lead to high computational overhead of the N-FINDR algorithm. Firstly, the algorithm is an iterative procedure, and every pixel in the data set must be evaluated to refine the estimate of endmembers, looking for the set of pixels that maximizes the volume of the simplex defined by the selected endmembers. Secondly, the computation is done for every single element in the input data set, and the replacement step is repeated for all the pixel vectors in the dataset. Thirdly, the calculation of the determinants is particularly time consuming.

B Results Obtained With Synthetic Hyperspectral Images

The synthetic hyperspectral images used in this work were generated with the Hyperspectral Imagery Synthesis toolbox for MATLAB available at [10], which allowed us to create a hyperspectral image of a spatial size defined by the user from p spectral signatures selected from the U.S. Geological Survey (USGS) digital spectral library. In particularly utilized images of 400 × 400 pixels and 431 spectral bands each, which were generated with the default options given by the tool for generating abundances according to a Legendre distribution or to two different types of Gaussian functions (Matérn and Spheric).

More concretely, generated three images for each type of function which correspond to three different values of p (7, 9, and 11). As it is observed in, where a snapshot of the images is depicted for the case of p = 7, the generated images have spatially homogenous areas as well as edges between them, as it is the case of the majority of real remotely sensed hyperspectral images.

C Results Obtained With the Cuprite Data Set

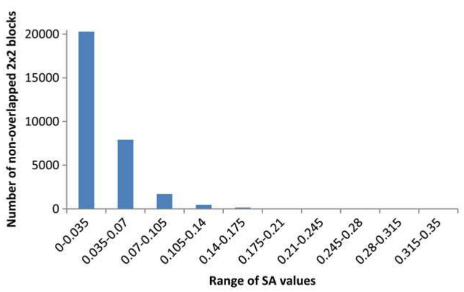
In order to test the proposed SE2PP algorithm in a more realistic scenario, the AVIRIS Cuprite image has also been used in this work. This scene is well understood mineralogical and has been widely used to validate the accuracy of endmember extraction algorithms. It consists of 350 × 350 pixels and 224 spectral bands between 0.4 and 2.5 μm. Prior to the analysis, different bands have been removed due to water absorption and low SNR resulting in a total of 188 spectral bands. In order to determine the number of endmembers of the image, the virtual dimensionality has been estimated by means of the noise-whitened Harsanyi–Farrand–Chang eigen thresholding method using the Neyman–Pearson test with the false alarm probability set to, resulting in a total number of 19 different pure materials.

This value is in agreement with the estimates provided by the well-known hyperspectral subspace identification method. It introduces the

SE2PP algorithm proposed in this letter, whose ultimate goal is to select a subset of pixels from a given hyperspectral image that will be the input to a subsequent endmember extraction algorithm.

IV. EXPERIMENTAL RESULTS

The proposed SE2PP algorithm is based on selecting a reduced set of pixels from the original image in order to process only those pixels in a posterior endmember extraction stage. In order to do so, it first computes the R_{avg} image from the R image and then divides R_{avg} into non-overlapped square blocks of pixels with the same spatial dimensions, which means that (m2 − m1) = (n2 − n1) = M. The SA of each of these blocks is further obtained, selecting only the pixels corresponding to blocks with a very high SA value, i.e., with a SA value above a predefined threshold (SAth), which has been defined in this work as follows:



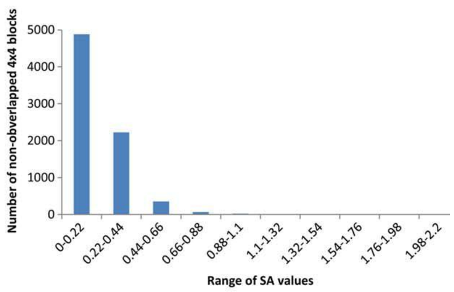


Fig.2 Graph value for the test images

V.CONCLUSION

In this letter, a novel spatial–spectral PP algorithm called SE2PP has been presented. The proposed method can be combined with existing endmember extraction algorithms (without modification of such algorithms) to produce solutions with reduced computational complexity while maintaining similar endmember identification accuracies. The results obtained with artificial and real hyperspectral images show that this PP method clearly outperforms other similar algorithms in terms of computational performance, obtaining similar average levels of extraction accuracy. Last but not least, the simplicity of the operations involved as well as their ease of parallelization validates the proposed SE2PP algorithm for current and future hardware-based real-time onboard systems.

REFRENCES

[1.] J. M. Bioucas-Dias, J.Chanussot, N. Dobigeon, M. Parente, Q. Du, P. Gader, and A.Plaza, “Hyperspectral unmixing overview: Geometrical, statistical and sparse regression-based approaches,” IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens., vol. 5, no. 2, pp. 354–379, Apr. 2012.

[2.] G.Martin and A. Plaza, “Spatial–spectral preprocessing prior to endmember identification and unmixing of remotely sensed hyperspectral data,” IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens., vol. 5, no. 2, pp. 380–395, Apr. 2012.

[3.] J. Alves, J. Bioucas-Dias, J. Nascimento, A.Plaza, and V. Silva “Parallel implementation of vertex component analysis for hyperspectral endmember extraction,” in Proc. IEEE IGARSS, 2012, pp. 4078–4081.

[4.] G.Martin and A. Plaza, “Region-based spatial preprocessing for endmember extraction and spectral unmixing,” IEEE Geosci. Remote Sens. Lett., vol. 8, no. 4, pp. 745–749, Jul. 2011.

[5.] A. Plaza and M. Zortea “Spatial preprocessing for endmember extraction,” IEEE Trans. Geosci. Remote Sens., vol. 47, no. 8, pp. 2679–2693, Aug. 2009.

[6.] Hyperspectral Imagery Synthesis Toolbox for MATLAB. [Online]. Available: http://www.ehu.es/ccwintco/index.php/Hyperspectral_Imagery_Synthesis_tools_for_MATLAB

[7.] J. Feng, J. Harris, D. M. Rogge, B. Rivard, A. Sanchez, and J. Zhang “Integration of spatial–spectral information for the improved extraction of endmembers,” Remote Sens. Environ., vol. 110, no. 3, pp. 287–303, Oct. 2007.

[8.] J. M. Bioucas-Dias and J. M. P. Nascimento “Vertex component analysis: A fast algorithm to unmix hyperspectral data,” IEEE Trans. Geosci. Remote Sens., vol. 43, no. 4, pp. 898–910, Apr. 2005.

[9.] C.-I. Chang and Q. Du, “Estimation of number of spectrally distinct signal sources in hyperspectral imagery,” IEEE Trans. Geosci. Remote Sens., vol. 42, no. 3, pp. 608–619, Mar. 2004.

[10.] P.Martinez, A.Plaza, J.Plaza, R.Perez, “Spatial/spectral end-member extraction by multidimensional morphological operations,” IEEE Trans. Geosci. Remote Sens., vol. 40, no. 9, pp. 2025–2041, Sep. 2002.