

Dimensionality reduction and hierarchical clustering in framework for hyperspectral image segmentation

K. Mallikharjuna Rao¹, B. Srinivasa Rao², B. Sai Chandana³, J. Harikiran⁴

^{1,2,4}School of Computer Science and Engineering, Vellore Institute of Technology (VIT-AP), Amaravathi, India

³Department of CSE, Kallam Haranadreddy Institute of Technology, Guntur, India

Article Info

Article history:

Received Dec 5, 2018

Revised Jan 7, 2019

Accepted Apr 5, 2019

Keywords:

Hierarchical clustering

Hyperspectral images

Image processing

Image segmentation

Remote sensing

ABSTRACT

The hyperspectral data contains hundreds of narrow bands representing the same scene on earth, with each pixel has a continuous reflectance spectrum. The first attempts to analyse hyperspectral images were based on techniques that were developed for multispectral images by randomly selecting few spectral channels, usually less than seven. This random selection of bands degrades the performance of segmentation algorithm on hyperspectral data in terms of accuracies. In this paper, a new framework is designed for the analysis of hyperspectral image by taking the information from all the data channels with dimensionality reduction method using subset selection and hierarchical clustering. A methodology based on subset construction is used for selecting k informative bands from d bands dataset. In this selection, similarity metrics such as Average Pixel Intensity [API], Histogram Similarity [HS], Mutual Information [MI] and Correlation Similarity [CS] are used to create k distinct subsets and from each subset, a single band is selected. The informative bands which are selected are merged into a single image using hierarchical fusion technique. After getting fused image, Hierarchical clustering algorithm is used for segmentation of image. The qualitative and quantitative analysis shows that CS similarity metric in dimensionality reduction algorithm gets high quality segmented image.

*Copyright © 2019 Institute of Advanced Engineering and Science.
All rights reserved.*

Corresponding Author:

K. Mallikharjuna Rao,
School of Computer Science and Engineering,
Vellore Institute of Technology (VIT-AP),
Amaravathi, India.
Email: rao.mkrao@gmail.com

1. INTRODUCTION

The process of acquiring information about an object on the earth using satellites without making any physical contact is called remote sensing [1]. The segmentation of objects on the earth by using electromagnetic radiations reflected or emitted by the surface is the main goal of remote sensing technology. New opportunities to use remote sensing data have arisen, with the increase of spatial and spectral resolution of recently launched satellites. Image segmentation is a key step in remote sensing applications [2]. In remote sensing, sensors are available that can generate hyperspectral data, involving many narrow bands in which each pixel has a continuous reflectance spectrum. Unsupervised image segmentation is an important research topic in hyperspectral imaging, with the aim to develop efficient algorithms that provide high segmentation accuracy [3].

This paper presents a framework for hyperspectral image segmentation using a clustering algorithm. The framework consists of three stages in segmenting a hyperspectral data set. In the first stage, dimensionality reduction algorithm is used to remove the bands that convey less information or redundant data. The data acquired from hyperspectral sensor usually contains hundreds of bands. This large data, fine details, redundancy and hyperspectrality differentiates the features of hyperspectral data and multispectral

data.

Due to this intrinsic dimensionality of data and spectral correlation between bands, dimensionality reduction is important for hyperspectral image segmentation. The dimensionality reduction step decreases many requirements for processing the hyperspectral data set such as storage space, computational load, communication bandwidth etc, thus increasing the efficiency of segmentation algorithm. In this paper a new method of dimensionality reduction using subset selection method is designed to select only informative bands removing redundancy. In the second stage, the informative bands which are selected in the first stage are merged into a single image using hierarchical fusion technique. The main goal of image fusion is to create a single image combining all the features in the selected image bands. After getting a single image, in the third stage, the image is segmented using hierarchical clustering algorithm. The flow diagram of proposed framework is shown in Figure 1. Section 2 presents dimensionality reduction using subset selection method, Section 3 presents hierarchical image fusion, Section 4 presents Hierarchical clustering algorithm, Section 5 presents Experimental results and Section 6 presents conclusions.

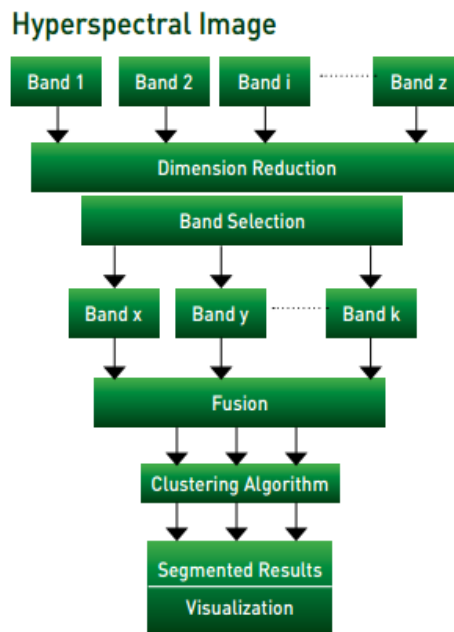


Figure 1. Framework for hyperspectral image segmentation

2. RESEARCH METHOD

Hyperspectral data set consists of stack of images which are strongly correlated means that there is huge amount of redundant information and that data need to be removed before segmentation. The dimensionality reduction step decreases many requirements for processing the hyperspectral data set, thus increasing the efficiency of segmentation algorithm [4]. The dimensionality reduction can be done using transformation based methods or selection based methods [5]. In transform based methods matrix transformations are used to project the data into lower dimension space which changes the physical meaning of spectral data. The selection based methods directly measures the information content in each individual band. The band selection is done by choosing the bands with higher information content. Selection based methods are better than transform based methods, as did not change the meaning of original dataset [6]. In transform based methods, as the original data is transformed, some useful information required for segmentation may be distorted, changing the physical meaning of data. This paper presents a new methodology of selection based dimensionality reduction of hyperspectral data.

Given hyperspectral data set with d bands. From these d bands, proposed methodology finds k bands that give us the most information for segmentation, discarding $(d-k)$ bands. These set of k bands contains the least number of dimensions that most contribute to accuracy of segmentation. Objective of this work is to create subsets of bands based on similarity criteria. Most similar bands are grouped into one subset. All bands in the same subset are similar with respect to metric M and the bands in different subsets are dissimilar with the same metric M . In this way k subsets are created, and we select one representative band from each subset.

These k bands from k subsets are used for further processing in the framework for hyperspectral image segmentation. The remaining (d-k) are discarded, thus achieving dimensionality reduction. In-order to create subsets, three parameters are required. A metric 'M', a reference band 'I_{ref}', and Threshold 'T'. Metric 'M': A metric M is a measure that directly shows how similar two images are. The metrics used in creation of subsets are Average Pixel Intensity [API], Histogram Similarity [HS], Mutual Information [MI] and Correlation Similarity [CS] [7].

The Average Pixel Intensity [API] of an image is calculated according to the given equation:

$$API_i = \frac{1}{M * N} \sum_{x=1}^M \sum_{y=1}^N I(x, y) \quad (1)$$

$$M = |I_{ref} - I_i|$$

Where M is the similarity metric and I_{ref} is the first image of a group.

The Mutual Information [MI] between two images A=I_{ref} and B=I_i can be given by equation:

$$MI = \sum_{a \in A} \sum_{b \in B} p(a, b) \log \frac{p(a, b)}{p(a) \cdot p(b)} \quad (2)$$

The Histogram Similarity [HS] between two images A= I_{ref} and B=I_i can be given by equation:

$$HS = \sum_{k=1}^K \sqrt{I_{ref}(k) \cdot I_i(k)} \quad (3)$$

Where K is the number of bins chosen for the histogram computation. If the value of HS is closer to zero, indicating the two histograms are highly disjointed.

The Correlation Similarity [CS] between two images is defined as

$$CS = \frac{\sum_x \sum_y (I_{ref}(x, y) - \bar{I}_{ref})(I(x, y) - \bar{I})}{\sqrt{\sum_x \sum_y (I_{ref}(x, y) - \bar{I}_{ref})^2 (I(x, y) - \bar{I})^2}} \quad (4)$$

Reference band 'I_{ref}': It always denote the first band of each subset. It is compared with the bands in the subset keeping similar images in the subset and discarding dissimilar images.

Threshold T: This is threshold that shows the similarity of images in a subset. Higher the threshold, smaller bands in each subset, increasing the total number of subsets.

Start with subset_i, (i=1) containing the original data set I_j (j=1, ..., d). Assume I_{ref} = I₁ first band in the data set. Add each band into subset_i, if the similarity between reference and selected band is less than threshold T. Otherwise move to subset_i+1. Start with subset_i = Φ (empty set) and bands are added to the subset_i if M(I_{ref}, I_j) ≤ T. The subset_i+1 is created according to the following formula

$$\begin{aligned} &\leq T, \text{ assign } I_j \text{ to subset}_i \\ &M(I_{ref}, I_j) = \\ &> T, \text{ assign } I_j \text{ to subset}_i+1, \text{ and first band in subset}_i+1 = I_{ref} \text{ for subset}_i+1. \end{aligned} \quad (5)$$

Where M is the similarity metric between two bands and j ranges from 2 to d. The same process is repeated until k subsets are created. From these k subsets, k bands are selected one from each subset.

3. HIERARCHICAL IMAGE FUSION TECHNIQUE

In hierarchical image fusion technique [8], the entire data set is partitioned into P subsets of hyperspectral, where P is given by $P = \frac{K}{M}$, K number of bands in data set and M bands in each subset. First image fusion is carried out independently on these P subsets, to form P fused images. These P images are used as input for second stage fusion again by dividing into subsets. This procedure is repeated in a hierarchical manner to generate the final result of fusion in a few stages. The flow diagram of hierarchical image fusion is shown in Figure 2. The fused image F at any stage is a linear combination of input images as shown below:

$$F(x, y) = \sum_{k=1}^M w_k(x, y) I_k(x, y)$$

and

$$\sum_{k=1}^M w_k(x, y) = 1, \forall(x, y)$$

(6)

where $w_k(x, y)$ is the normalized weight for the pixel at location (x, y) , $F(x, y)$ is the fused image [9, 10].

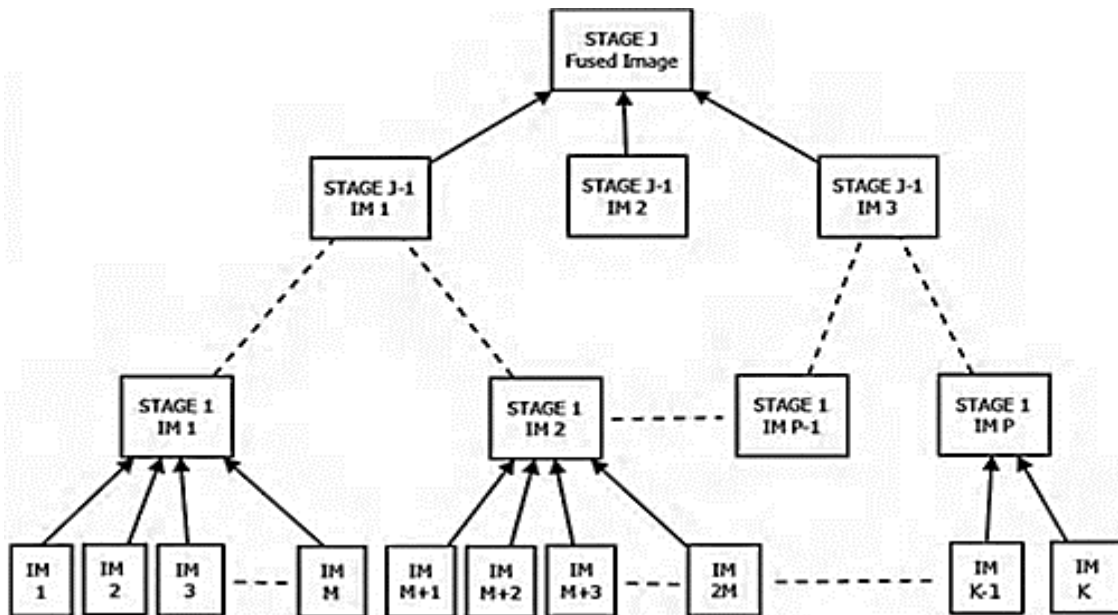


Figure 2. Hierarchical image fusion

4. HIERARCHICAL CLUSTERING TECHNIQUE

Hierarchical clustering technique segments the data in the form of multilevel hierarchy, where the clusters at one level are combined with the clusters at the next level forming hierarchical cluster tree or dendrogram. The scaling of the level of hierarchy depends on the application on which hierarchical clustering algorithm is implemented and the dendrogram shows the cluster tree [11]. The basic steps in the hierarchical clustering algorithm on hyperspectral image obtained after fusing with implementation in Matlab is shown:

- Compute the similarity or dissimilarity between pixels in the image.
- This can be done using 'pdist' function in Matlab with different distance metrics such as 'euclidean', 'seuclidean', 'cityblock', 'minkowski', 'cosine', etc.,
- After finding the similarity distances group the pixels, into binary hierarchical cluster tree.
- This can be done using 'linkage' function in Matlab with a method such as 'average', 'centroid', 'complete', 'median', 'single' and 'weighted' for computing distance between clusters.
- The linkage function uses the distance information generated in (i) to determine the proximity of pixels to each other. As pixels are paired into binary clusters, the newly formed clusters are grouped into larger clusters until a hierarchical tree is formed.
- After construction of cluster tree, next step is to determine where to cut the hierarchical tree into clusters. This can be done by using 'cluster' function in Matlab. The cluster function can create these clusters by detecting natural groupings in the hierarchical tree or by cutting off the hierarchical tree at an arbitrary point. This function has a parameter 'maxclust' to create specific number of clusters.
- The hierarchical cluster tree generated by above three functions can be visualized graphically by dendrogram plots which can be done by using the function 'dendrogram'.

5. EXPERIMENTAL RESULTS

The proposed methodology is tested on Indian Pines hyperspectral image data set collected from [12] containing 220 spectral bands. The dimensionality reduction is done using proposed subset methodology with different similarity metrics and 80 bands selected from 220 bands. After dimensionality reduction, hierarchical image fusion is carried out to create a single image. This image is segmented using Hierarchical Clustering Algorithm with maximum 16 clusters. In the tree generated by hierarchical clustering technique, any two pixels in the original image are linked together at some level and the height of the link denotes the distance between two clusters that contain those pixels. This height is known as cophenetic distance. This distance will measure the correlation between pdist and linkage functions, known as cophenetic correlation coefficient. The closer the value of the cophenetic correlation coefficient is to 1, the more accurately the clustering solution reflects your data [13]. Table 1 shows the cophenetic correlation coefficient measures with different metrics in pdist and linkage functions used in hierarchical clustering technique on hyperspectral image. Figure 3 shows the code snippet of hierarchical clustering. Table 2 shows the classification accuracies obtained by hierarchical clustering with different similarity metrics used in dimensionality reduction compared with ground truth information available in [14, 15]. The qualitative analysis of the proposed method on Indian Pines hyperspectral data set is shown in Figure 4.

Table 1. Cophenetic correlation measures

Pdist	Linkage	Cophenetic correlation measure
Euclidean	Single	0.8742
Cityblock	Average	0.9197
Squaredeuclidean	Centroid	0.9015
Cosine	Median	0.8698

Table 2. Classification accuracies of different similarity metrics used in dimensionality reduction step

Similarity metric in dimensionality reduction algorithm	Classification accuracy
API	89.6%
MI	90.1%
HS	92.3%
CS	94.9%

```

Y=pdist(double(i4),'squaredeuclidean');
squareform(Y);
Z=linkage(Y,'centroid');
dendrogram(Z);
c = cophenet( Z, Y);
T = cluster( Z,'maxclust', 2);

```

Figure 3. Code Snippet of Hierarchical clustering

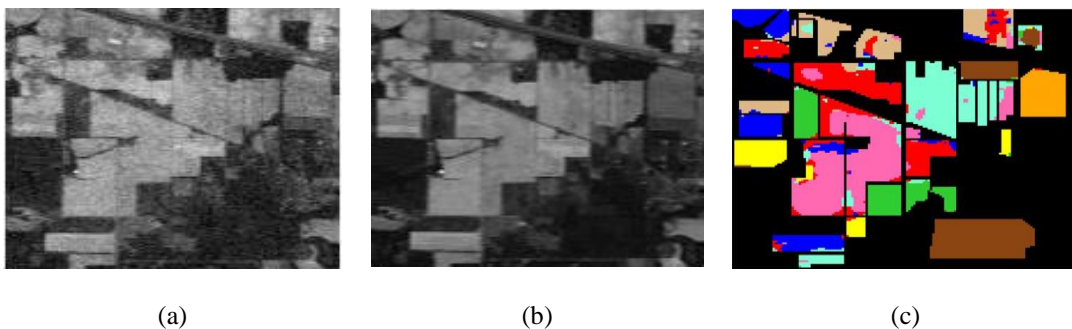


Figure 4. Segmentation of hyperspectral image using FCM: (a) Original image band 100 (Indian pines dataset), (b) Fused Image, (c) Segmented Image

6. CONCLUSIONS

In this paper a framework for hyperspectral image segmentation is presented. The framework is carried out in three stages. First stage contains dimensionality reduction method using subset selection method to select informative bands leaving the bands that convey less descriptive information, second stage contains hierarchical image fusion to generate a single informative band and in the third stage, segmentation using Hierarchical clustering algorithm. Existing methods for hyperspectral data sets is done by selecting limited number of bands normally less than seven. The accuracy of any segmentation algorithm decreases if the number of spectral bands increases. The framework presented in this paper provides a methodology for segmenting the hyperspectral data set by incorporating all the information existing in the original bands rather than selecting some spectral bands. The methodology presented in this paper shows the performance of hierarchical clustering algorithm by using different similarity metrics in dimensionality reduction algorithm.

REFERENCES

- [1] Y. Ma, J. Zhang and Y. Gao, "High Resolution Remote Sensing Image Classification of Coastal Zone and Its Automatic Realization," *2008 International Conference on Computer Science and Software Engineering*, Hubei, 2008, pp. 827-829.
- [2] B.Saichandana et.al. "Hyperspectral Image Enhancement Using Evolutionary Algorithm" *International Journal of Advanced Research*, 3(4): 934-938. March 2016.
- [3] B.RaviTeja, J.Harikiran et.al. "A Framework for Hyperspectral Image Segmentation using Unsupervised Algorithm", *Indian Journal of Science and Technology*, 9(47): 1-6. December 2016.
- [4] B.Raviteja, Dr.M.S.Prasad Babu, Dr.K.Venkata rao, J.Harikiran "A New Methodology of Hierarchical Image Fusion in Framework for Hyperspectral Image Segmentation", *Indonesian Journal of Electrical Engineering and Computer Science (IJECS)*, Volume 6, Number 1, pp. 58-65, April 2017.
- [5] A. Martínez-UsóMartinez-Uso, F. Pla, J. M. Sotoca and P. García-Sevilla, "Clustering-Based Hyperspectral Band Selection Using Information Measures," in *IEEE Transactions on Geoscience and Remote Sensing*, vol. 45, no. 12, pp. 4158-4171, Dec. 2007.
- [6] Ozan Arslan et.al. "A Comparative analysis of Classification methods for hyperspectral images generated with conventional dimension reduction methods", *Turkish Journal of Electrical Engineering and Computer Sciences*, 25:58-72. 2017.
- [7] P. Shah, M. Drumetz, S. N. Merchant and U. B. Desai, "Efficient hierarchical fusion using adaptive grouping techniques for visualization of hyperspectral images," *2011 8th International Conference on Information, Communications & Signal Processing*, Singapore, 2011, pp. 1-5.
- [8] Dr.R.Kiran Kumar, B.Saichandana, Dr.K.Srinivas, "Dimensionality Reduction and Classification of Hyperspectral Images using Genetic Algorithm", *Indonesian Journal of Electrical Engineering and Computer Science (IJECS)*; 3(3): 503-511. 2016.
- [9] B.Saichandana et.al. "Image Fusion in Hyperspectral Image Classification using Genetic Algorithm", *Indonesian Journal of Electrical Engineering and Computer Science (IJECS)*; 2(3): 703-711. 2016.
- [10] K. Kotwal and S. Chaudhuri, "Visualization of Hyperspectral Images Using Bilateral Filtering," in *IEEE Transactions on Geoscience and Remote Sensing*, vol. 48, no. 5, pp. 2308-2316, May 2010.
- [11] Meng-Xin Li, Cheng-Dong Wu, Zhong-Hua Han and Yong Yue, "A hierarchical clustering method for attribute discretization in rough set theory," *Proceedings of 2004 International Conference on Machine Learning and Cybernetics (IEEE Cat. No.04EX826)*, Shanghai, China, 2004, pp. 3650-3654 vol.6.
- [12] http://www.ehu.eus/ccwintco/index.php?title=Hyperspectral_Remote_Sensing_Scenes.
- [13] B.Saichandana, et.al, "Clustering Algorithm combined with Hill climbing for Classification of Remote Sensing Image", *International Journal of Electrical and Computer Engineering (IECE)*; 4(6): 923-930. 2014.
- [14] B.Saichandana, J.Harikiran, Dr.K.Srinivas, Dr.R.KiranKumar, "Application of BEMD and Hierarchical Image Fusion in Hyperspectral Image Classification", *International Journal of Computer Science and Information Security*; 14(5): 437-445. 2016.
- [15] J.Harikiran, et.al. "Multiple Feature Fuzzy C-means Clustering Algorithm for Segmentation of Microarray image", *International Journal of Electrical and Computer Engineering (IJECE)*; 5(5): 1045-1053. 2015.

BIOGRAPHIES OF AUTHORS



K.Mallikharjuna Rao ,currently working as Assistant Professor, Department of CSE, VIT AP . He has published 10 research papers in international conferences and journals. He can be reached at rao.mkrao@gmail.com



B. Srinivasarao, currently working as Associate Professor, Department of CSE, VIT AP. He has published 10 research papers in international conferences and journals. He can be reached at sreenivas.battula@gmail.com



B. Sai Chandana, currently working as Professor, Department of Computer Science and Engineering, Kallam Haranadhareddy Institute of Technology, Guntur. She has published 20 research papers in international conferences and journals. She can be reached at bschandana@gmail.com



J. Hari Kiran, currently working as Assistant Professor, School of CSE, VIT-AP. He has published 25 research papers in international conferences and journals. He can be reached at jhari.kiran@gmail.com