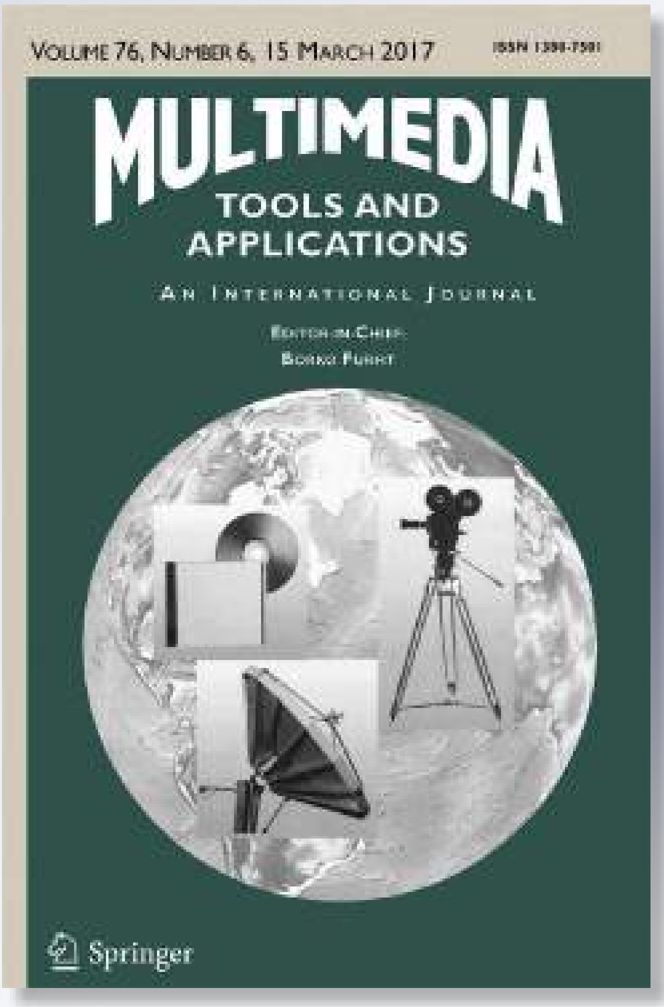
Clustering based band selection for endmember extraction using simplex growing algorithm in hyperspectral images

Veera Senthil Kumar G & Vasuki S

Multimedia Tools and Applications
An International Journal

ISSN 1380-7501
Volume 76
Number 6

Multimed Tools Appl (2017)
76:8355-8371
DOI 10.1007/s11042-016-3420-4



Multimed Tools Appl (2017) 76:8355–8371
DOI 10.1007/s11042-016-3420-4

Clustering based band selection for endmember extraction using simplex growing algorithm in hyperspectral images

Veera Senthil Kumar G [1] **· Vasuki S** [1]

Received: 22 August 2015 / Revised: 10 January 2016 / Accepted: 29 February 2016 /
Published online: 2 April 2016

Abstract With the advancement in technology, hyperspectral images have potential applications in the field of remote sensing due to their high spectral resolution. Despite the hyperspectral image providing abundant information, its analysis suffers from the problem of high dimensionality. Hence, Dimensionality Reduction (DR) is an essential task in all hyperspectral image analysis. Band Selection, which is one of the DR techniques, is still a challenging issue even though many algorithms have been developed. To provide remedy for this issue, this paper explores a novel approach for band selection using K-means clustering on statistical feature in hyperspectral images. The proposed method of clustering based band selection for DR is simple and accurate. A reliable estimate of number of bands to be selected is provided by Virtual Dimensionality (VD). Informative bands preserving maximum information are selected based on the statistical feature, the variance using K-means Clustering technique. Further, our proposed work involves the utilization of the effectiveness of Simplex Growing Algorithm (SGA) on endmember extraction in association with clustering based band selection. Using Fully Constrained Least Squares (FCLS) method, abundance fraction is estimated based on endmember signatures, which are derived using Endmember Extraction Algorithm (EEA). The proposed work is investigated and compared with that of N-FINDR and Vertex Component Analysis (VCA) algorithms. The performance of the proposed algorithm is evaluated using Root Mean Square Error (RMSE), Spectral Angle Distance (SAD) and computation time. Experimental results show that the proposed clustering based band selection with SGA endmember extraction algorithm reduces the average SAD by 8 to 10 % and the average RMSE by nearly 1 %, compared to that of N-FINDR and VCA algorithms. In terms of computation time, the proposed band selection based DR with SGA algorithm is seven times faster than conventional transform based DR with SGA algorithm.

✉ Veera Senthil Kumar G
gvs@vcet.ac.in

Vasuki S
sv@vcet.ac.in

[1] ECE Department, Velammal College of Engineering and Technology, Madurai, Tamil Nadu, India

Keywords Dimensionality reduction · Virtual dimensionality · Band selection · K-means · Endmember extraction · SGA · FCLS

1 Introduction

Hyperspectral image is an image cube which is composed of hundreds of spectral bands with fine resolution. It contains abundant information which is essential for accurate object identification. However, there is a curse of dimensionality in hyperspectral image analysis. Based on the literature, DR techniques are broadly categorized into two types namely transform based DR and band selection DR. In the former approach, the image is transformed to a lower dimensional space [13] in which the originality of the data is perturbed. Also, the transform based DR for extracting endmembers from the hyperspectral image has been used in [6], which increases the computational complexity. But, in latter approach, a subset of bands is selected without changing their originality [9]. The accuracy of object detection is a significant factor in the analysis of hyperspectral images. Hence, the subset of bands should be selected preserving high accuracy while reducing the dimension.

Many band selection methods have been found in the literature. Band Selection methods are broadly categorized into two types namely Supervised and Unsupervised. Supervised band selection requires prior knowledge on data set, which may not be available practically. But unsupervised band selection selects informative bands from the larger band set, without prior information. Irrespective of the methods, there exist two issues in band selection. The first issue is estimating the number of bands, which is resolved by the concept of VD [5] using Noise-Whitened Harsanyi-Farrand-Chang (NWHFC) method. The second issue is identifying criteria to select the bands. Many unsupervised band selection techniques exist in the literature are based on information evaluation [4, 5, 11, 12, 19, 23], in which the selected bands are generally extreme ones and sometimes outliers. Band selection is interpreted as a clustering problem in order to select representative bands instead of extreme bands. Under these circumstances, clustering based methods [1, 18, 21] provide a better strategy for band selection.

Further, to prove the effectiveness of proposed band selection method, several EEAs have been considered. Pixel Purity Index (PPI) is one of the popular EEAs, developed by Boardman, used in hyperspectral images. But it requires the prior knowledge of number of skewers. N-FINDR algorithm [24] is another widely used EEA using the concept of maximum simplex volume, but it also suffers from the issues that require prior knowledge of number of endmembers and initialization of endmembers. VCA [20] is recently developed EEA, successful in unmixing hyperspectral data. Simplex Growing Algorithm (SGA) [6] is a simplex based approach like N-FINDR, but overcomes the drawback of inconsistency in endmember determination present in both VCA and N-FINDR. Real-Time SGA (RT-SGA), a version of SGA without DR has been proposed in [7], which is computationally more expensive than SGA with DR. Field-Programmable Gate Array (FPGA) implementation for SGA algorithm has been proposed in [8], in which the drawbacks of N-FINDR algorithm have been resolved.

FCLS algorithm [16] is used to estimate the abundance fractions of endmembers. The FCLS method is similar to the one found in [15] which includes the Abundance Sum-to-one Constraint (ASC) in nonnegative least squares algorithm [17].

The rest of the paper is organized as follows: Estimation of number of endmembers using VD is explained in section 2. Statistical feature based K-means clustering algorithm for band selection is described in section 3. Endmember extraction algorithms are presented in section 4.

The proposed work is explained in section 5. Experimental results are discussed in section 6 and conclusion and future work is given in section 7.

2 Virtual dimensionality

VD estimation is a reliable measure for finding number of distinct spectral signatures available in hyperspectral image [2]. Eigen threshold based method, referred to as the Harsanyi-Farrand-Chang (HFC) method, has been previously developed in [14] to find the number of endmembers in AVIRIS data. It is based on Neyman-Pearson detection theory and the number of bands are found by testing the number of failures for all bands, for a given false alarm rate P_F. Noise-Whitened HFC (NWHFC) method [3] is an alternative for estimating VD accurately because HFC does not have noise-whitening process.

3 Clustering based band selection

Clustering based band selection is proposed to identify the informative bands by grouping bands based on minimum variance within the cluster and maximum variance between the clusters [18]. K-means clustering is one of the popular unsupervised algorithms for solving the problem of clustering. The algorithm starts with K number of centroids for each cluster. The data points in the data set are associated to nearest centroid resulting K groups. Then new centroids are calculated for each cluster and again the association process is done. This process is repeated until there is no change in the centroids between previous and current step in the iteration. The centroids of the clusters in each iteration are calculated by minimizing the sum of squared errors. The steps involved in K-means algorithm are given in Algorithm I.

Algorithm I: K-Means Clustering

```
1. Initialize the clusters and cluster centroids.
2. Associate each object to the nearest centroid based on minimum
distance.
3. Recalculate the centroids and form new clusters.
4. Repeat the steps (2) and (3) until the centroids are fixed.
```

Variance for each band is calculated and the bands are clustered using these variance values. The bands possessing maximum variance from each cluster are selected. For L-dimensional hyperspectral image of size M x N, the variance for K^{th} band, d_k is calculated from the data [22] as follows:

$$d_k = \frac{1}{MN} \sum_{i=1}^{MN} \left(b_i - \overline{B}_k \right)^2 \tag{1}$$

where $\overline{B}_k$ is the mean of the band.

4 End member extraction algorithms

An endmember means signature of a pure pixel [6] which refers to the spectrum of a single material. There are two ways by which endmembers can be identified: Endmember Extraction

Algorithm (EEA), in which pure pixels are determined directly from the data and Endmember Generation Algorithm(EGA), in which pure signatures are generated from available pure pixels. EEA algorithms have been discussed in this section.

A. N-FINDR Algorithm

N-FINDR algorithm [24] which deploys the concept of finding the simplex of the maximum volume with a given number of vertices, but it has been briefly described. The detailed step-by-step algorithm is explained in [6], which is given in Algorithm II.

Algorithm II: N-FINDR Algorithm

```
1. VD is estimated to find the number of endmembers in the image.
2. Using this VD estimate, informative bands are selected
through K-means clustering technique.
3. Initialize the endmembers by selecting a set of p pixels
randomly from the data.
4. Calculate the volume of the simplex from the initial endmembers.
5. Check for resulting larger simplex volume by calculating
volumes of the simplexes formed by substituting each initial
endmember with pixel under test.
6. If the above condition is met, the pixel being tested replaces
the initial endmember and the process is repeated until potential
endmember is found.
7. The pixels which are found as potential endmembers are the
final endmembers.
```

B. Vertex component analysis Algorithm

Another popular algorithm in the sequence of endmember extraction in hyperspectral data is VCA algorithm [20]. The algorithm iteratively projects data onto a direction orthogonal to the subspace spanned by the endmembers already determined. The new endmember signature corresponds to the extreme of the projection. The algorithm iterates until all endmembers are exhausted. The steps involved in VCA algorithm are given in Algorithm III.

Algorithm III: VCA Algorithm

```
1. SNR threshold is calculated.
2. SNR and threshold SNR is then compared.
3. If SNR > SNR threshold, the data is projected on to a lower
dimensional subspace using SVD.
4. Else the data is projected on to the lower dimensional sub-
space using PCA.
5. Auxiliary matrix is initialized to store the endmember sig-
natures' projections.
6. Then, the projection extremes are found iteratively by
projecting data on to a vector orthonormal to the subspace spanned
by the columns of Auxiliary matrix. The stopping criteria is the
number of endmembers required.
7. Finally, the condition SNR > SNR threshold condition is again
checked to find the endmember matrix.
```

C. **Simplex growing Algorithm**

It is a simplex based EEA which finds simplex with maximum volume for each vertex [6]. The process of finding maximum volume simplexes is repeated till the number of vertices equals the number of endmembers, which are provided by VD estimate. The algorithm requires random selection of initial endmember pixel, which is determined by the randomly generated target pixel. The randomness of target pixel has no effect on final endmembers. Moreover, the generated first endmember remains as one of the final endmembers. Hence, SGA is consistent in producing desired endmembers. The detailed algorithm has been explained in Algorithm IV.

Algorithm IV: Simplex Growing Algorithm

Initialization of first endmember:

1. First a target pixel t is randomly generated.

2. Then a pixel e_1 yielding maximum of absolute determinant of the matrix over all sample vectors r is found using the formula

$$e_1 = \arg\left\{\max_r \left[\left|\det\begin{bmatrix} 1 & 1 \\ t & r \end{bmatrix}\right|\right]\right\} \tag{2}$$

where a DR technique is used to reduce the original dimensionality of data from L to dimension 2 and t is the target pixel.

Algorithm:

1. Calculate VD to estimate the number of endmembers p.

2. Set e_1 as initial endmember pixel and set n = 1.

3. For $n \geq 1$, calculate the volume of the simplex $V(e_1, e_2, \ldots, e_n, r)$ specified by the vertices e1,e2,…en, r, for each sample vector r using

$$V(e_1, e_2, \ldots, e_n, r) = \frac{\left|\det\begin{bmatrix} 1 & 1\ldots\ldots\ldots1 & 1 \\ e_1 & e_2\ldots\ldots e_n & r \end{bmatrix}\right|}{n!} \tag{3}$$

where a DR technique is used to reduce the original dimensionality of data from L to dimension n.

4. Find e_{n+1} that yields maximum of (3), ie,

$$e_{n+1} = \arg\left\{\max_r \left[V(e_1, e_2, \ldots e_n, r)\right]\right\} \tag{4}$$

5. If n < p, then increment n value by one and proceed from step 3. Else, the obtained set of $\{e_1, e_2, \ldots, e_p\}$ in step 4 is final endmembers.

5 Proposed methodology

In the proposed work, as a preprocess, water absorption and low SNR bands are removed from hyperspectral image. After preprocessing, VD is calculated using NWHFC method to determine the number of endmembers. The VD estimate can also be used to decide the number of bands to be selected for the subsequent process of band selection. Since

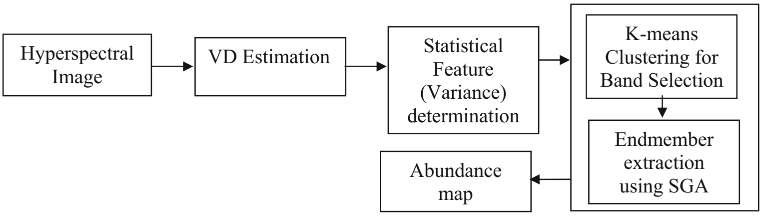


Fig. 1 Block diagram of proposed clustering based band selection with SGA

the hyperspectral image cube is composed of hundreds of bands, dimensionality reduction should be performed prior to all the hyperspectral image analysis. Dimensionality reduction is accomplished in this work using band selection method based on VD estimate. In the band selection process, bands are clustered using K-means algorithm based on the variance. K-means algorithm is a familiar cluster algorithm and is adopted especially because it clusters bands into distinct groups exactly. The variance is calculated using (1) for each band and the bands are clustered based on these variance values. The bands possessing minimum variance are selected from each cluster and used as representative bands for the subsequent process of endmember extraction. In endmember extraction phase, SGA algorithm is adopted since it provides superior and consistent result in the determination of endmembers, compared to the other algorithms such as N-FINDR and VCA. Moreover, SGA algorithm resolves the two issues, which are found in N-FINDR and VCA algorithm, the first one is random initialization of endmember pixels and the second one is inconsistency in producing the final endmembers. Hence, clustering based band selection with SGA is proposed for effective determination of endmembers from the hyperspectral image. The block diagram of the proposed clustering based band selection with SGA is given in Fig. 1.

Normally, in a hyperspectral image, a pixel is a mixture of number of materials. There are two ways by which a pixel can be modeled: Linear mixing and Non-linear mixing. Linear Spectral Mixture Analysis (LSMA) is deployed in this paper for calculating abundance fractions of materials present in the image. In LSMA, FCLS method has been used for estimating the abundance fractions of materials accurately, because this method satisfies the basic two requirements of LSMA ie, non-negativity constraint and sum-to-one constraint.

6 Experimental results

Two real hyperspectral datasets are used in our experimentation. First one is a 224-band Jasper Ridge image as shown in Fig. 2 having size 512 x 614 pixels ranging from 0.38 to 2.5 μm, a popular hyperspectral data used in [10]. To reduce the computational complexity, a subimage of size 100 x 100 pixels starting from the (105,269)-th pixel in the original image used in [26] has been deployed in this experiment. There are four endmembers present in the image: Water, Tree, Road and Soil. The ground truth of endmembers spectral signatures is shown in Fig. 3.

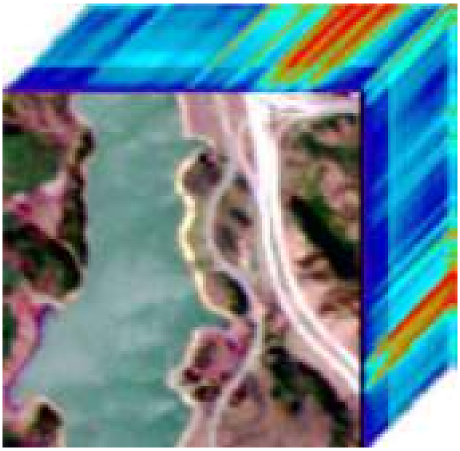


Fig. 2 Jasper Ridge

After removing 1–3, 108–112, 154–166 and 220–224 low SNR and water absorption bands, 198 bands are retained for the subsequent process.

The number of bands required for this dataset is estimated using the concept of VD by NWHFC method with false-alarm probability set to $P_F = 10^{-5}$, beyond which the value of VD remains constant at 9. The VD estimation for different false–alarm rate (P_F) is shown in Table 1.

It is inferred from Table 1 that the VD estimate is 9. Variance is calculated for 198 bands. Then the bands are clustered using K-means algorithm based on the variance. The number of clusters is equal to the number of endmembers. Bands having maximum variance value in each cluster are selected and considered to be the informative bands for subsequent process. It is observed in the experimentation that the endmembers

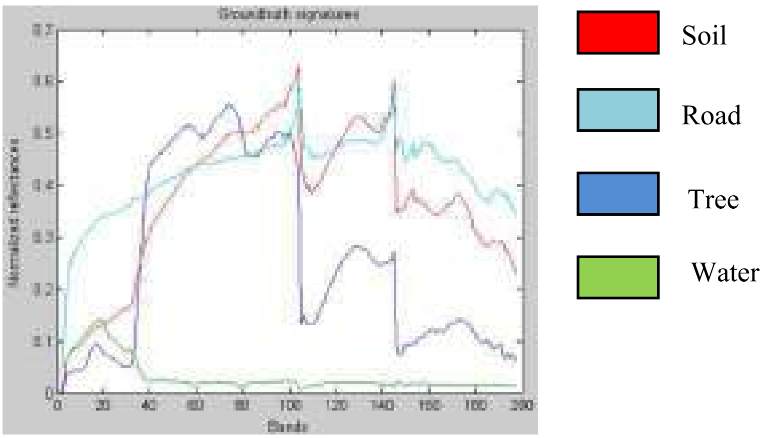


Fig. 3 Groundtruth spectral signatures of endmembers

Table 1 VD estimation using NWHFC method for different false-alarm rate

P_F	10^{-1}	10^{-2}	10^{-3}	10^{-4}	10^{-5}
VD	21	17	12	10	9

present in the Jasper Ridge image can be determined accurately with four bands instead of nine. The selected bands based on variance are shown in Table 2.

In endmember extraction phase, three endmember extraction methods such as N-FINDR, VCA and SGA are considered and their performance on selected bands is evaluated. The spectral signatures of endmembers in Jasper Ridge are found using the three algorithms and they are plotted in Fig. 4a–c. The abundance fractions of endmembers are estimated using FCLS method and their corresponding abundance maps are shown in Fig. 5a–d along with Groundtruth Abundance map for comparision.

Two benchmark metrics namely Spectral Angle Distance (SAD) and Root Mean Square Error (RMSE) [25] have been used to assess the results quantitatively. SAD measures the spectral similarity between the estimated endmember using EEA and its groundtruth whereas RMSE evaluates the estimated abundance map. Lower the values of these measures indicate better performance of the algorithm. They are defined as

$$SAD\left(m_k, \hat{m}_k\right) = \arccos\left(\frac{{m_k}^T \hat{m}_k}{\|m_k\| . \left\|\hat{m}_k\right\|}\right) \tag{5}$$

where $\hat{m}_k$ is the kth estimated endmember and m_k is the corresponding groundtruth.

$$RMSE\left(z, \hat{z}\right) = \left(\frac{1}{N}\left\|z-\hat{z}\right\|^2\right)^{1/2} \tag{6}$$

where N is the number of pixels in the image, $\hat{z}$ (a row vector in the abundance matrix) is the estimated abundance map and z is the corresponding ground truth.

The performance comparison of proposed clustering based band selection with SGA and that of N-FINDR and VCA algorithm for Jasper Ridge is given in Tables 3 and 4 using (5) and (6).

It has been observed from Tables 3 and 4 that the obtained SAD and RMSE values for each endmembers are different using different endmember extraction algorithm. Compared to clustering based band selection with N-FINDR and VCA, the average error percentage obtained for the proposed algorithm is lower, which is shown bold in Tables 3 and 4. From Table 3, it is clearly shown that there is a considerable improvement in endmember determination as a result of the proposed algorithm for this Jasper ridge image. Eventhough the average RMSE values obtained for each EEA

Table 2 Selected bands for Jasper Ridge

Feature	Bands selected
Variance	182,118, 53,104

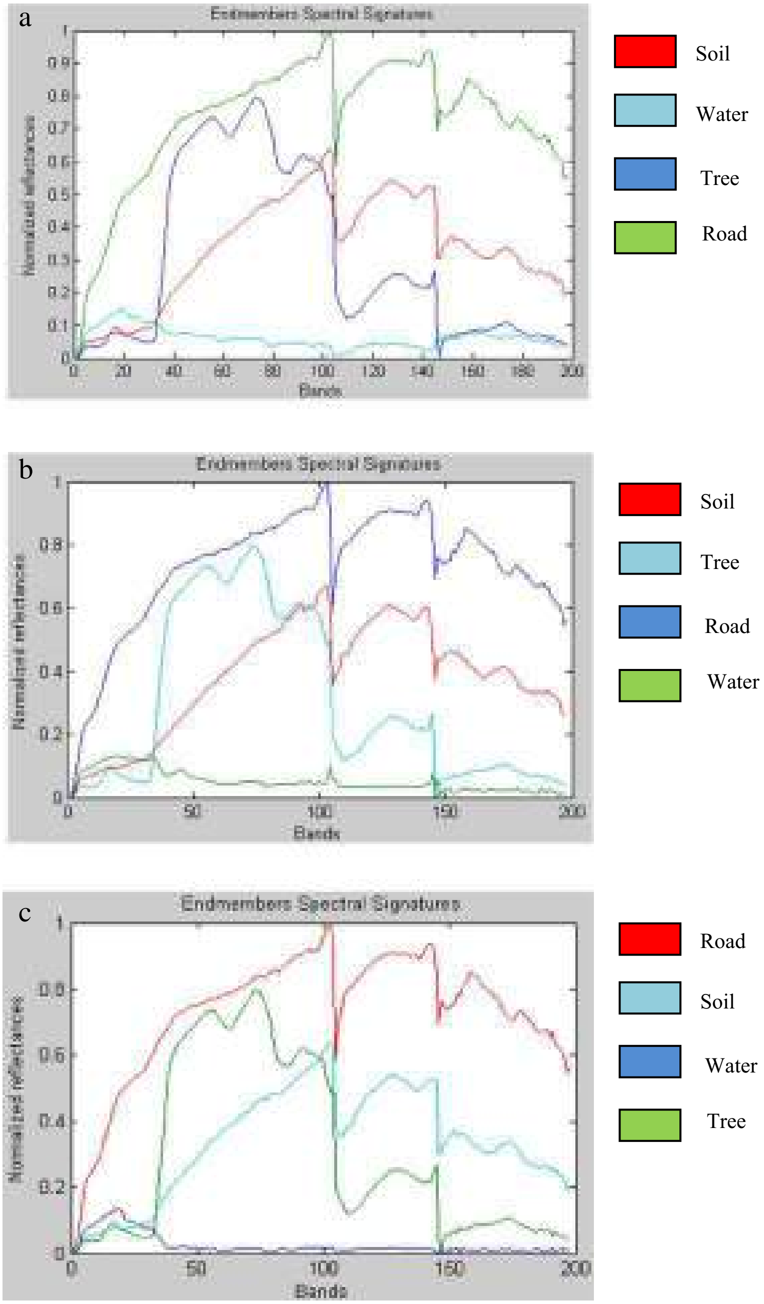


Fig. 4 **a** Spectral signatures of endmembers using N-FINDR. **b** Spectral signatures of endmembers using VCA. **c** Spectral signatures of endmembers using SGA

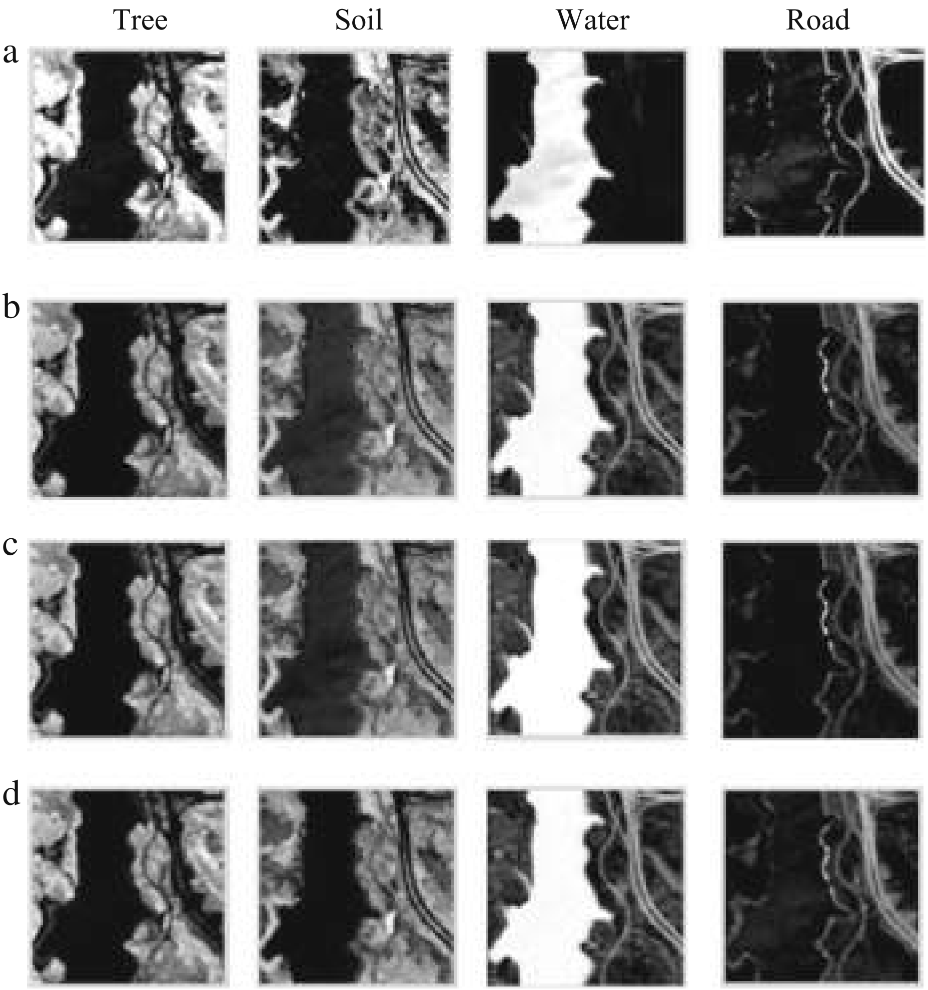


Fig. 5 **a** Groundtruth abundance maps of endmembers. **b** Abundance maps of endmembers using N-FINDR. **c** Abundance maps of endmembers using VCA. **d** Abundance maps of endmembers using SGA

Table 3 Comparison of performance of proposed clustering based band selection with SGA and that of N-FINDR and VCA using SAD

Endmembers	N-FINDR	VCA	SGA
	SAD (%)		
Soil	11.14	28.46	11.14
Tree	15.59	15.59	15.59
Water	46.89	40.31	12.54
Road	10.69	10.69	10.69
Average	21.08	23.76	**12.42**

Table 4 Comparison of performance of proposed clustering based band selection with SGA and that of N-FINDR and VCA using RMSE

Endmembers	RMSE(%)		
	N-FINDR	VCA	SGA
Soil	14.35	14.58	13.72
Tree	17.25	16.24	17.07
Water	19.81	21.25	20.04
Road	12.60	11.78	10.99
Average	16.00	15.96	**15.46**

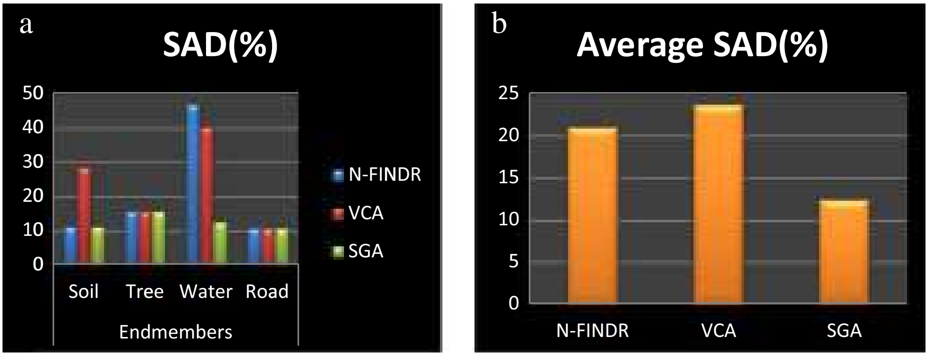


Fig. 6 **a** SAD(%) for different endmembers using clustering based band selection with N-FINDR,VCA and SGA algorithm. **b** Average SAD(%) obtained using clustering based band selection with N-FINDR,VCA and SGA algorithm

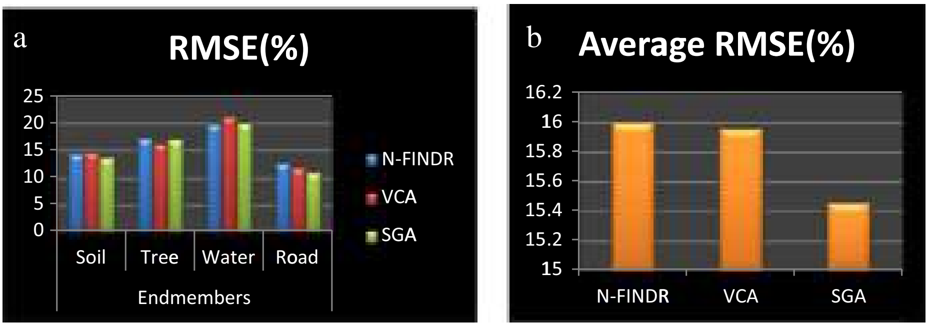


Fig. 7 **a** RMSE(%) for different endmembers using clustering based band selection with N-FINDR,VCA and SGA algorithm. **b** Average RMSE(%) obtained using clustering based band selection with N-FINDR,VCA and SGA algorithm

Table 5 Computer environment used for implementing N-FINDR, VCA and SGA

CPU	Memory	OS
Intel (R) Core (TM)2 Duo E7500@2.93GHz	2.00 GB	Windows 7

are slightly varied, this small variation leads to highly significant result in spectral unmixing. Moreover, the RMSE values obtained for each endmemembers using the proposed algorithm are comparatively low. The reason for superior result in finding endmembers in the proposed work is due to the nature of SGA ie, non-randomness in the initial endmember selection as well as the consistency in producing final endmembers. Clustering based band selection also enhances the performance of the proposed method by selecting the more informative bands. The comparison of performance of the clustering based band selection for endmember extraction using SGA with N-FINDR and VCA algorithm in terms of SAD and RMSE for Jasper Ridge image are clearly depicted in the Figs. 6a, b and 7a, b respectively.

From Figs. 6 and 7, it is very clear that our proposed clustering based band selection with SGA outperforms the clustering based band selection with N-FINDR and VCA algorithms, by extracting endmembers with lower average percentage of SAD and RMSE.

Further, Computation time is also considered for comparison among these EEA. The computer environment used for this implementation in MATLAB has been described in Table 5. The computation time for the implementation of clustering based band selection with three EEAs for Jasper Ridge dataset has been given in Table 6.

It has been already observed in [6] and [7] that VCA algorithm has shown considerable savings in time compared to N-FINDR and SGA. It is proved in Table 6 that VCA has lower computational complexity. But VCA is projection based approach which is entirely different from N-FINDR and SGA, which are simplex based approaches. Hence, VCA algorithm cannot be considered for comparison with N-FINDR and SGA using the computation time. It is also observed from Table 6 that SGA is computationally less expensive than N-FINDR algorithm.

The second dataset is AVIRIS Cuprite image, which is available on the United States Geological Survey (USGS) Website http://aviris.jpl.nasa.gov/. The image scene used in our work is a 224-band subimage of size 350 x 350 pixels. After removing 1–3, 105–115, and 150–170 low SNR and water absorption bands, remaining 189 bands are used for the experimentation. The four number of minerals: Alunite(A), Buddingtonite(B), Kaolinitie(K), and Muscovite(M), which are available in the image scene are considered in this work. The Cuprite image and the groundtruth spectral signatures of four minerals are shown in Figs. 8 and 9.

Table 6 Computation time of clustering based band selection with N-FINDR, VCA and SGA

Endmember extraction algorithm	N-FINDR	VCA	SGA
Computation time (seconds)	4.48	0.16	1.95

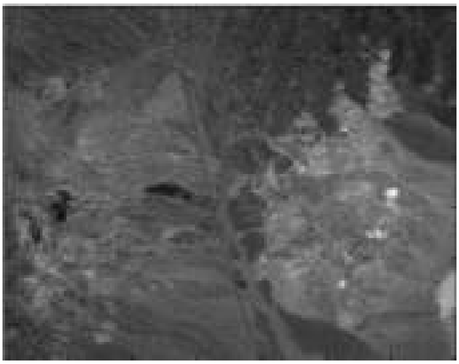


Fig. 8 Cuprite

The number of bands, estimated using the concept of VD for this Cuprite dataset with false alarm probability of 10^{-4} [6], is 22. The bands are selected using the proposed band selection method and the selected bands are shown in Table 7. The selected bands in the Table 7 are used for the subsequent process of endmember extraction. The endmembers are extracted using N-FINDR, VCA and SGA separately and the comparative results are shown in Table 8. Since the groundtruth for abundance fraction of Cuprite dataset is unavailable, the results of EEAs are reported in terms of SAD, which is sufficient for comparison. The comparison of performance of the proposed clustering based band selection for endmember extraction with SGA and that of N-FINDR and VCA algorithms in terms of SAD are shown in Figs. 10 and 11.

It is very clear from Table 8 that the average SAD obtained using proposed clustering based band selection with SGA algorithm is lower than that of VCA and same as that of N-FINDR. But the computational complexity of N-FINDR algorithm is more than that of SGA. The computation time for the implementation of clustering based band selection with three EEAs for Cuprite dataset has been given in Table 9. The computer environment used for this implementation is same as described in Table 5. It is obvious from Table 9 that the proposed clustering based band selection with SGA is nearly five times faster than that of N-FINDR.

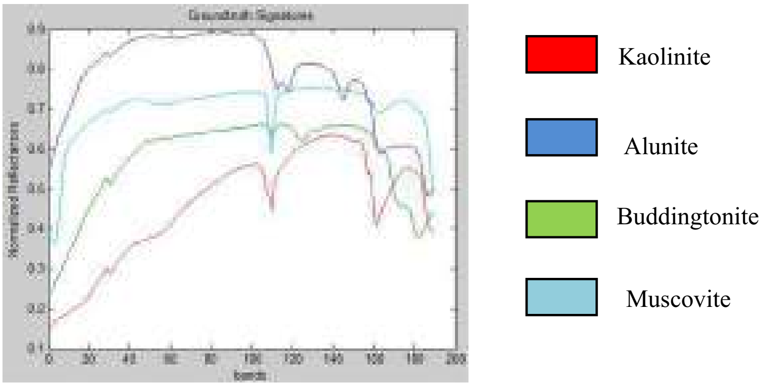


Fig. 9 Groundtruth spectral signatures of four minerals

Table 7 Selected bands for Cuprite

Feature	Bands selected
Variance	5,183,10,157,14,138,145,17,18,20,21,23,102,31,33,131,49,55,125,66,74,189

Further, to prove the significance of our proposed work for real time applications, for Cuprite dataset, the computation time of proposed clustering based band selection with SGA algorithm and transform based DR with SGA [6] is compared in Table 10. Maximum Noise Fraction (MNF) is used prior to EEA for implementing transform based DR technique. The computer environment used for implementing this transform based DR is the same as described in Table 5. The comparison of computation time between transform based DR with SGA and proposed algorithm has been given in Table 10, which shows that our proposed algorithm works faster than transform based DR with SGA.

7 Conclusion and future work

Statistical feature based clustering method is adopted in this work for selection of bands from a high dimensionality hyperspectral image cube. K-means clustering based on variances, which are computed for each band, is used for band selection. SGA algorithm is adopted in endmember extraction. Further, the experimentation with SGA, N-FINDR and VCA has been carried out on selected bands. SGA provides superior performance in endmember extraction than N-FINDR and VCA algorithms because the later cases suffer from the problem of inconsistency in providing final endmembers. It has been observed from the experimental results that the proposed clustering based band selection with SGA produces lower average SAD by 8 to 10 %, compared to that of N-FINDR and VCA algorithms for Jasper Ridge Dataset. Hence, it is proved that our proposed clustering based band selection with SGA yields superior performance in accurate finding of the endmembers. Also, it has been proved in experimentation that our proposed algorithm reduces computation time by 143.92 seconds when compared with conventional transform based DR with SGA for Cuprite dataset. Hence, our proposed algorithm is well suited for real time applications. The future scope of this work is to develop an efficient hardware design for implementing our proposed algorithm in real-world applications.

Table 8 Comparison of performance of proposed clustering based band selection with SGA algorithm and that of N-FINDR and VCA using SAD

Endmembers	N-FINDR	VCA	SGA
	SAD(%)		
Alunite	10.10	18.06	10.10
Buddingtonite	11.34	12.24	11.34
Kaolinite	19.46	14.97	19.46
Muscovite	5.55	14.22	5.55
Average	11.61	14.87	11.61

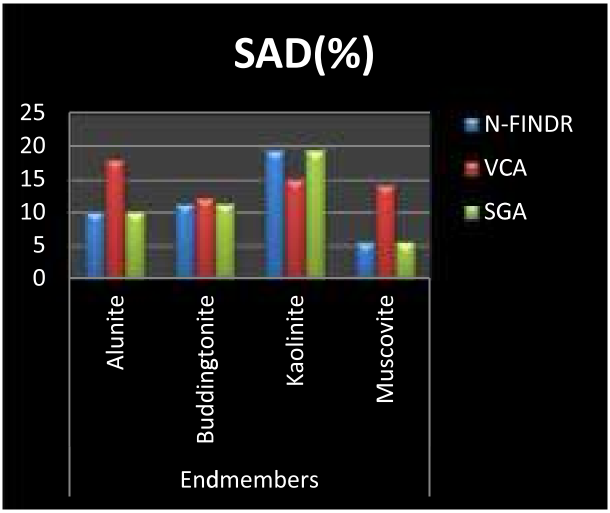


Fig. 10 SAD(%) for different endmembers using clustering based band selection with N-FINDR,VCA and SGA algorithm

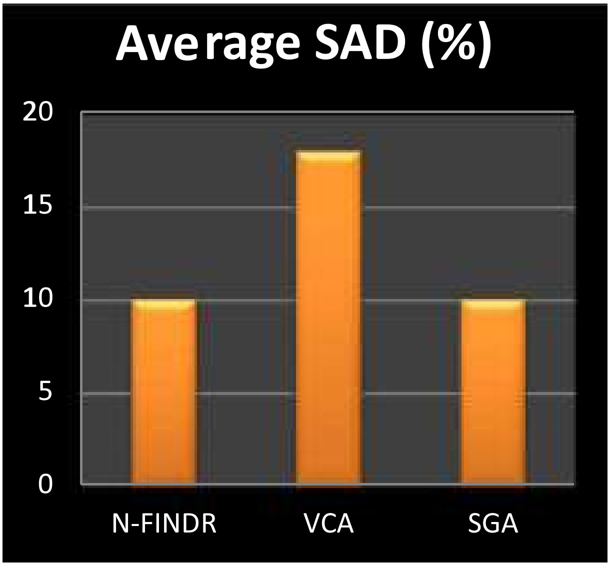


Fig. 11 Average SAD(%) obtained using clustering based band selection with N-FINDR,VCA and SGA algorithm

Table 9 Computation time of clustering based band selection with N-FINDR, VCA and SGA

Endmember extraction algorithm	N-FINDR	VCA	SGA
Computation time (seconds)	90.02	0.41	20.75

Table 10 Computation time of proposed algorithm and transform based DR with SGA

Algorithm	Transform based DR with SGA	Proposed algorithm
Computation time (seconds)	164.67	20.75

Acknowledgments The authors would like to thank Mr. Feiyun Zhu who provided the hyperspectral dataset with ground truth.

References

1. Ahmad M, Haq DIU, Mushtaq Q, Sohaib M (2011) A new statistical approach for band clustering and band selection using K-means clustering. IACSIT Int J Eng Technol 3(6):606–614
2. Chang C-I (2003) Hyperspectral imaging: techniques for spectral detection and classification. Plenum, New York
3. Chang C-I, Du Q (2004) Estimation of number of spectrally distinct signal sources in hyperspectral imagery. IEEE Trans Geosci Remote Sens 42(3):608–619
4. Chang C-I, Du Q, Sun T-L, Althouse MLG (1999) A joint band prioritization and band-decorrelation approach to band selection for hyperspectral image classification. IEEE Trans Geosci Remote Sens 37(6):2631–2641
5. Chang C-I, Wang S (2006) Constrained band selection for hyperspectral imagery. IEEE Trans Geosci Remote Sens 44(6):1575–1585
6. Chang C-I, Wu C-C, Liu W-m, Ouyang Y-C (2006) A new growing method for simplex-based endmember extraction algorithm. IEEE Trans Geosci Remote Sens 44(10):2804–2819
7. Chang C-I, Wu C-C, Lo CS, Chang M-L (2010) Real-time simplex growing algorithms for hyperspectral endmember extraction. IEEE Trans Geosci Remote Sens 48(4):1834–1850
8. Chang C-I, Xiong W, Wu C-C (2013) Field-programmable gate array design of implementing simplex growing algorithm for hyperspectral endmember extraction. IEEE Trans Geosci Remote Sens 51(3):1693–1700
9. Du Q, Yang H (2008) Similarity-based unsupervised band selection for hyperspectral image analysis. IEEE Geosci Remote Sens Lett 5(4):564–568
10. Envi-Tutorials (2013) An overview of hyperspectral remote sensing. http://www.cossa.csiro.au/hswww/Overview.htm
11. Geng X, Sun K, Ji L, Ji L (2014) A fast volume-gradient based band selection method for hyperspectral image. IEEE Trans Geosci Remote Sens 52(11):7111–7119
12. Ghamary Asl M, Mobasheri MR, Mojaradi B (2014) Unsupervised feature selection using geometrical measures in prototype space for hyperspectral imagery. IEEE Trans Geosci Remote Sens 52(7):3774–3787
13. Harsanyi JC, Chang C-I (1994) Hyperspectral image classification and dimensionality reduction: an orthogonal subspace projection approach. IEEE Trans Geosci Remote Sens 32(4):779–785
14. Harsanyi J, Farrand W, Chang C-I et al. (1993) Determining the number and identity of spectral endmembers: an integrated approach using Neyman–Pearson eigen thresholding and iterative constrained RMS error minimization. Proc. 9th Thematic Conf Geologic Remote Sensing
15. Haskell KH, Hanson RJ (1981) An algorithm for linear least squares problems with equality and nononegativity constraints generalized. Math Prog 21:98–118
16. Heinz DC, Chang C-I (2001) Fully constrained least squares linear spectral mixture analysis method for material quantification in hyperspectral imagery. IEEE Trans Geosci Remote Sens 39(3):529–545
17. Lawson CL, Hanson RJ (1995) Solving least squares problems. classification in applied mathematics, PA: SIAM, 15
18. Martinez-Uso A, Pla F, Sotoca JM, Garcia-Sevilla P (2007) Clustering-based hyperspectral band selection using information measures. IEEE Trans Geosci Remote Sens 45(12):4158–4171
19. Mitra P, Murthy C, Pal SK (2002) Unsupervised feature selection using feature similarity. IEEE Trans Pattern Anal Mach Intell 24(3):301–312
20. Nascimento JMP, Dias JM (2005) Vertex component analysis: a fast algorithm to unmix hyperspectral data. IEEE Trans Geosci Remote Sens 43(4):898–910
21. Qian Y, Yao F, Jia S (2009) Band selection for hyperspectral imagery using affinity propagation. IET Comput Vis 3(4):213–222
22. Sohaib M, Haq I-U, Mushtaq Q (2013) Dimensional reduction of hyperspectral image data using band clustering and selection based on statistical characteristics of band images. Int J Comput Commun Eng 2(2):101–105
23. Sun K, Geng X, Ji L, Lu Y (2014) A new band selection method for hyperspectral image based on data quality. IEEE J Sel Topics Appl Earth Observ Remote Sens 7(6):2697–2703

24. Winter ME (1999) N-finder: an algorithm for fast autonomous spectral endmember determination in hyperspectral data. Imag Spectrometry V, Proc SPIE 3753:266–277
25. Zhu F, Wang Y, Fan B, Xiang S, Meng G, Pan C (2014) Spectral unmixing via data-guided sparsity. IEEE Trans Image Process 23(12):5412–5427
26. Zhu F, Wang Y, Xiang S, Fan B, Pan C (2014) Structured sparse method for hyperspectral unmixing. ISPRS J Photogramm Remote Sens 88:101–118

G.Veera Senthil Kumar received his B.E degree from Mohamed Sathak Engineering College, Kilakarai, Madurai Kamaraj University, TamilNadu, India and M.E. Degree from Mepco Schlenk Engineering College, Sivakasi, Tamil Nadu, India. He is doing his research in Hyperspectral Image Processing at Anna University, Chennai. He is working as Assistant Professor in Electronics and Communication Engineering Department, Velammal College of Engineering and Technology, Madurai, Tamil Nadu, India. He has published over 20 Technical papers in International Journals, International Conferences and National Conferences. He is a Life member of Indian Society of Remote Sensing, Life member of Indian Society for Technical Education, and Life member of Broadcast Engineering Society(India).

Dr.S.Vasuki received her B.E degree from Government College of Technology, Coimbatore and obtained her M.E Degree from A.C.College of Engineering and Technology, Karaikudi, TamilNadu, India and Ph.D degree in Color Image Processing at Anna University, Chennai. She is working as Professor and Head in Electronics and Communication Engineering Department, Velammal College of Engineering and Technology, Madurai, Tamilnadu, India. She has published over 110 Technical papers in International Journals, International/ National Conferences. She is a Life member of Indian Society for Technical Education and Fellow of Institution of Electronics and Telecommunication Engineers and member in Institute of Electrical and Electronics Engineers.