



A Superpixel and Mutual Information Based Band Selection Method for Hyperspectral Images

Sonia Sarmah¹, Sanjib Kr. Kalita²

¹Assistant Professor, Assam Don Bosco University, India, sarmahsonia07@gmail.com

²Assistant Professor, Gauhati University, India, sanjib959@gauhati.ac.in

ABSTRACT

Dimension reduction is an important task in hyperspectral image (HSI) processing to minimize computational complexity. This work presents an efficient band selection method to select a set of highly relevant yet non-redundant bands using class-band mutual information (MI) and clustering technique. In HSI, the bands comprise continuous reflectance values, while the class labels in the ground truth are discrete. Thus before calculating class-band MI, the continuous reflectance values are converted to several discrete values using a novel graph-based superpixel generation method based on the spectral angle divergence (SAD). To ensure non-redundancy of the selected bands, clustering is performed to using correlation distance and then from each cluster only one representative band is selected using the calculated class-band MI. Experiments are carried out over three hyperspectral images to evaluate the efficacy of the proposed band selection technique. For each image, classification is performed using Support Vector Machine (SVM) classifier over both the compressed image with only the selected bands and the original image. The comparison results reveal the capability of the proposed method to maintain the classification accuracy with a substantially compact set of selected bands.

Key words: Band selection, clustering, correlation, hyperspectral image, mutual information, SAD, superpixel.

1. INTRODUCTION

HSIs contain hundreds of spectral bands with relatively narrow bandwidths (5-10 nm). A HSI is represented as a data cube with spatial information stored in the first two dimensions and spectral information represented along the third dimension. These high dimensional data cubes contain ample information about the objects and thus enable the classification of spectrally unique materials with high precision. However, due to the increased number of spectral bands, more processing time is required for analyzing such images. Therefore, dimension reduction of HSI by selecting only the

significant bands, without compromising the information content, has been an active area of research [1], [2]. In literature, many criteria such as- divergence, Bhattacharya distance, entropy has been used for the selection of bands that are crucial and significant in terms of information conservation [3], [4]. In [5], Nakamura et al proposed a band selection method based on optimum path forest (OPF) classifier [6] as an optimization function. In another technique, A.C. S Santos et al used k-means clustering with entropy filtering for band selection and classification in hyperspectral images [7]. They used correlation distance to cluster the similar bands, from each of which, they selected the band closest to the cluster center. Next, they performed classification over the reduced image constructed with only the selected bands. In addition, they also reduced the image spatially by using 9X9 entropy filter before classification.

In this work, we have proposed an algorithm that aims to select a set of non-redundant bands having high relevance to the class information. Non-redundancy is taken care of by selecting only the bands that are loosely correlated with each other. In [3], authors effectively utilized the correlation that exists among the adjacent bands of HSI to select non-redundant bands. Class-band MI, a widely used criterion [8], has been used for measuring the relevance of the bands with the class information. But, the calculation of the MI is based on probability, which tends to get more involved for datasets with continuous values. In [9] authors have explained this problem by considering two random variables $X = \{1, 1.1, 0.9, 5, 1.12, 5.1, 1.13, 0.89, 0.87, 1.15\}$ and $Y = \{a, a, b, a, a, a, b, b, a, b\}$ of type continuous and categorical respectively. Mathematically, $P(1)$, the probability of occurrence of 1 in X is $1/10$. Similarly, $P(5)$ in X also evaluates to $1/10$. But, intuitively, 1 is more likely to repeat in X than 5 as the majority of the values are much closer to 1. This sort of problem does not arise in categorical datasets such as Y .

In HSI, each pixel has an associated reflectance vector consisting of N reflectance values, one for each of the N spectral bands. The reflectance values are continuous

and thus the reflectance vectors, for pixels belonging to the same object, tend to be similar but may not be the same. On the other hand, the class labels in the ground truth are of type categorical. In [10] Swarnajyoti et al, to convert the continuous reflectance values to discrete values, used a basic equal-width interval binning approach (uniform quantization) [11]. In this algorithm, the minimum and maximum values of the discrete variables were calculated and the range was then divided into user-defined number discrete intervals of equal-width. In a subsequent step, the relevance and importance of the bands were determined using the rough set theory [12]. The authors then selected the k highly informative bands based on the relevance score and significance of the bands.

In this work, to address this issue with MI calculation in HSI, we have proposed a graph-cut based superpixel generation method. Superpixel is a state of the art technology to find the patches of similar neighboring pixels and label those as being of the same type [13]. The authors in [14] introduced a graph-based superpixel generation method for HSI, where the input HSI was represented as an undirected weighted graph. The primary advantages of this algorithm were that it did not require the number of superpixels to be pre-defined and also could be directly applied to 3-D HSI without any transformation. Owing to these advantages, the proposed superpixels generation method also adapts a similar approach using graph-cut and spectral angle divergence (SAD). After the generation of the superpixels, all the pixels belonging to a superpixel are assigned with the same reflectance vector. Though this step results in some amount of information lost, this simplifies the calculation of MI by using the same reflectance vector for similar pixels. Section 2 presents a detailed description of the proposed method.

2. PROPOSED METHODOLOGY

The proposed band selection algorithm is divided into three phases - a) generation of superpixels, b) calculation of class-band MI and c) band clustering and selection. A detailed description of each phase is presented in the following sub-sections. Figure 1 presents the block diagram of the proposed methodology.

2.1 Generation of Superpixels

In this phase, superpixels are generated for the input HSI using a graph-cut based method. In the proposed method, initially, a graph $G(V, E)$ is constructed using the input HSI, I . Let I consists of ' n ' pixels and ' N ' spectral bands. Each pixel is treated as a node in G and from each node; edges are created to its 8-connected spatial neighbourhood. The weight, $w(e_{ij})$, of an edge e_{ij}

connecting two vertices v_i and v_j , is calculated using a divergence function $d(v_i, v_j)$ based on SAD which is defined as [14]-

$$w(e_{ij}) = d(v_i, v_j) = 1 - \sum_{\lambda} \frac{\rho_{i\lambda} \rho_{j\lambda}}{|\rho_i| |\rho_j|} \quad (1)$$

Where, $\rho_{i\lambda}$ is the reflectance value of the i^{th} pixel at a band with wavelength λ . In the next step, the edges whose weights exceed a predefined threshold, δ , are deleted from the graph. Next, the connected components of the graph are extracted. Let the number of connected components be n' , where n' may vary from 1 to n . Each component results in a superpixel consisting of the nodes (pixels) belonging to that component. Thus, a total of n' ($n' \leq n$) superpixels are generated for I . Next, for each superpixel, the mean spectral reflectance vector is calculated by averaging the reflectance vectors of all the pixels belonging to that superpixel. The calculated mean reflectance vectors are then assigned to all the pixels of the corresponding superpixel. This results in a transformed image I' , having the same dimensions as I , but with modified reflectance values. Algorithm 1 presents the proposed graph-cut based superpixel generation method.

Algorithm 1: A graph-cut based superpixel generation method

Input: Hyperspectral Image (I), δ

Output: A superpixel based transformed image I'

Step 1: Construct a weighted undirected graph $G(V, E)$ by considering each pixel of I as a vertex. Connected each vertex to its 8-connected spatial neighborhood vertices by an edge.

Step 2: For each edge e in E , Calculate its weight $w(e)$ using equation 1.

Step 3:

for each edge e belonging to E **do**

if $w(e) > \delta$, **then**

 delete the edge e ;

end

end

Step 4: Find the set of all connected components of the graph G .

Step 5: Construct a superpixel map from G and a transformed image I'

for each component C_i in G **do**

 Mark all the vertices in C_i as the same superpixel number S_i

r_i = average reflectance vector of all the pixels in S_i

for all the pixels in S_i **do**

 set reflectance vector of the pixel to r_i

end

end

Step 6: Stop

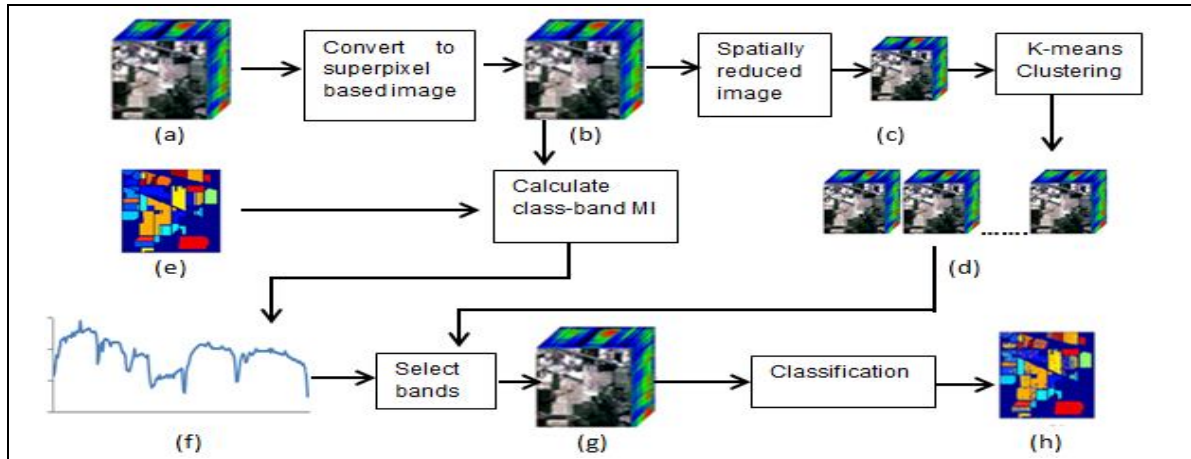


Figure 1: a) Input image with n number of pixels and N number of spectral bands b) Image with n number of pixels, n' number of superpixels and N number of spectral bands c) Spatially reduced image with n' number of pixels and N spectral bands d) k clusters of spectral bands e) Ground truth f) Normalized class-Band MI g) Image with n number of pixels and k selected bands and h) Classified image

2.2 Calculation of Class-Band MI

In this step, the normalized mutual information (NMI) of the bands from Γ are calculated with the referenced ground. Mutual information measures the amount of information that two variables X and Y share. Mathematically, the MI between two discrete random variables X and Y is defined as [15]-

$$I(X;Y) = \sum_{x \in X} \sum_{y \in Y} p(x,y) \log \frac{p(x,y)}{p(x)p(y)} \quad (2)$$

Here $p(x)$, $p(y)$ and $p(x,y)$ are the probability distribution of X , probability distribution of Y and the joint probability distribution of X and Y respectively. Though MI is a good criterion for similarity measure, it is not bounded. Thus, for our work, we have used the normalized mutual information as defined by authors in [16]. Mathematically, it can be expressed as-

$$NMI(X,Y) = \frac{I(X,Y)}{\min(H(X), H(Y))} \quad (3)$$

Where, $H(X)$ and $H(Y)$ are the entropies of the variables X and Y respectively. Entropy for a given discrete random variable X with probability distribution $p(x)$ is defined as [15]-

$$H(X) = - \sum_{x \in X} p(x) \log(p(x)) \quad (4)$$

2.3 Band Clustering and Band Selection Phase

In this phase, highly correlated bands are grouped to form k ($k \leq N$) clusters using the k-means clustering algorithm [17]. We have used correlation as the distance metric, which for two vectors X and Y may be defined as [7]-

$$DC(X,Y) = 1 - \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2 \sum_{i=1}^n (y_i - \bar{y})^2}} \quad (5)$$

Where, n is the length of the vectors X and Y . \bar{x} and \bar{y} are the means of the vectors X and Y respectively. Clustering is performed by considering each of the N spectral bands as a sample and the pixels as feature vectors. However, instead of using all the pixels, we are considering only one pixel from each superpixel. Thus, the length of the feature vector for each band is n' . This decreases the computation time of the band selection phase as generally n' is much smaller than n . Next, from each of the k clusters, the band having the highest class-band mutual information score (calculated in the second phase) is selected. Thus the number of selected bands by the algorithm is equal to the number of clusters. This set of selected bands is later used for training and testing the classifier.

3. EXPERIMENTAL SETUP

3.1 Dataset Description

For the experimental analysis of the proposed work, we have used three hyperspectral images – Botswana, Indian Pines (IP) and Salinas.

The Botswana image was acquired by NASA Earth Observing-1 satellite over the Okavango Delta, Botswana, using 242 bands. Noisy bands that cover water absorption features were removed, and the remaining 145 bands were included as candidate features. The image contains a total of 377856 pixels, out of which only 3248 are labelled. The ground truth consists of 14 classes.

Table 1: Number of labelled samples (class wise) for the different datasets

Class#	#Samples Botswana	#Samples IP	#Samples Salinas
C1	270	46	2009
C2	101	1428	3726
C3	251	830	1976
C4	215	237	1394
C5	269	483	2678
C6	269	730	3959
C7	259	28	3579
C8	203	478	11271
C9	314	20	6203
C10	248	972	3278
C11	305	2455	1068
C12	181	593	1927
C13	268	205	916
C14	95	1265	1070
C15	N/A	386	7268
C16	N/A	93	1807
# labeled Samples	3248	10249	54129
# Unlabeled samples	374608	10776	56975

The Indian Pines scene was acquired by the AVIRIS (Airborne Visible/Infrared Imaging Spectrometer) sensor over the Indian Pines test site in North-western Indiana and consists of 145×145 pixels and 224 contiguous spectral bands. A corrected version of the original dataset containing only 200 bands (by removing the 20 bands covering water absorption region) is also available. The reference ground truth consists of 10249 labeled samples from 16 classes.

The Salinas scene was also acquired by the AVIRIS sensor over Salinas Valley, California using 224 spectral bands. The image contains 512×217 pixels. The corrected image consists of 200 bands by discarding the 20 water absorption bands. The ground truth contains 54129 labelled samples from 16 classes [18]. Table 1 presents the different classes and the corresponding number of pixels for the three dataset.

3.2 Threshold Selection and Superpixel generation

The number of superpixels generated in the first phase of the proposed algorithm is directly dependent on the input threshold, δ . A very low threshold results in over-segmentation by grouping similar pixels into different superpixels. Alternatively setting a high threshold produces a relatively small number of superpixels at the cost of merging of pixels from different classes. Due to this, the selection of δ is a crucial task to generate a suitable number of superpixels. For our experiment, for a dataset, we have first calculated the average intra-class SADs for each class using

the ground truth. Table 2 presents the average intra-class divergences for the datasets. Then, inter-class SADs were computed by finding differences in average intra-class SADs for each pair of classes. The threshold, δ , for the dataset was then set to the average of the inter-class SADs. The calculated threshold values and the corresponding number of superpixels are reported in table 3.

3.3 Setting the value of k

As already stated in the proposed methodology, the number of selected bands is equal to the number of clusters formed in the band clustering phase. In k-means clustering, the number of desired clusters has to be pre-defined. For this work, we have repeated the band clustering phase of the algorithm for different k values ranging from 5 to 50 in a step of 5. For each k value, a reduced dataset was constructed by retaining only the selected bands from the original image.

Table 2: Average intra-class divergence of the different datasets

Class#	Botswana	IP	Salinas
C1	0.013912	0.00150189	0.9790839
C2	0.00461	0.00300793	0.9623952
C3	0.00091	0.00210921	0.5843338
C4	0.001271	0.00416295	0.8327617
C5	0.00262	0.00892097	0.7406048
C6	0.002996	0.00286835	0.4632312
C7	0.00496	0.00084361	0.5910876
C8	0.001036	0.00190302	0.460925
C9	0.001757	0.00101102	0.5912938
C10	0.000811	0.00234199	0.6797362
C11	0.000639	0.00246688	0.7773878
C12	0.000659	0.00464918	0.8678898
C13	0.001057	0.0009779	0.9712606
C14	0.002622	0.001883	0.9790839
C15	N/A	0.00605556	0.9623952
C16	N/A	0.00292895	0.5843338

Table 3: Threshold values and the corresponding number of superpixels the datasets

Dataset	#Pixels (n)	Threshold (δ)	#Superpixels (n')
Botswana	21025	0.00324	20137
IP	111104	0.00231	2130
Salinas	377856	0.17216	9468

3.4 Designing of the Classification Model

To examine the effectiveness of the selected bands, classification was performed over the compressed image. The pixels were treated as samples and the reflectance values of a pixel, across the selected bands, were considered as the features. The classification was performed using multiclass Support Vector Machine (SVM) with the one-against-one

strategy with the polynomial kernel. SVM has been used successfully for classification of hyperspectral images owing to its capability of handling large input space [19]. Two parameters, C (soft margin cost function) and d (degree of the polynomial) were tuned by performing grid search over the normalized IP dataset using 5-fold cross-validation. For this purpose, the ranges of C and d were set to 0.01–100 and 1–15, respectively. The best combination reported by grid search was $C=40$ and $d=5$, which was used for the other two datasets as well. The model was then trained and tested using stratified 10-fold cross-validation. The classification results were reported in terms of overall accuracy (OA) and kappa score.

4. RESULTS AND DISCUSSION

Initially, for all the three datasets, we constructed the superpixel based transformed images (I') using the thresholds given in table 3. As can be seen from the table, the numbers of generated superpixels were much less than the actual number of pixels in the original images. Next, for each dataset superpixel based transformed images were constructed. Figure 2 shows the band number 25 of the actual (I) and the transformed (I') images of the Botswana dataset. Figure 3 shows the band number 178 from the actual and the transformed images of the IP dataset. Similarly, for the Salinas dataset, figure 4 shows the band number 178 of the original and the transformed images. The differences between the original and the transformed images are evident from these figures.

Then, we calculated the normalized class-band NMI of the bands from transformed images with the corresponding ground truth. We have measured the same for the bands from the original images for comparison. Figures 5-7 display the calculated normalized class-band MI for the three datasets. The values for the original images are shown using green lines and that for the transformed images are shown using blue lines.

Next, we performed band clustering and band selection. The process was repeated for both the original and the transformed images. The resultant sets of selected bands were different in both cases. Two reduced datasets, I_{red} and I'_{red} were created with the selected bands. The first one was created using the selected bands for the original image and the second using the selected bands for the transformed image. The classification was carried out independently on both I_{red} and I'_{red} . The classification results in terms of overall accuracies and kappa scores for the different datasets are presented in table 4. Figures 8-10 present the results in the form of graphs, where the blue and the green lines reflect the results obtained for I_{red} and I'_{red} respectively.

Figure 8 presents the classification results obtained for the Botswana dataset with a different number of selected bands. It was observed that the performance of the I'_{red} was much better than I_{red} . With only 30 selected bands, for I'_{red} , we achieved

an overall accuracy of 98%. Table 5 reveals that it was an enhancement over the result (94.66%), obtained by Swarnajyoti et al [10], using the same number of bands.

Figure 9 displays the measured overall accuracy and kappa score for the reduced datasets with a different number of selected bands. It was noted that the accuracy obtained by I'_{red} outperformed the accuracy obtained by I_{red} . For I'_{red} , with only 25 bands, we could gain an accuracy of 89.86%. This accuracy, as can be observed from table 5, was better than the accuracy obtained by the authors in [5] and [10]. In [5], authors achieved the best accuracy of 85.40% using 25 selected bands. In [10], authors achieved an accuracy of 89.66% using 40 selected bands. However, the results reported by A.C. S Santos et.al. [7], for IP dataset was still better compared to the proposed method.

The last experiment was conducted over the Salinas dataset. From table 3, it is noticed that the number of superpixels produced for the dataset was very small, barely 0.08% of the original number of pixels. Figure 10 gives the overall accuracy and kappa score achieved for the reduced datasets. It was seen that the overall accuracy and kappa score was much better in the case of I_{red} than I'_{red} . This might be due to the inadequate number of superpixels in the transformed image, which resulted in a substantial amount of information loss. Nevertheless, the performance of the selected bands by the proposed band selection algorithm, for the original image, was still better than the results obtained in [5] and [7]. In [5], authors achieved accuracy of 97.70% using almost 100 selected bands and in [7], using 30 bands authors achieved an accuracy of 97.10%. However, for the original image, with only 25 selected bands by the proposed method we could achieve an accuracy of 97.02%.

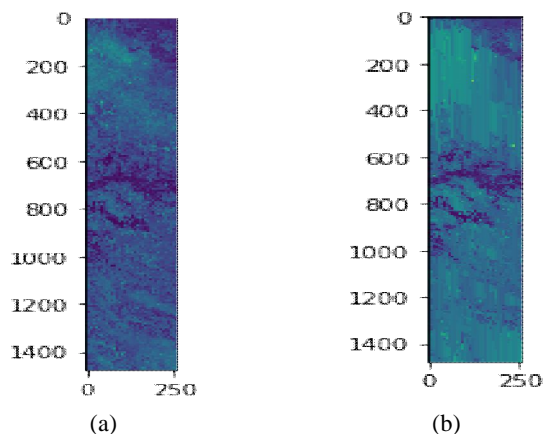


Figure 2: Band no. 25 of the Botswana Dataset- (a) Original image (I) and (b) Transformed image (I') for the threshold δ

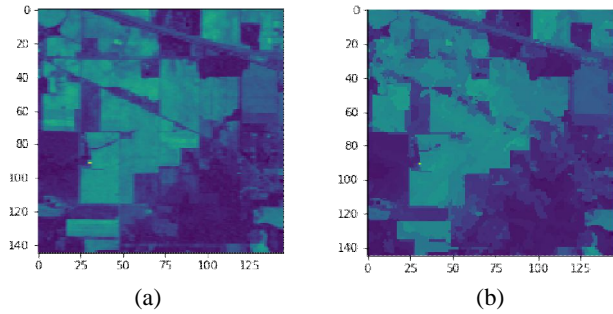


Figure 3: Band no. 178 of the IP Dataset- (a) Original image (I) and (b) Transformed image (I') for the threshold δ .

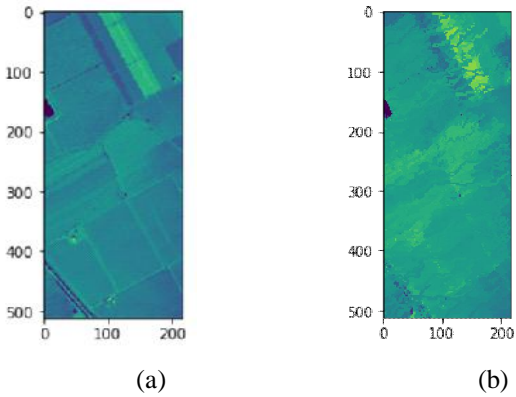


Figure 4: Band no. 178 of the Salinas Dataset- (a) Original image (I) and (b) Transformed image (I') for the threshold δ .

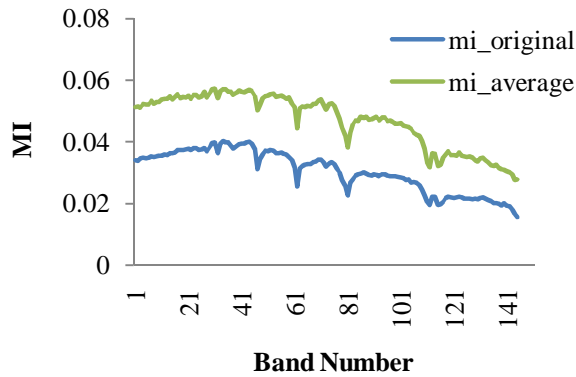


Figure 5: Normalized Class-Band MI for Botswana Dataset

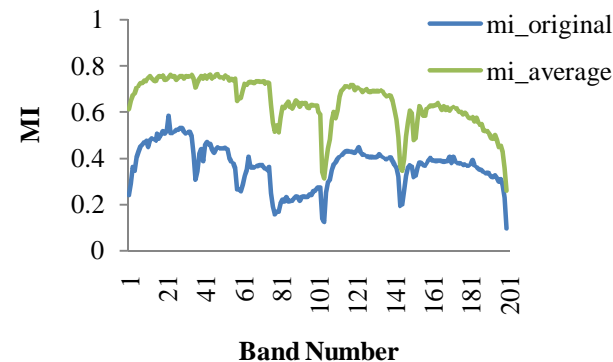


Figure 6: Normalized Class-Band MI for the IP Dataset

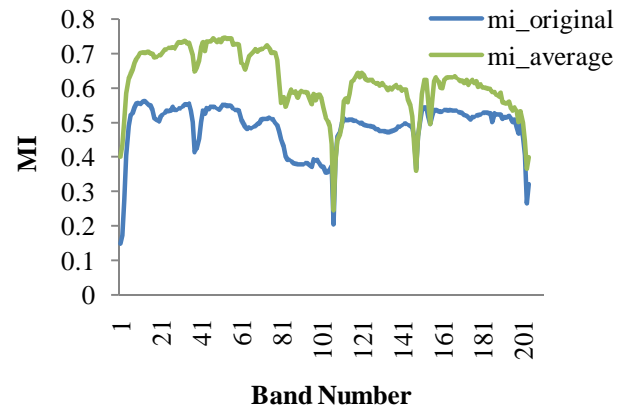
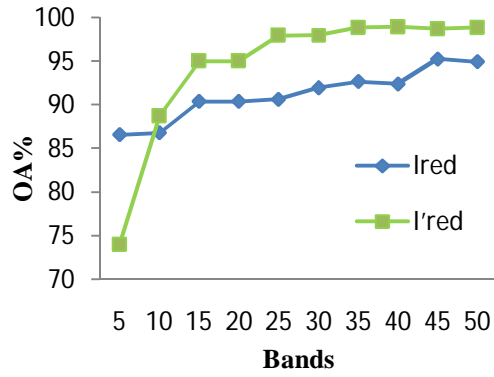


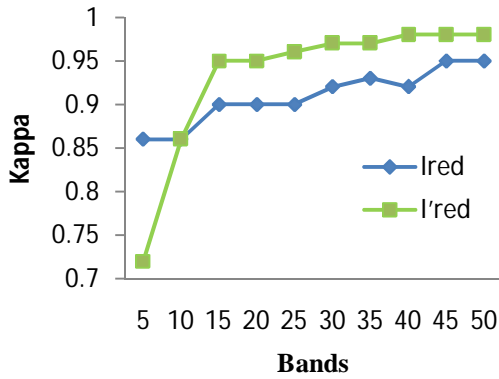
Figure 7: Normalized Class-Band MI for the Salinas Dataset

Table 4: Classification results in terms of overall accuracy and kappa score for the different dataset

Dataset	#Selected Bands	OA(%)		Kappa	
		I_{red}	I'_{red}	I_{red}	I'_{red}
Botswana	5	86.58	73.98	0.86	0.72
	10	86.79	88.73	0.86	0.86
	15	90.39	95.00	0.90	0.95
	20	90.39	95.00	0.90	0.95
	25	90.67	97.94	0.90	0.96
	30	91.96	98.00	0.92	0.97
	35	92.67	98.87	0.93	0.97
	40	92.40	98.94	0.92	0.98
	45	95.28	98.72	0.95	0.98
	50	94.95	98.87	0.95	0.98
IP	5	65.04	65.17	0.58	0.60
	10	73.66	77.92	0.78	0.78
	15	75.62	85.92	0.71	0.86
	20	79.70	88.29	0.76	0.88
	25	79.52	89.86	0.76	0.90
	30	81.50	89.40	0.78	0.90
	35	81.98	89.47	0.79	0.90
	40	82.57	90.02	0.80	0.92
	45	84.72	91.33	0.82	0.92
	50	85.70	91.48	0.83	0.92
Salinas	5	77.68	91.64	0.75	0.88
	10	89.46	94.47	0.88	0.92
	15	95.35	95.92	0.89	0.93
	20	90.38	96.64	0.89	0.94
	25	90.46	97.02	0.89	0.95
	30	90.75	97.21	0.89	0.95
	35	90.68	97.3	0.89	0.95
	40	90.79	97.75	0.90	0.96
	45	91.81	98.14	0.91	0.96
	50	91.25	98.42	0.91	0.97

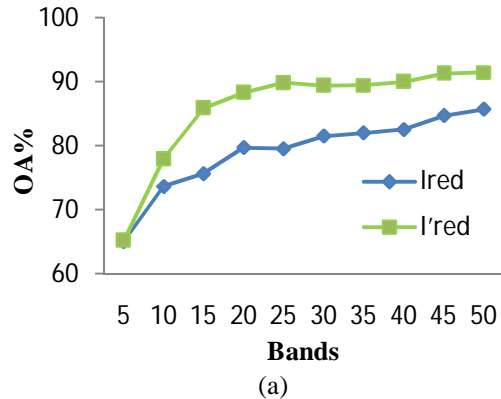


(a)

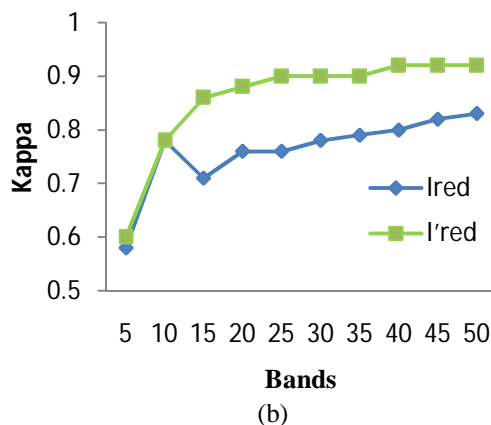


(b)

Figure 8: Classification results for different number of bands of Botswana dataset - (a) Overall Accuracy (b) Kappa

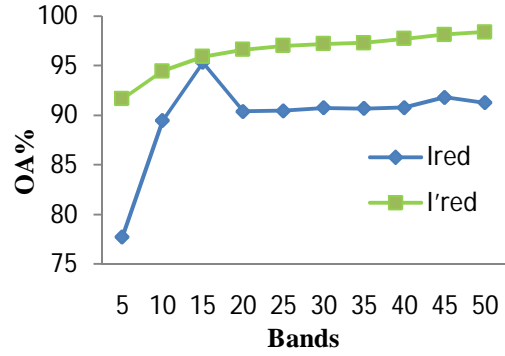


(a)

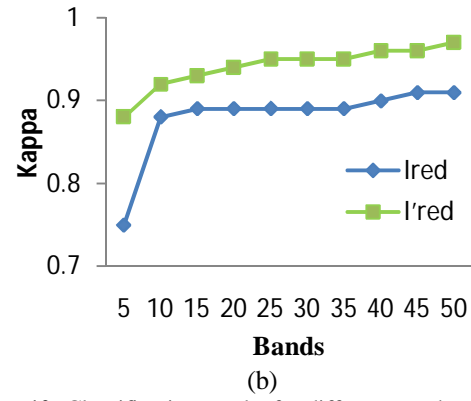


(b)

Figure 9: Classification results for different number of bands of IP dataset - (a) Overall Accuracy (b) Kappa



(a)



(b)

Figure 10: Classification results for different number of bands of Salinas dataset - (a) Overall Accuracy (b) Kappa

Table 5: Comparison of the proposed band selection method with other approaches available in the literature

Dataset	Authors	#Selected Bands	OA (%)
Botswana	SwarnajyotiPatra et al	30/145	94.66
	Proposed	30/145	98.00(I'red)
IP	Nakamura et al.	25/220	85.4
	SwarnajyotiPatra et al	40/220	89.66
	A. C. S. Santos et al	30/220	97.1
	Proposed	25/200	89.86(I'red)
Salina	Nakamura et al.	≈100/204	93.7
	A. C. S. Santos et al	30/204	97.1
	Proposed	25/204	97.02(I'red)

5. CONCLUSION

In this paper, we proposed a band selection method based on k-mean clustering, correlation, mutual information and superpixels. The paper attempted to address the issues involved with the calculation of mutual information for datasets having continuous values. As the pixels of HSI also comprise of continuous reflectance values, similar issues also occur in such datasets. A superpixel based solution was proposed in this work, using the spectral angle of divergence, to address these issues in HSI. The proposed solution strived

to uniformize the reflectance vectors of neighboring pixels that carry similar information. The modified dataset was then used for the proposed band selection algorithm. This also added to the spatial reduction of the dataset, as in the final phase of the algorithm, i.e. band clustering, instead of considering all the pixels, only one pixel from each superpixel was considered. This reduced the time required for the band clustering and thus sped up the band selection process. It was noted that the proposed technique was capable of selecting a set of relevant bands and thus achieved high classification accuracy with a relatively low number of bands. The experimental results for Botswana and IP datasets demonstrated that the classification accuracy was much better in case of the reduced dataset constructed using the superpixel based transformed image. However, it was also observed that if the number of superpixels was too less, as in the case of Salinas dataset, the classification accuracy degraded significantly. This is due the reason that the given threshold resulted in merging of pixels from different classes. Hence, the threshold selection is a sensitive issue in the applied graph-based superpixel generation method. For binary classification problem finding a threshold is not that difficult. However, for multi-class problems, selecting one threshold to separate all the classes is a challenging task, as each class has different margins with each of the remaining classes. Thus, this is an open area for future research.

REFERENCES

- [1]. C.I. Chang, Q. Du, T.L. Sun and M. L. G. Althouse, **A joint band prioritization and band-decorrelation approach to band selection for hyperspectral image classification**, IEEE Trans. Geosci. Remote Sens, vol. 37, PP. 2631-2641, 1999.
- [2]. M. Lahlimi and M. A. Kerroumand Y. Faikhri, **Band Selection by Divergence Distance Based on Gaussian Mixture Model for Hyperspectral Image Classification**, Int. J. of Advanced Trends in Computer Science and Engineering, vol. 8, PP. 2330-2338, 2019. <https://doi.org/10.30534/ijatcse/2019/72852019>
- [3]. S. Kumar, J. Ghosh, and M. M. Crawford, **Best-bases feature extraction algorithms for classification of hyperspectral data**, IEEE Trans. Geosci. Remote Sens, vol. 39, PP.1368-1379, 2001.
- [4]. S. D. Stearns, B. E. Wilson, and J. R. Peterson, **Dimensionality reduction by optimal band selection for pixel classification of hyperspectral imagery**, in Applications of Digital Image Processing XVI, Int. Society for Optics and Photonics, PP. 118-127, 1993.
- [5]. R. Y. M. Nakamura, L. M. G. Fonseca, J. A. D. Santos, R. S. Torres, X. S. Yang and J. P. Papa, **Nature-inspired framework for hyperspectral band selection**, IEEE Trans. Geosci. Remote Sens, vol. 52, PP. 2126-2137, 2013.
- [6]. J. P. Papa, A. X. Falcão, and C. T. N. Suzuki, **Supervised pattern classification based on optimum-path forest**, Int. J. Imag. Syst. Technol, PP. 120–131, 2009.
- [7]. A. C. S. Santos and H. Pedrini. **A combination of k-means clustering and entropy filtering for band selection and classification in hyperspectral images**, Int. J. Remote Sens., vol. 37, PP. 3005-3020, 2016.
- [8]. B. Guo, S. Gunn, B. Damper, and J. Nelson, **Adaptive band selection for hyperspectral image fusion using mutual information**, 7th Int. Conf. on Information Fusion, IEEE, vol. 1, PP. 630-637, 2005.
- [9]. N. Hoque, H. A. Ahmed, D. K. Bhattacharyya and J. K. Kalita, **A fuzzy mutual information-based feature selection method for classification**, J. of Fuzzy Information and Engineering, vol. 8, PP. 355-384, 2016. <https://doi.org/10.1016/j.fiae.2016.09.004>
- [10]. S. Patra, P. Modi, and L. Bruzzone, **Hyperspectral band selection based on rough set**, IEEE Trans. Geosci. Remote Sens, vol. 53, PP. 5495-5503, 2015.
- [11]. S. Kotsiantis and D. Kanellopoulos, **Discretization techniques: A recent survey**, GESTS Int. Trans. Comput. Sci. Eng., vol. 32, PP. 47–58, 2006.
- [12]. J. Komorowski, Z. Pawlak, L. Polkowski, and A. Skowron, **Rough-fuzzy hybridization: a new trend in decision making**. Published by Springer-Verlag Berlin and Heidelberg GmbH & Co. K, 1999.
- [13]. R. Achanta, A. Shaji, K. Smith, A. Lucchi, P. Fua, and S. Süsstrunk, **SLIC superpixels compared to state-of-the-art superpixel methods**, IEEE transactions on pattern analysis and machine intelligence, vol. 34, PP. 2274-2282, 2012.
- [14]. A. M. Saranathan and M. Parente, **Uniformity-based superpixel segmentation of hyperspectral images**, IEEE Trans. Geosci. Remote Sens, vol. 54, PP. 1419-1430, 2016. <https://doi.org/10.1109/TGRS.2015.2480863>
- [15]. J. Novovičová, P. Somol, M. Haindl, and P. Pudil, **Conditional mutual information based feature selection for classification task**, In proceedings of Iberoamerican Congress on Pattern Recognition, Springer, Berlin, Heidelberg, PP. 417-426, 2007.
- [16]. Z. Dawy, J. Hagenauer, P. Hanus, and J. C. Mueller, **Mutual information based distance measures for classification and content recognition with applications to genetics**, IEEE Int. Conf. on Communications, vol. 2, PP. 820-824, 2005.
- [17]. J. MacQueen, **Some methods for classification and analysis of multivariate observations**, in Proceedings of the fifth Berkeley symposium on mathematical statistics and probability, vol. 1, PP. 281-297, 1967.
- [18]. **Hyperspectral Remote Sensing Scenes**, www.ehu.es/ccwintco/index.php/Hyperspectral_Remote_Sensing_Scenes.
- [19]. Puruwadi and N. Suryana, N. A. Abu and B. A. Kusuma, **A Comprehensive Review: Classification Techniques on Hyperspectral Remote Sensing**, Int. J. of Advanced Trends in Computer Science and Engineering, vol. 8, PP. 156-164, 2019. <https://doi.org/10.30534/ijatcse/2019/3181.52019>