# Information content vs. class separabilityat optimal spectral regions

Aria, S. E.Hosseini; Menenti, M.; Gorte, B. G.H.

**DOI**
10.1109/WHISPERS.2013.8080747

**Publication date**
2017

**Document Version**
Accepted author manuscript

**Published in**
2013 5th Workshop on Hyperspectral Image and Signal Processing: Evolution in Remote Sensing, WHISPERS 2013

# INFORMATION CONTENT VS. CLASS SEPARABILITYAT OPTIMAL SPECTRAL REGIONS

*S.E. Hosseini Aria, M. Menenti, B.G.H. Gorte*

Geoscience and Remote sensing Department, Delft University of Technology, Delft, Netherlands

## ABSTRACT

One of the main steps in hyperspectral image classification is the selection of bands that provide the best separability among classes. It is usually understood that the selected bands for classification must contain a large amount of information, and the correlation among selected bands should be small to avoid redundancy. At the same time for optimal classification, class separability should be at maximum value. The question arises whether the most informative spectral regions are really the same as the most discriminant ones for a given set of classes. Answering the question, we developed a new method named Spectral Region Splitting (SRS) to identify interesting spectral regions. This article concludes that the optimal informative and the optimal separable spectral regions are not identical. Furthermore, the cause of the difference is proven theoretically.

***Index terms***—information content, separability, spectral region, hyperspectral

## 1. INTRODUCTION

Hyperspectral imagery provides extremely thorough and complete spectral coverage with high spectral resolution. Although the high dimensionality of hyperspectral data increases the ability of image classification and pattern recognition, data redundancy leads to an increase in computational cost. One way to mitigate the so-called curse of dimensionality of hyperspectral data is selection of optimal bands. From a spectral point of view, two factors are noticeable in band selection algorithms, which are also important in image classification and target detection issues. First, the spectral bands have a large amount of information; Second, class separability should be maximized [1, 2]. Since in hyperspectral data, lots of narrow bands lie next to each other; adjacent bands are highly correlated and cause data redundancy. So, the former criterion tries to select the bands which have low correlation among each other in order to decrease redundant data and obtain the optimal information. The other criterion, mostly used to select optimal bands, is the separability for a given set of classes in a scene by supervised feature selection. In many studies, bands are chosen based on maximizing separability between predefined classes [3-5].

This article answers the fundamental question whether the most informative bands and the most separable bands are identical sets. In this research instead of selecting bands, we merge adjacent bands into spectral regions, which should have either the most information or the largest discrimination respectively.

The optimal sets of bands are obtained by a newly developed method called Spectral Region Splitting (SRS). This method is a top-down, iterative algorithm, starting with a single very wide band covering the entire spectrum and recursively splitting it into interesting spectral regions. Two different criteria are used in the SRS method to identify the most informative spectral regions and the most discriminant ones: 1) using correlation coefficient to identify the most informative spectral regions, 2) using transformed divergence (a separability measure) to determine the most discriminant spectral ones.

The SRS method is applied to a famous hyperspectral image, Indiana Pine, collected by the Airborne Visible/Infrared Imaging Spectrometer (AVIRIS) between 400 to 2500 nm wavelength in more than two hundred narrow (in the order of 10 nm) contiguous bands [6]. The scene includes agriculture, forest or other perennial vegetation, and some man-made objects. Before applying SRS noisy bands should be removed.

The result of SRS on Indiana Pine consists of two sets: the informative spectral regions and the discriminant spectral regions. The results show that the two optimal sets are not identical. More details are presented in section 4. Section 2 and 3 give more details of the data set and provide a full description of the spectral region splitting or SRS method respectively. Section 5 discusses the reason for the differences between two spectral band sets, and section 6 summarized the study.

## 2. DATASET

The scene was collected by the AVIRIS sensor over the Indian Pines test site in North-western Indiana. It consists of 145*145 pixels and 224 spectral reflectance bands in the wavelength range 0.4–2.5 nanometers. Two-thirds of the Indian Pines scene is covered by agriculture, and one-third by forest and other natural perennial vegetation. The ground truth available is designated into sixteen classes, which unfortunately are not all mutually exclusive. Since the classes almost cover three fourth of the whole scene, we use these parts of the scene to identify the most informative and discriminant spectral regions. We first reduce the number of bands to 185 by removing bands covering the region of water absorption (104-108, 150-163, 220) and noisy bands (1-3, 103, 109-112, 148-149, 164-165, 217-219). Indian Pines data are available through Purdue's university website:https://engineering.purdue.edu/~biehl/MultiSpec/hyp-erspectral.html.

## 3. METHOD

In this section, subdividing the spectrum for identifying the optimal informative spectral regions and the optimal separable ones is investigated by the SRS method. As already mentioned the SRS method is a top-down algorithm; it starts with a single very wide spectral band covering the complete measured spectral range. The signal in this band consists of the average of all spectral channels. It has no spectral detail. In the second step, the wide band is split into two parts, and the location of split is determined based on the criterion introduced to the program: either for finding informative or discriminant spectral regions. So the split scans all possible situations and finds the best dividing location concerning the criteria. The algorithm continues recursively, and each time gives finer divisions until it reaches to a termination point. So, by determining a criterion and a termination point based on the interested spectral regions, the SRS algorithm can be summarized as follows:

*1) Average all the spectral bands.*

*2) A split scans the entire cases and computes parameters based on the criterion.*

*3) Find the best location for the split.*

*4) Check the termination point; if not, return to step 2.*

*5) End; the interested spectral regions is obtained by the splits minus the last one.*

The first criterion for extracting the most informative spectral regions is based on information theory [7], which emphasizes that by having more independent data, information content is increasing. The optimal location of the split is where the largest information increase is obtained. When the wide band is split into two parts, the information on spectral detail will improve; most if the new bands have minimum dependency. In this research, we used correlation coefficient to compute the relationship between two spectral divisions. In fact, correlation coefficient investigates linear dependency for two bands. The mathematical representation of correlation coefficient ($R_{ij}$) for bands *i* and *j* is as follows:

$$R_{ij} = \frac{\sum_{p=1}^{n}(x_{ip}-\mu_i)(x_{jp}-\mu_j)}{\sqrt{\sum_{p=1}^{n}(x_{ip}-\mu_i)^2}\sqrt{\sum_{p=1}^{n}(x_{jp}-\mu_j)^2}} \qquad (1)$$

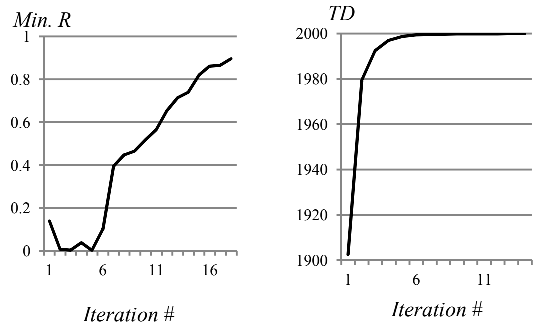


Fig.1. Convergence iterative scheme towards the termination point for correlation coefficient (left) and transformed divergence (right).

where, $x_{ip}$ and $x_{jp}$ are the *pth* pixel value of bands *i* and *j*, $\mu_i$ and $\mu_j$ represent the mean of the bands. In the case of comparison and finding minimum dependency, our case, the absolute value of *R* is used. Since *R* just computes the dependency between two bands, sometimes new selected spectral regions may highly correlated with the adjacent regions which already obtained. Therefore a checking step is added to SRS for identifying informative spectral regions after step 3 as below:

*3.a) Check the correlations between adjacent bands with the new one; if highly correlated, they are averaged.*

The termination point is also identified based on the dependency concept of the produced bands by the SRS method. As mentioned, in this algorithm when a very wide band, made up of the entire measured spectrum, is split into two parts, the algorithm scans all the situations again to find the location of the new split which has the minimum dependency, and then three regions are obtained. By continuing this procedure, increasingly refined spectral regions will be given. In each iteration the minimum correlation, which is at the location of split, is mostly higher than in the previous one (Fig. 1. left). Eventually, the algorithm reaches the point where there are just highly correlated spectral regions, and this point is the termination point. To identify a threshold for highly correlated spectral regions, it is assumed that the correlation coefficient distribution of all band combinations is normal. Then, correlation coefficient matrix of the scene is computed (Fig. 2). In this figure, the bright parts reveal more correlation, which are mostly placed at neighbouring bands. Therefore, we considered that 25% of all combinations in the correlation matrix are highly dependent which is equal to 0.879 for Indiana pine scene. This value is calculated based on standardization of the correlations and use of normal distribution table. The threshold is also used for merging highly correlated bands in step 3.a. of the algorithm.

In the second implementation of the SRS method, the optimal spectral bands are the ones which give the maximum separability for per-defined classes. This

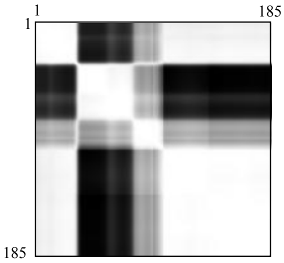

Fig 2. Correlation matrix of bands

time a separability measure, Transformed Divergence (TD), is considered as the criterion for determining locations of the splits. This measure considers variances of class distribution. (Eq. 2).

$$TD = 2000 * [1 - e^{D/8}] \qquad (2)$$

where $D$is divergence distance computed for two classes $a$ and $b$ as follows:

$$D = \frac{1}{2} tr[(V_a - V_b)(V_b^{-1} - V_a^{-1})] + \frac{1}{2} tr[(V_a^{-1} + V_b^{-1})(\mu_a - \mu_b)(\mu_a - \mu_b)^T] \qquad (3)$$

where $V_a$ and $\mu_a$ are the covariance matrix and the mean value for class $a$ respectively[8]. Splitting based on the separability measure means after making the wide band, a split scans all of situations and computes the separability. The best location of split is obtained wherever the discrimination for given classes is maximum. The TD value is between zero and 2000, and in each iteration, this value is not decreasing (by considering that the number of bands is increasing each time; Fig.1.right); hence, the difference of TD in two consecutive iterations is used as the termination point. If this difference is less than a unit, the program comes to a stop.

## 4. RESULTS

SRS was applied to the prepared Indiana pine dataset with two criteria to extract the most informative and the most separable spectral regions. Using the Indiana pine reference map, 13 main classes were observed (three classes with less than 50 pixels were not considered).The most discriminant spectral regions have the best separability for the given classes on basis of TD. To identify the most informative spectral regions, it does not need to know the classes, and the entire image can be considered. However, we used almost the same areas that cover all the classes.
SRS terminated after 18 iterations to identify the most informative spectral regions; while some of the splits had been removed in step 3.a. of the algorithm. So, the final dataset has seven bands. On the other hand, SRS needed just 7 iterations to identify the optimal discriminant spectral regions for the given classes; the final image has 7 bands (the last split is removed, as it increased the TD distance less than a unit). Fig. 3 shows the results.

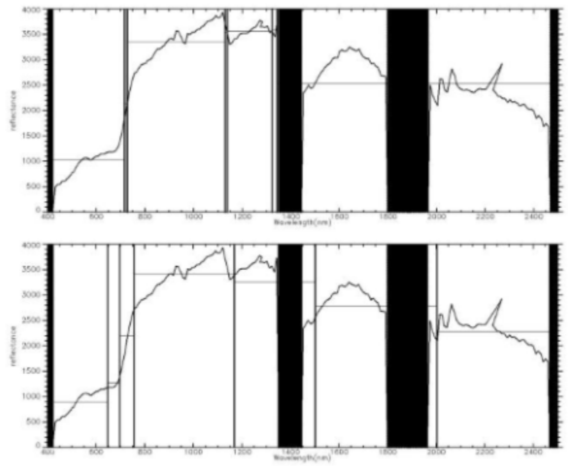

Fig. 3. Up: The most informative spectral regions. Down: the most separable spectral regions for the Indian Pine scene. The vertical lines show the location of splits over the spectrum, and noisy bands are showed by solid black areas. The mean-reflectance values of the new band sets are shown by the horizontal lines.

As it is seen the spectral regions are not the same. For instance, about 700nm wavelength, called red edge for vegetation areas, we have one 10nm band as an informative spectral region; while for separability two wider divisions (about 50nm or more) are selected; or more obvious at middle infrared region. Generally speaking, we could say that the most informative spectral regions and the most discriminant bands are not similar, and they do not have any relation to each other.

## 5. DISCUSSION

To answer the question concerning the similarity between spectral regions in terms of information content and class separability, SRS was applied over the scene which results in different band combinations.
Comparing the most informative spectral regions and the most discriminant ones as given by SRS, it was found that they are not the same. The reason is that the most informative bands are the bands which have a small correlation with the adjacent bands. As already mentioned, the informative spectral regions are based on information theory, which emphasizes that with less dependency in a set, the information content increases. The algorithm for identifying the most separable spectral regions selects the bands which lead to maximizing the distance between mean values of classes. It may select bands which have large correlation with other bands, contrary to the most informative bands selection algorithm. In fact, by having more correlated bands, separability may increase. We can prove this assertion by a simple example.

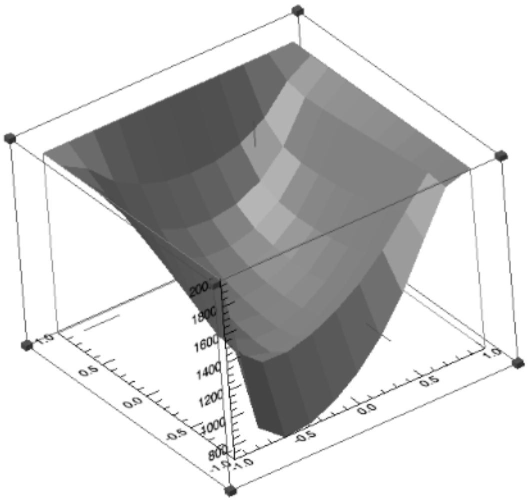


Fig 3. Variation in TD distance for two hypothetical classes with constant means and standard deviation and changeable correlation. *X* and *Y* axes are correlations. *Z* axis is the separability measure.

Let us assume that two hypothetical classes exist in a two-dimension scatter plot. If means and standard deviations of both classes in two dimensions are constant, and the correlation between two bands for each class is changing between -1 and 1, we can see that the maximum separability between two classes is obtained when the correlation of the bands in one or both classes is at the maximum value. The examined samples in different situations show that the maximum value of separability measures, including covariance matrix of the classes for given classes, is affected by two factors: the relative location of class's means, and the type of separability measure. However, in all cases the separability measures reach the maximum, when at least two bands are highly correlated in one class; sometimes negatively, and sometimes positively.

For example, Fig. 3 shows the maximum transformed divergence distance for two hypothetical classes with constant means and standard deviations when the correlations between the bands are changing. The *X* and *Y* axes are the correlation between bands for each class, and *Z* is the separability measure. In this example, when one of the classes has the maximum positive correlation (+1), the discriminant function reaches the maximum value.

To be more certain, the correlation matrix for the most separable bands shown in Fig. 2 (down) was computed. Apparently, some of the selected bands are highly correlated; even, some adjacent bands like bands 1&2, 5&6, or 6&7. Table 1 reveals the correlation matrix of the most discriminant bands for the given classes.

Table 1. Correlation matrix for the most discriminant

| ***Band*** | ***1*** | ***2*** | ***3*** | ***4*** | ***5*** | ***6*** | ***7*** |
|---|---|---|---|---|---|---|---|
| ***1*** | 1.00 | 0.99 | 0.27 | -0.42 | 0.83 | 0.92 | 0.93 |
| ***2*** | 0.99 | 1.00 | 0.23 | -0.45 | 0.85 | 0.95 | 0.96 |
| ***3*** | 0.27 | 0.23 | 1.00 | 0.73 | 0.48 | 0.16 | 0.11 |
| ***4*** | -0.42 | -0.45 | 0.73 | 1.00 | -0.07 | -0.45 | -0.51 |
| ***5*** | 0.83 | 0.85 | 0.48 | -0.07 | 1.00 | 0.91 | 0.86 |
| ***6*** | 0.92 | 0.95 | 0.16 | -0.45 | 0.91 | 1.00 | 0.99 |
| ***7*** | 0.93 | 0.96 | 0.11 | -0.51 | 0.86 | 0.99 | 1.00 |

## 7. CONCLUSION

In this paper, information content and class separability for a hyperspectral scene were investigated. These two concepts usually use in band selection algorithms to achieve the best image classification. The results obtained by newly developed method, SRS, prove that the most informative spectral regions and the most discriminant regions are not the same. The research shed light on the cause of the difference which emphasizes that informative bands have low correlation among each other; while highly correlated bands give the best separability between classes.

## 8. REFERENCES


[1] Y. Jihao, W. Yisong, and Z. Zhanjie, "Optimal Band Selection for Hyperspectral Image Classification Based on Inter-Class Separability." *Proc. Symp.Photon. Optoelectron.,* pp.1-4, 2010.

[2] X. Li, and J. Liu, "An adaptive band selection algorithm for dimension reduction of hyperspectralimages.", *Proc. Conf.Image Analysis and Signal Processing,* pp. 114-118, 2009.

[3] S. Kumar, J. Ghosh, and M. M. Crawford, "Best-bases feature extraction algorithms for classification of hyperspectral data," *Ieee Transactions on Geoscience and Remote Sensing,* vol. 39, no. 7, pp. 1368-1379, Jul, 2001.

[4] C. I. Chang, Q. Du, T. L. Sun *et al.*, "A joint band prioritization and band-decorrelation approach to band selection for hyperspectral image classification," *Ieee Transactions on Geoscience and Remote Sensing,* vol. 37, no. 6, pp. 2631-2641, Nov, 1999.

[5] A. Ifarraguerri, and M. W. Prairie, "Visual Method for Spectral Band Selection," *Ieee Geoscience and Remote Sensing Letters,* vol. 1, no. 2, pp. 101-106, Apr, 2004.

[6] T. M. Lillesand, R. W. Kiefer, and J. W. Chipman, *Remote sensing and image interpretation*, 6th ed., Hoboken, NJ: John Wiley & Sons, 2008.

[7] C. E. Shannon, "A Mathematical Theory of Communication," *Bell System Technical Journal,* vol. 27, no. 3, pp. 379-423; no. 4, pp.623-656, 1948.

[8] R. A. Schowengerdt, *Remote sensing, models, and methods for image processing*, 2nd ed., San Diego: Academic Press, 1997.