IOP Conf. Series: Earth and Environmental Science **17** (2014) 012064 doi:10.1088/1755-1315/17/1/012064

Assessing reliability of classification in the most informative spectral regions of hyperspectral images

S E Hosseini Aria[1], M Menenti[2] and B G H Gorte[3]
Department of Geoscience and Remote Sensing, Delft University of Technology, Netherlands

[1]S.E.HosseiniAria@tudelft.nl, [2]M.Menenti@tudelft.nl, [3]B.G.H.Gorte@tudelft.nl

Abstract. Reliability analysis is usually applied to evaluate classification procedures with different classes. In this research, we have applied the analysis to two different band sets to find out which one is more reliable. These band sets provide the most informative spectral regions covered by hyperspectral images. The informative regions are identified by minimizing two dependency measures between bands: correlation coefficient and normalized mutual information. The implementations are done by a newly developed top-down method named Spectral Region Splitting (SRS) resulting in two sets of bands which are almost identical at critical spectral regions. A reliability analysis based on the thresholding technique of the two sets of bands was performed. A technique was applied to discard those pixels that are not correctly classified at the given confidence level. The results show that the informative spectral regions selected by normalized mutual information was more reliable.

1. Introduction

Although the high dimensionality of hyperspectral data increases the potential performance of image classification and pattern recognition, data redundancy is a big challenge leading to an increase in computational cost. To reduce the high dimensionality of hyperspectral images, selection of optimal bands is usually applied. There are several methods for decreasing the number of bands, which are categorized in various types, e.g.: feature extraction [6], feature/band selection [17], supervised [3, 11] and unsupervised [9] algorithms, top-down and bottom-up techniques [11] etc. Other criteria and models to tackle the high dimensionality problem include spectral reconstruction [20], statistical analysis of the reflectance [19], partial least squares [14], and mutual information [8]. Concerning applied techniques for band reduction, the number of bands and their spectral locations are different. This research investigates the reliability of classifications with various numbers of bands to show the quality of image classification outputs. Usually, reliability analysis is performed for classification procedures with different sets of classes. But in this paper, we use a single set of classes to analyse reliability of classification with two different band sets. These band sets are the optimal informative subdivisions of the part of the spectrum, that is covered by hyperspectral images. So, in the first step, the optimal set of bands that yields the most informative regions of the spectrum (between 400nm-2500nm) is formed by a newly developed method called Spectral Region Splitting (SRS). This method is a top-down recursive algorithm starting with a very wide band covering the spectrum, and splitting it into the interesting spectral regions. The method can handle different criteria to form spectral regions. The criteria we used in this paper are two dependency measures information theory [15, 16], which emphasizes that by having more independent data, information content is increasing. The dependence measures are 1) correlation coefficient (R) 2) normalized mutual information (NI). The

algorithm splits the spectrum into regions. When increasing the number of regions, the dependency between them usually gets higher. The algorithm continues splitting until it reaches a termination point which gives the optimal informative regions. The SRS method has been evaluated on a well-known 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 (on the order of 10 nm) contiguous bands [13]. The scene includes agriculture, forest or other perennial vegetation, and some man-made objects. Noisy bands were removed beforehand. The implementation of SRS on Indiana Pine using the two dependency measures *R* and *NI* gives two sets of optimal informative spectral regions; these are very similar in critical parts of the spectrum, but with a different number of regions.

In the second step, we assess the reliability of classifications applied to the two datasets derived by SRS. For measuring the reliability a *thresholding* technique is applied, which counts those pixels that are not classified at the given confidence level. The percentage of rejected pixels at a given confidence level is taken as a measure of classification reliability. So, the two band sets, selected based on *R* and *NI*, are fed into the thresholding method to identify which one gives more reliable classifications. More details about two dependency measures are presented in section 6. Section 2 gives some information of the dataset and Section 3 provides a full description of the spectral region splitting (SRS method). Reliability analysis is described in Section 4. Section 5 shows the results of SRS implementation and reliability analysis. The conclusion is presented in section 7.

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 nanometres. Two-thirds of the Indian Pines scene is covered by agricultural crops, and one-third by forest and other natural perennial vegetation. There are two major dual lane highways, a railway, as well as some low density housing, other built-up structures, and smaller roads. Since the scene is taken in June some of the crops present, corn, soybeans, are in early stages of growth with less than 5% coverage. The ground truth available is designated into sixteen classes. Since the classes almost cover three fourth of the whole scene, we use these parts of the scene to identify the most informative spectral regions. We first reduced 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 Univeristy website: https://engineering.purdue.edu/~biehl/MultiSpec/hyp-erspectral.html.

3. Most informative spectral regions

This section, describes identifying the most informative spectral regions by the SRS method. It is a top-down algorithm; it starts with a single very wide spectral band covering the complete measured spectral range (without the bands that were removed). The signal in this band consists of the average of all spectral bands. The wide band is then split into two parts, where the location of the split is determined by the criterion introduced to the program. So the split scans all of the situations and finds the best dividing location concerning the criterion. The algorithm continues recursively and each time gives finer divisions until it reaches a termination point.

According to information theory we should find spectral regions which have small dependency between each other to obtain the informative regions of the spectrum. Two criteria are used to measure dependency. The correlation coefficient is the first one which indicates the strength of a linear relationship between two variables with a random distribution. The correlation coefficient (R_{ij}) of bands *i* and *j* is:

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

where, x_{ip} and x_{jp} are the *p*th pixel value (radiance) of bands *i* and *j*, μ_i and μ_j represent the mean radiance in the bands *i*, *j*. In the case of comparison and finding minimum dependency, our case, the absolute value of *R* is used.

The second dependence measure is normalized mutual information (*NI*). Mutual information *(I)* is a quantitative measurement on two random variables that can be thought of as the reduction in uncertainty about one random variable when given knowledge about the other. Less mutual information between two random variables indicates smaller reduction of uncertainty, so the minimum value of mutual information, zero, means the variables are totally independent. Since *I* can get any positive value and does not have a maximum, we use *NI* which is always between zero and one, which allows for comparison. Its formula is as follows:

$$NI(X_i, X_j) = \frac{2 \cdot I(X_i, X_j)}{H(X_i)+H(X_j)} \tag{2}$$

where X_i and X_j are bands number *i* and *j*. *I* is the mutual information for the two bands and *H* is the entropy, i.e. the average amount of information for a band. Both *I* and *H* are functions of probability distributions:

$$I(X_i, X_j) = \sum_{x_i \in X_i} \sum_{x_j \in X_j} p(x_i, x_j) log_2 \frac{p(x_i, x_j)}{p(x_i)p(x_j)} \tag{3}$$

$$H(X_i) = \sum_{x_i \in X_i} P(x_i)\, log_2 \frac{1}{P(x_i)} \tag{4}$$

where, $p(x_i, x_j)$ is the joint probability distribution function of X_i and X_j, $p(x_i)$ and $p(x_j)$ are the marginal probability distribution function of X_i and X_j. More details are in [2].

For extracting the most informative spectral regions, these two dependence measures have been used in the SRS method. Therefore, the wide band covering all the spectrum has been split to maximize spectral information with the minimum number of bands. The optimal location of the split is where the maximum information is obtained. When the wide band is split into two parts, the information on spectral detail will improve; especially, if the new bands have little dependency, resulting in high information content. Finally, we have two band sets, one for each dependency measure.

As SRS is an iterative technique to obtain informative spectral regions, there should be a termination point, where it reaches the optimal informative sampling of the spectrum. This point is identified based on the dependency of the produced bands. After a very wide band, made up of all the measured spectrum, has been split into two parts, the algorithm scans all the situations again to find the location of a new split that gives minimum dependency, and then three divisions are obtained. By continuing this procedure, increasingly refined spectral divisions will be obtained. In each iteration the minimum dependency, which is the location of the split, is usually higher than in the previous one. Eventually, the algorithm reaches the point where it can only create two highly dependent spectral regions, and this point is the termination point. determining the threshold to identify high spectral dependency, it is assumed that the distribution of dependency measures over all combinations is normal. Then, the correlation coefficient matrix and the normalized mutual information matrix, involving all dependencies between bands, are computed. In figure 1. the bright parts indicate higher dependency, which mostly occurs between neighbouring bands.

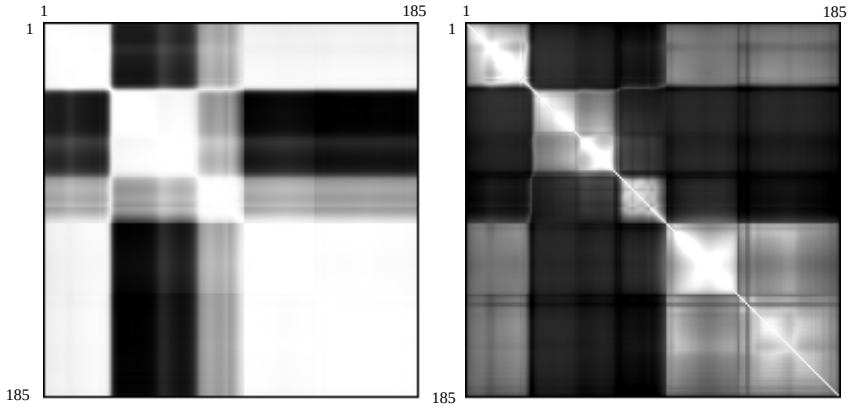

Figure 1. Correlation matric (right) and normalized mutual information matrix (left) of bands.

When we consider 25% of all combinations in the dependency matrices as highly dependent, this corresponds to values larger than 0.879 for the *R* matrix and 0.406 for the *NI* matrix. These value are calculated from the histograms of the two matrices.

As the dependency measures are computed for pairs of bands, the algorithm sometimes produces several narrow bands next to each other, which may be highly dependent. Therefore a post processing step is done, during which the dependencies between adjacent regions are computed and highly dependent adjacent regions are merged. For merging, the same threshold for the termination point is used as well. Section 5 gives the results.

4. Reliability analysis

The error rate and the reject rate are commonly used to describe the performance level of classification systems. The option to reject is introduced to safeguard against excessive misclassification; i.e. it converts potential misclassification into rejection [1]. Based on the Bayes decision rule in pattern classification, a pattern at a pixel *x* is assigned to a class ω_i if the posterior probability of the pixel for one class is more than its posterior probability for other classes [4]. For example, by having just two classes, ω_1 and ω_2, the decision rule says:

$$\text{Decide } \omega_1 \text{ if } P(\omega_1 |x) > P(\omega_2 |x) \text{ ; otherwise decide } \omega_2 \tag{5}$$

where $P(\omega_i |x)$ is a posteriori probability of a pixel *x* belonging to class ω_i. This decision minimizes the probability of error. In this scheme all pixels are classified, whereas there may be pixels not belonging in any class, and classification reliability decreases. To improve classification reliability, a threshold is introduced to reject a pixel if the maximum of the posteriori probabilities of a pixel is less than the threshold: the pixel is rejected and considered as unclassified. The advantage of rejecting a sample instead of running the risk of misclassifying is to boost classification reliability. Larger threshold values will increase the number of unclassified pixels, named rejection class. The threshold can be based on a confidence level. The percentage of rejected pixels at a given confidence level is a measure of classification reliability known as “thresholding” technique [5]. A plot of the percentage of rejected image pixels versus confidence level can be made by using different threshold values. It is shown in [5] that a low reliability classification procedure has a rapid drop in the plot when the confidence level is increasing, since most pixel are rejected at low confidence level. On the opposite side, the plot is going down gradually by increasing the confidence level for a classification procedure with high reliability.

Many classifiers are based on a pixel-to-class proximity measure in an n-dimensional space. An example is the Mahalanobis distance $D_{x,i}^2$:

$$D_{x,i}^2 = (x - \mu_i)V_i^{-1}(x - \mu_i) \quad (6)$$

where *x* represents a pixel in a *n*-dimensional space (*n* is the number of bands), μ_i and V_i are the mean and the covariance matrix of class *i*. When assuming that the attribute values in each class are normally distributed a small value of $D_{x,i}^2$ corresponds to a high posteriori probability of the given pixel. When we use Mahalanobis distance in the thresholding method rejected pixels at a given confidence level are obtained if the $D_{x,i}^2/2$ is larger than the corresponding threshold [5]. Swain [18] remarks that when data are normally distributed the frequency of Mahalanobis distances approaches a chi-square distribution with *n* degrees of freedom (dimensionality of data). Under this assumption, each threshold value corresponds to a confidence level which can be determined from a chi-square table.

5. Implementation and result

SRS was applied to the prepared Indiana pine dataset to extract the most informative spectral regions. It terminated after 18 iterations with correlation coefficient as the dependency criterion, and after 45 iterations with normalized mutual information. As mentioned, a post processing step was done to eliminate possible highly dependent adjacent bands reducing the number of bands to 7 and 14 respectively. Figure 2 shows the results.

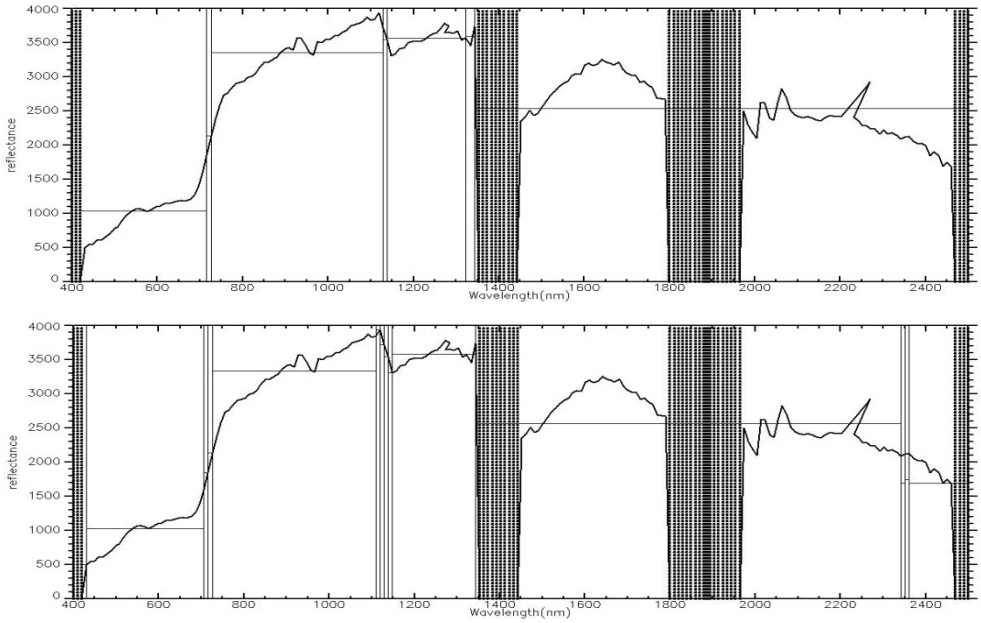

Figure 2. The most informative spectral regions Up: based on correlation coefficient. Down: based on normalized mutual information. The vertical lines illustrate the location of splits over the spectrum, and noisy bands are in solid dotted areas. The mean-reflectance values of the new band set are shown by the horizontal lines.

The sets regions obtained are almost the same, although SRS based on *NI* selects more narrow bands. For instance, around 700nm wavelength, called red edge for vegetated areas, we have one 10nm band based on *R*; whereas the *NI* algorithm has two narrow bands. The difference may arise

because correlation coefficient only takes linear relations into account only, whereas normalized mutual information also considers non-linear relations between two bands.

The area that was used to identify the most informative spectral region includes 16 classes based on the Indiana pine reference map. For reliability analysis 13 of them were selected (three classes consisting of less than 50 pixels are not considered). Based on the number of bands, critical values were extracted from chi-square distribution table corresponding to the confidence levels which are shown in table 1. In each band set for each pixel, the Mahalanobis distance was computed. Then, half of the minimum distance was compared with the critical values shown in table 1. The resulting curve for each dataset is illustrated in figure 3.

Table 1. Critical values for different confidence levels with 7 and 14 degrees of freedom.

Confidence level	99%	91%	74%	55%	41%	19%	9%	3%	0.01%
df = 7	1.24	2.72	4.34	5.92	7.18	9.97	12.38	15.51	29.88
df = 14	4.66	7.58	10.30	12.70	14.54	18.38	21.48	25.49	42.58

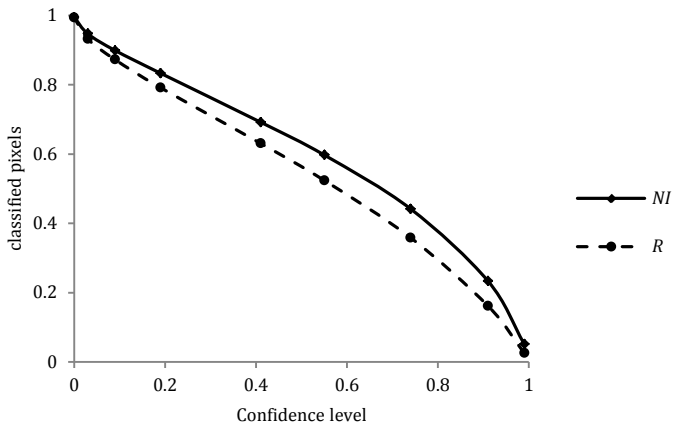

Figure 3. Reliability curve for two band sets obtained from normalized mutual information (*NI*) and correlation coefficient (*R*).

As expected, the percentage of classified pixels decreases when increasing the confidence level. Figure 3 shows that the reliability of the band set selected by normalized mutual information is higher than the one with correlation coefficient. It is not just because of the number of bands, but also it may relate to the characteristic of *NI* which considers non-linearity dependency between bands. More discussion about the difference between *NI* and *R* is in the next section.

6. Evaluating the dependency measures

In the SRS method, measuring the dependency between two bands is the fundamental mechanism for identifying the location of splits over the spectrum between informative regions. Two dependence measures are used: correlation coefficient and normalized mutual information, which result in two different sets of spectral bands. Useful bands are selected by both algorithms, like the red edge; but some spectral bands are not identical.

The correlation coefficient measures the linear dependency of two bands as the quotient of their covariance and the product of their standard deviations. Two independent random variables are surely uncorrelated; however, two uncorrelated random variables are not necessarily independent. Normalized mutual information takes nonlinear dependency into account. This dependency measure is

symmetric and lies between zero and one. This property is useful when comparing the dependency among variables.

Noise usually has an influence on dependency between bands. Experience over different hyperspectral scenes shows that noisy bands have more influence on normalized mutual information than on correlation coefficient. i.e. *NI* decreases significantly when one or two bands are contaminated by noise. For example, two noise-free adjacent bands were selected from the dataset (bands 50 and 51) which are highly dependent. Their correlation is 0.999 and the *NI* is 0.847. Then some noise was added to band 50. The result was 0.984 for correlation and 0.528 for *NI*, which shows that the given noise decreased the correlation less than 2%, while the *NI* reduced more than 35%. Therefore, noisy channels are more selected by when using normalized mutual information and they were not even merged at the post-processing step.

Correlation coefficient and normalized mutual information have similar behaviour when the marginal and joint probability distribution of bands are normal. The normal distribution is a linear distribution, in the sense that the linear correlation coefficient captures the overall dependence of two random variables [10]. It is proven if the joint probability distribution is Gaussian, the correlation and mutual information are analytically related to each other [7, 12].

It is well understood that normalized mutual information measures general dependency, while the correlation measures linear dependency. Therefore we expect that spectral partitions defined by *NI* are more accurate than the ones based on correlation, although *NI* is more sensitive to noisy bands. As the final result, it could be said that *NI* is a better measure of spectral dependency than the correlation coefficient and also it leads to higher reliability of classification.

7. Conclusion

By using a newly developed method, spectral region splitting, optimal informative spectral regions can be obtained from a hyper-spectral image. The criteria that were used to obtain informative spectral regions are correlation coefficient and normalized mutual information, which should be minimized to reduce the dependency between spectral divisions and extract informative spectral regions. The results over Indian pine dataset show that both algorithms selected similar spectral regions as the most informative ones, which *NI* algorithm has more bands. Reliability analysis was done for two produced band sets based on the thresholding method. By increasing the confidence level, the number of rejected pixels are going up. The reliability graphs show that at each level the correlation coefficient criterion gives more rejected pixels than the *NI* criterion. Therefore, the band set selected by *NI* algorithm is more reliable than the other set chosen by *R* algorithm. It could be because of higher number of bands and behaviour of normalized mutual information which calculates linear and non-linear dependency between bands.

8. References

[1] Chow C 1970 On optimum recognition error and reject tradeoff *Ieee T Inform Theory* **16** 41-6
[2] Cover T M and Thomas J A 1991 *Elements of information theory* (New York: Wiley)
[3] Du Z, Jeong M K and Kong S G 2007 Band selection of hyperspectral images for automatic detection of poultry skin tumors *Ieee T Autom Sci Eng* **4** 332-9
[4] Duda R O and Hart P E 1973 *Pattern classification and scene analysis* (New York,: Wiley)
[5] DUrso G and Menenti M 1996 Performance indicators for the statistical evaluation of digital image classifications *ISPRS Journal of Photogrammetry and Remote Sensing* **51** 78-90
[6] Fauvel M, Chanussot J and Benediktsson J A 2009 Kernel Principal Component Analysis for the Classification of Hyperspectral Remote Sensing Data over Urban Areas *Eurasip J Adv Sig Pr*
[7] Fraser A M 1989 Reconstructing Attractors from Scalar Time-Series - a Comparison of Singular System and Redundancy Criteria *Physica D* **34** 391-404

[8] Guo B F, Gunn S R, Damper R I and Nelson J D B 2006 Band selection for hyperspectral image classification using mutual information *Ieee Geosci Remote S* **3** 522-6
[9] Jia S, Ji Z, Qian Y T and Shen L L 2012 Unsupervised Band Selection for Hyperspectral Imagery Classification Without Manual Band Removal *Ieee J-Stars* **5** 531-43
[10] Kraskov A, Stögbauer H and Grassberger P 2004 Estimating mutual information *Phys Rev E* **69** 066138
[11] Kumar S, Ghosh J and Crawford M M 2001 Best-bases feature extraction algorithms for classification of hyperspectral data *Ieee T Geosci Remote* **39** 1368-79
[12] Li W T 1990 Mutual Information Functions Versus Correlation-Functions *J Stat Phys* **60** 823-37
[13] Lillesand T M, Kiefer R W and Chipman J W 2008 *Remote sensing and image interpretation* (Hoboken, NJ: John Wiley & Sons)
[14] Lorente D, Aleixos N, Gomez-Sanchis J, Cubero S, Garcia-Navarrete O L and Blasco J 2012 Recent advances and applications of hyperspectral imaging for fruit and vegetable quality assessment *Food Bioprocess Tech* **5** 1121-42
[15] Shannon C E 1948 A Mathematical theory of communication *At&T Tech J* **27** 379-423
[16] Shannon C E 1948 A Mathematical theory of communication *At&T Tech J* **27** 623-56
[17] Sotoca J M, Pla F and Sanchez J S 2007 Band selection in multispectral images by minimization of dependent information *Ieee T Syst Man Cy C* **37** 258-67
[18] Swain P H and Davis S M 1978 *Remote Sensing: The Quantitative Approach* (New York, NY: McGraw-Hill)
[19] Taylor S, Kumar L, Reid N and Lewis C R G 2012 Optimal band selection from hyperspectral data for Lantana camara discrimination *Int J Remote Sens* **33** 5418-37
[20] Wang C, Menenti M, Stoll M P, Belluco E and Marani M 2007 Mapping mixed vegetation communities in salt marshes using airborne spectral data *Remote Sens Environ* **107** 559-70