Classification of Hyperspectral Data for Land Cover Mapping: Is There Any Significant Improvement?

Alvin Lau Meng Shin and Mazlan Hashim

Department of Remote Sensing
Faculty of Geoinformation Science & Engineering
Universiti Teknologi Malaysia, UTM Skudai,81310 Johor, Malaysia
{alvinlau,mazlan}@fksg.utm.my

Abstract

This paper highlights the results of classification of an airborne MASTER hyperspectral data for land cover mapping in Redang Island, Malaysia. Two addressed issues in the study are : (1) whether or not hyperspectral would increase classification accuracy over medium spatial resolution (10m) of MASTER data for land cover mapping, and (2) radiometric normalization still required in hyperspectral data Three classification algorithms examined in this study, namely Binary encoding, Spectral Angle Mapper and Linear spectral Unmixing. The topographic-effect normalization was applied to the test site prior data classification. Results of study indicated that Linear Spectral Unmixing classification technique gives the best overall classification accuracy of the hyperspectral data for land cover in the study area. The result of this study also clearly indicated that hyperspectral data could not improve classification accuracy significantly especially when the .mixed pixels are abundant

Key words: Hyperspectral Remote Sensing, Binary Encoding, Spectral Angle Mapper, Linear Spectral Unmixing, Topographic Normalization

1. Introduction

Hyperspectral data is one of the recent development in remote sensing technology, an advancement in sensor development that have allowed voluminous of data be collected in 50 or more channels over the same target area. Parallel with this development, emphasis on data analysis dedicated for hyperspectral data has also been stressed. There are at least two broad categories of how specific information can be extracted from hyperspectral data, either by feature extraction technique or by hyperspectral data classification approach. In the context of the feature extraction, the information were determined using the basis of specific pattern within the signatures of the target-of-interest, hence, allow features be extracted once the appropriate spectral bands were identified. In the latter, all the spectral data were subjected to classification algorithms where the main task is to whether or not the spectral classes identified within hyperspectral data can be transformed to the corresponding labels in the real world. Both of these two processing techniques have shortcomings as well as advantages, but comprehensive analysis for utilization of such data and techniques for land cover mapping have not so far being reported. This leads to the objective of this study to map land cover of study area using (1) Binary encoding, (2) Spectral Angle Mapper, and (3) Linear spectral Unmixing methods.

2. Materials and Methods

2.1 Study Area

The study area is located in Redang Island (lat 5° 44′ -5° 50′ N and long 102° 5′-102° 59′ E) as shown in Figure 1. About 75 percent of the study area is a high

Meng Shin and Hashim/Geoinformation Science Journal, Vol. 3, No. 1, pp. 74-79

Classification of Hyperspectral Data for Land Cover Mapping: Is There Any Significant Improvement?

Alvin Lau Meng Shin and Mazlan Hashim

Department of Remote Sensing
Faculty of Geoinformation Science & Engineering
Universiti Teknologi Malaysia, UTM Skudai,81310 Johor, Malaysia
{alvinlau,mazlan}@fksg.utm.my

Abstract

This paper highlights the results of classification of an airborne MASTER hyperspectral data for land cover mapping in Redang Island, Malaysia. Two addressed issues in the study are : (1) whether or not hyperspectral would increase classification accuracy over medium spatial resolution (10m) of MASTER data for land cover mapping, and (2) radiometric normalization still required in hyperspectral data Three classification algorithms examined in this study, namely Binary encoding, Spectral Angle Mapper and Linear spectral Unmixing. The topographic-effect normalization was applied to the test site prior data classification. Results of study indicated that Linear Spectral Unmixing classification technique gives the best overall classification accuracy of the hyperspectral data for land cover in the study area. The result of this study also clearly indicated that hyperspectral data could not improve classification accuracy significantly especially when the .mixed pixels are abundant

Key words: Hyperspectral Remote Sensing, Binary Encoding, Spectral Angle Mapper, Linear Spectral Unmixing, Topographic Normalization

1. Introduction

Hyperspectral data is one of the recent development in remote sensing technology, an advancement in sensor development that have allowed voluminous of data be collected in 50 or more channels over the same target area. Parallel with this development, emphasis on data analysis dedicated for hyperspectral data has also been stressed. There are at least two broad categories of how specific information can be extracted from hyperspectral data, either by feature extraction technique or by hyperspectral data classification approach. In the context of the feature extraction, the information were determined using the basis of specific pattern within the signatures of the target-of-interest, hence, allow features be extracted once the appropriate spectral bands were identified. In the latter, all the spectral data were subjected to classification algorithms where the main task is to whether or not the spectral classes identified within hyperspectral data can be transformed to the corresponding labels in the real world. Both of these two processing techniques have shortcomings as well as advantages, but comprehensive analysis for utilization of such data and techniques for land cover mapping have not so far being reported. This leads to the objective of this study to map land cover of study area using (1) Binary encoding, (2) Spectral Angle Mapper, and (3) Linear spectral Unmixing methods.

2. Materials and Methods

2.1 Study Area

The study area is located in Redang Island (lat 5° 44′ -5° 50′ N and long 102° 5′-102° 59′ E) as shown in Figure 1. About 75 percent of the study area is a high land area, situated in the center of the island. The land cover classes, among others include the primary forest, scrub forest, grassland, rubber, coconut trees, wetland and sand area.

For the purpose of this study, all these land cover classes were generalized into five major classes which are: primary forest, scrub forest, wetland, rubber and sand area. The categorization considered the relevancy of the spatial resolution sensor of sensor where data are acquired.

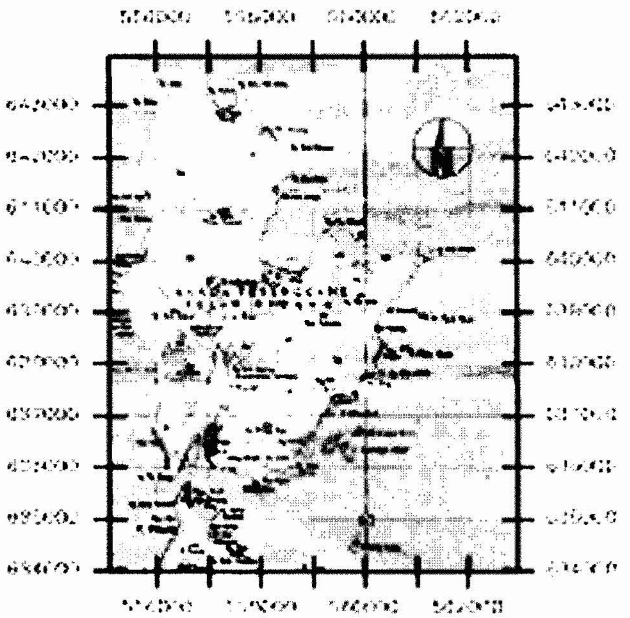

Figure 1: The study area- Redang Island

2.2 Satellite and Ancillary Data

The MASTER (MODIS/ASTER) airborne data was used in this study. The data were acquired under joint project of JPL-NASA Pacrim II campaign of the AIRSAR/TOPSAR mission in the South East Asian Countries and Australia. Table 1 tabulates the specification of data.

The ancillary data of the study area, namely the topographic map (1:50,000) and the land cover map were used as base for geometric correction of data and ground truth for classification of data. Both the ancillary data were published by Malaysian Dept of Surveying & Mapping, and Dept of Agriculture, respectively.

Table 1: Specifications of MASTER data

Date of Acquisition	19 September 2000
Location	Pulau Redang, Malaysia
Flight Number	00-020-29
Track Number	06
Aircraft Heading	269 °
Solar Zenith	42.4°
GPS Altitude	8405 m (MSL)
Wavelength range	0.4-13 micrometer
Number of channels	50
Number of pixels	716
Instantaneous Field of View	2.5 mradians
Total Field of View	82.92°
Pixel Size	10 m
Product	Radiance at sensor (level 1B)
Data Format	HDF
Digitization	16 bit

2.3 Digital Image Processing

The extraction process of land cover information is divided into three phases: (1) Data Pre-Processing, (2) Data Classification and (3) Accuracy Assessment. Figure 2 shows the flowchart of the processing activities undertaken in these three processing phases.

2.3.1 Data Pre-processing

The data preprocessing involved the followings: (a) Conversion of digital number (DN) to Reflectance, (b) Geometric Correction, and (c) Removal of Topographic Effect.

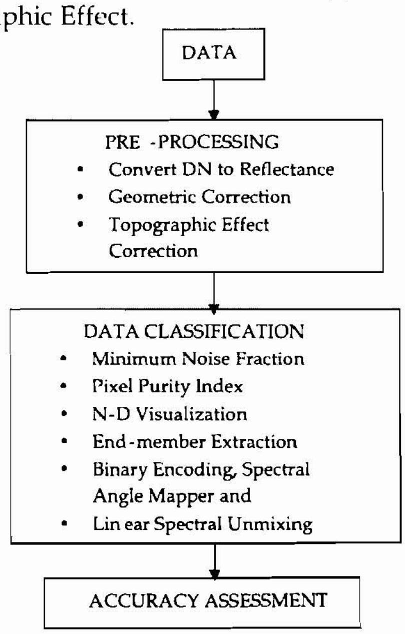

Figure 2: Methodology employed in data processing.

(a) Convert DN to Reflectance

The conversion of the DN of the MASTER data was carried out using Internal Average Relative Reflectance (IAR) technique. The output of this conversion is the relative reflectance and this can be achieved by dividing each pixel spectrum by the overall average spectrum.

(b) Geometric Correction

The image-to-map geometry rectification was adopted in the image geometric correction. A total number of 20 ground control points were established, the geometric transformation of data were carried out using a second order polynomial function. The RMSE (root mean square) error achieved is 0.4 pixel. The nearest neighbour resampling scheme was employed in the interpolation of the radiometric values of raw data into the transformed geometry to ensure the original reflectance variation values did not change after performing the image rectification.

(c) Removal of Topographic Effect

The Lambertian Reflection model was adopted in minimising the topographic effect. The main parameters used in the model are Digital Elevation Model (DEM) of the study area, the solar azimuth and elevation angle at each individual pixels where the data were acquired. The output is the topographic-normalized image.

2.3.2 Data Transformation (Preparation)

Four steps procedure carried out in the classification of data. The procedures are as follows: (i) Minimum Noise Fraction (MNF) transformation, (ii) Derivation of Pixel Purity Index, and (iii) n-D Visualization, and end-members Extraction.

(i) Minimum Noise Fraction (MNF) transform

The MNF Transform is applied to further segregate noises found in the data. With MNF, the MASTER data space was divided into two parts: (1) Those with large eigen values and coherent eigen images, and (2) those with near-unity eigen values and noise-dominated images. Both the eigen values and the MNF images are used to evaluate the dimensionality of the data. The eigen values for bands that contain information will be in an order of magnitude larger than those that contain only noise, while the corresponding eigen images will be spatially coherent. The noise images, not contain any spatial information, hence have low order magnitude of eigen values. Further explanation of MNF noise removal function can be found in Board and Kruse (1994).

When MNF transform was completed, an eigen value plot was calculated and 25 MNF-transformed bands were displayed. From the MNF transform, only the first seven transformed bands are found to be adequate to hold original information content of all bands.

(ii) Pixel Purity Index (PPI)

Within the feature space, the noise-free data occurs as a continuous class from the purest to variety of mixtures. The purest pixels within feature space is referred to as end-member, while mixtures are sub-compositions due to mixels, attributed from low to medium resolution of the pixel recorded.

The PPI is a "counting system" to which the number of times each pixel within the scene to be classified are designed as purest. Only MNF-transformed Band 1 and Band 2 were selected as input when running PPI. This is because other MNF-transformed bands have lower eigen values, only have little information and contains lot of noise which will decrease the accuracy of the classification result.

(iii) N-D Visualization and Endmembers Extraction

The MASTER data (or spectra) can be thought of as points in an n-dimensional scatterplot, where n is the number of bands. The data for a given pixel corresponds to a spectral reflectance for that given pixel. The distribution of the MASTER data in n-space was used to estimate the number of spectral end-member's and their pure spectral signatures and to help understand the spectral characteristics of the materials which make up that signature. The image generated from PPI was used as the input in this session. The Spectral library for land cover of study area was created with the n-D Visualization.
Different classes generated from n-Dimensional Visualizer were compared to spectral library to identify each class. After the spectra collected from n-Dimensional Visualizer were identified, the spectra for each class were saved into a new spectral library file which will be used in classification.

2.3.3 Data Classification

The data classification were performed using three different classifiers, namely binary encoding, Spectral Angle Mapper (SAM), and Linear Spectral Unmixing.

(i) Classification using Binary Encoding

In binary encoding classification technique, data and end-member spectra were encoded and into bit map (0 or 1) based on whether a band falls below or above the spectrum mean. An exclusive function is used to compare each encoded reference spectrum with the encoded data spectra and a classification image is produced. Each pixel is classified to the end-member with the greatest number of bands that matched the classes.

(ii) Classification using Spectral Angle Mapper (SAM)

The Spectral Angle Mapper (SAM) is an automated method for comparing image spectra to individual spectra or a spectral library. The SAM assumes the data to be classified have been reduced to apparent reflectance. The algorithm determines the similarity between two spectra by calculating the "spectral angle" between them, treating them as vectors in a spectral space with dimensionality equal to the number of bands. A simplified explanation of this can be given by considering a reference spectrum and an unknown spectrum from two-band data. The two different materials will be represented in the 2-D scatter plot by a point for each given illumination, or as a line (vector) for all possible illuminations. Please refer to Kruse et al.(1993) for detailed explanation of SAM.

(iii) Classification using Linear Spectral Unmixing

The set of spectrally unique surface materials existing within a scene are often referred to as the spectral end-members for that scene. Linear Spectral Unmixing exploits the theory that the reflectance spectrum of any pixel is the result of linear combination of the spectra of all end-members inside that pixel. A linear combination in this context can be thought of as a weighted average, where each end-member's weight is directly proportional to the area of the pixel containing that end-member. If the spectra of all end-members in the scene are known, then their abundances within each pixel can be calculated from each pixel's spectrum (Peg, 2002).

3. Results and Discussion

The normalized image for topographic effect is shown in Figure 3. The corrected image shows a consistency of the reflectance values surrounding hilly areas, this is evident at slopes fringing the shorelines.The results of classification MASTER for both original reflectance and normalized images are shown in Figure 4. Both the classification results from the original reflectance image and topographic effect corrected image were compared and analyzed. Digitized land cover of the corresponding area was used as ground truth of the classifications. The Kappa statistic was employed in the accuracy assessment.

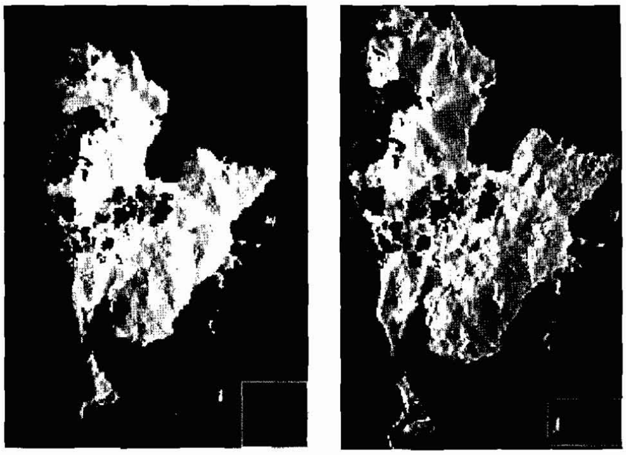

Figure 3: (a) Original reflectance image and (b) Image normalized for topographic effect

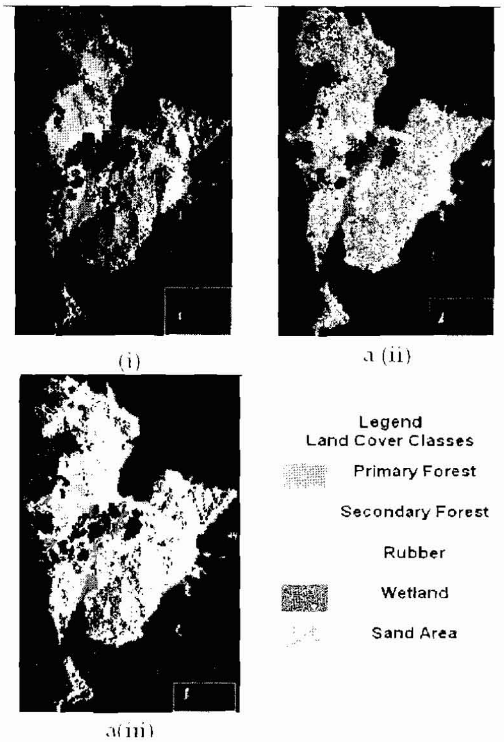

Figure 4(a) : Classification results of Original reflectance image using (i) Binary Encoding, (ii) SAM, (iii) Linear Spectral Unmixing.

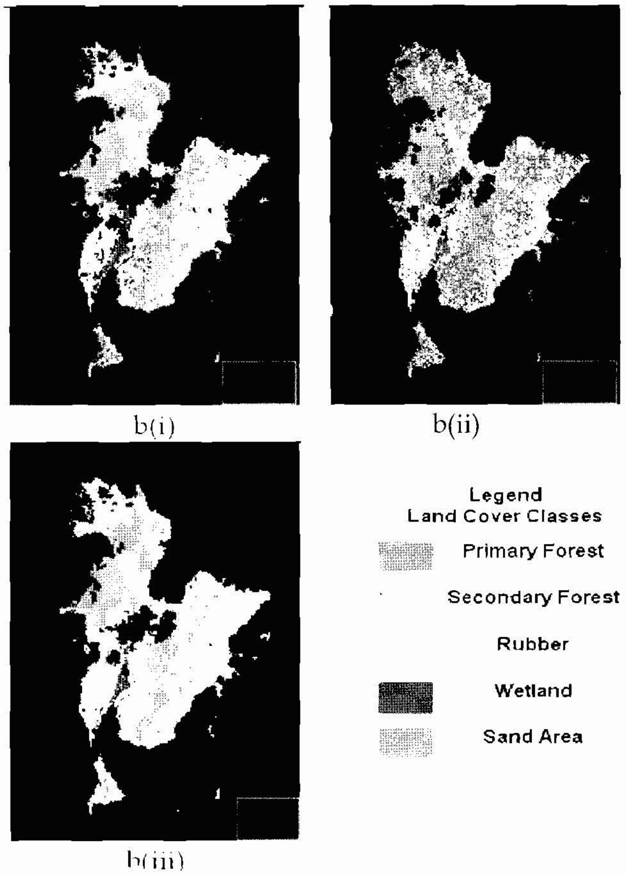

Figure 4: Classification results of Topographic corrected image using (i) Binary Encoding, (ii) SAM, (iii) Linear Spectral Unmixing.

The topographic normalized image using Linear Spectral Unmixing classification technique achieved the overall accuracy of 79% and Kappa coefficient of 0.74 while original reflectance image with the same classification technique recorded an overall accuracy of 72% and Kappa coefficient of 0.71. The classification of individual classes for both image set were tabulated in Table 2 and Table 3.

It is worth noting that Binary Encoding classified image for both classification sets (see Table 2) recorded the least producer accuracy compared to other classification techniques. After the topographic-effect normalization, the Linear Spectral Unmixing technique gives the best classification result followed by SAM classifier.

Table 2: Classification accuracy assessment for original image input

	Binary Encoding	
Class*	Producer Acc. (%)	User Acc. (%)
1	59.81	68.49
2	69.43	67.23
3	74.21	55.84
4	72.31	72.87
5	63.52	61.29

	SAM	
Class*	Producer Acc. (%)	User Acc. (%)
1	59.81	72.21
2	69.57	70.32
3	59.47	57.32
4	69.34	76.82
5	72.86	67.32

	Linear Spectral Unmixing	
Class*	Producer Acc. (%)	User Acc. (%)
1	72.83	71.18
2	74.11	72.31
3	52.38	62.12
4	82.43	78.39
5	68.25	71.19

*Note: 1= Primary Forest, 2=Secondary Forest, 3=Rubber, 4=Wetland, 5=Sand

Table 3: Classification accuracy assessment for topographic normalized image

	Binary Encoding	
Class*	Producer Acc. (%)	User Acc. (%)
1	70.89	67.42
2	71.43	75.24
3	68.34	76.21
4	74.91	77.81
5	68.73	63.27

	SAM	
Class*	Producer Acc. (%)	User Acc. (%)
1	79.48	74.07
2	69.88	75.68
3	63.15	76.88
4	84.96	79.27
5	76.64	73.34

	Linear Spectral Unmixing	
Class*	Producer Acc. (%)	User Acc. (%)
1	71.32	77.98
2	84.98	79.19
3	74.19	74.32
4	79.34	82.71
5	79.46	73.87

*Note: 1= Primary Forest, 2=Secondary Forest, 3=Rubber, 4=Wetland, 5=Sand

4. Conclusion

This paper has presented the results of classification of hysperspectral data (Master data) for land cover mapping. Three classifiers for hyperspectral data classification were examined. Apart from different classification approaches, this study also have addressed on whether or not the pre-processing of the radiometric defects could contribute in improving the classification of data. The results of the study clearly indicated that:

- The topographic normalized image gives a good performance in classification accuracy compared to original reflectance image, hence, it can be concluded that the topographic and slope effect will give the significant effect to the hyperspectral processing techniques. With the proper topographic normalization being applied to the hyperspectral data, the overall accuracy will increase significantly.

- The Linear Spectral Unmixing gives the best classification result in the study area, However, the superb spectral resolution of hyperspectral data with 50 channels does not contribute significantly to classification accuracy. With the nature of small and irregular land cover classes found in study area, the only best classification result be obtained by high spatial resolution data where the spatial variations of classes can be recorded.

Acknowledgments

The authors would like to acknowledge the Research Team of PACRIM II for providing MASTER data used in this study.

References

Boardman J. W., and Kruse, F. A. (1994), Automated spectral analysis: A geologic example using AVIRIS data, north Grapevine Mountains, Nevada: in Proceedings, Tenth Thematic Conference on Geologic Remote Sensing, Environmental Research Institute of Michigan, Ann Arbor, MI.

Cannon, M., Lehar A., and Preston F. (1983), Background Pattern Removal by Power Spectral Filtering, Applied Optics, Vol. 22, No. 6: p. 777-779.

Kruse, F. A., Lefkoff, A. B., Boardman, J. W., Heidebrecht, K. B., Shapiro, A. T., Barloon, J. P., and Goetz, A. F. H. (1993), The spectral image processing system (SIPS) - Interactive visualization and analysis of imaging spectrometer data, Remote Sensing of Environment, v. 44, 145 - 163.

Peg, S. (2002), Hyperspectral 101 Timely Tutorial, Geospatial Solution, February 2002.