LAND COVER CATEGORY DEFINITION BY IMAGE INVARIANTS FOR AUTOMATED CLASSIFICATION

Nguyen Dinh Duong
Environmental Remote Sensing Laboratory
Institute of Geography
Hoang Quoc Viet Rd., Cau Giay, Hanoi, Vietnam
duong.nd@hn.vnn.vn

Technical Commission TC VII-5

KEY WORDS: Automation, Classification, Remote Sensing, Land use/Land cover, Global change.

ABSTRACT

The definition of conventional land cover categories (legend) is usually formulated by statements which describe a land cover category by criteria derived from an interdisciplinary approach. Such a system of definition is suitable for conventional land cover mapping by ground observation or visual image interpretation when the interpreter combines image and auxiliary information to classify an object. Supervised image classification relies on statistical parameters of a class generated during training sampling which is in general nontransferable from one image to the other. Unsupervised classification with a clustering technique provides automated grouping, but there is no way to establish a fixed relation between a cluster code and a certain land cover category. Moreover the post-classification interpretation of results is time consuming and a subjective process that requires extensive ground truth data collection. Recently, research into automated classification of land cover was initiated by the author in the framework of the NASDA Research Announcement for the ADEOS-II satellite, which will carry a GLI sensor that has 6 spectral channels similar to those of LANDSAT TM. One of the issues of automated classification of GLI data is to develop a system for defining land cover in the image domain. This means a land cover category should be described by values derived from image data. The author has discovered several image invariants based on graphical analysis of the spectral reflectance curve of a pixel. The invariant found include modulation of the spectral reflectance curve, total reflected radiance index (TRRI) and spectral angles. These invariants seem to be quite stable for image data generated by the same sensor. In the paper the author reports on efforts to develop a digital definition of land cover using image invariants for automated classification. The proposed legend has been applied to a data set simulated for the future GLI sensor and the obtained classification result has proved the chosen approach to be correct.

1 INTRODUCTION

Land use and land cover change play a pivotal role in global environmental change. They contribute significantly to earth-atmosphere interactions and biodiversity loss, are major factors in sustainable development and human responses to global change, and are important for integrated modelling and assessment of environmental issues in general (IGBP Report No. 35 Land-Use and Land-Cover Change Science/Research Plan). This awareness has recently led to land use and land cover mapping activities at national, regional and global scales. One of the issues of land cover mapping is to standardise classification schemes from both technical and typological points of view. In conventional land cover mapping, each land cover unit is usually defined in the form of a statement that describes the land cover category by a set of attributes derived from an interdisciplinary approach (vegetation and soil sciences). Such a system of definition is suitable for conventional land cover mapping by ground observation (field survey) or visual image interpretation when the interpreter combines image and auxiliary information to classify an object. In computer processing there are two methods of analysis which are mostly used for land use/land cover classification. They are supervised and unsupervised classification. Supervised image classification relies on statistical parameters of a class generated during training sampling which are in general nontransferable from one image to another. Unsupervised classification with a clustering technique provides automated grouping but there is no way to establish a fixed relation between a cluster code and certain land cover category. Moreover the post-classification interpretation of results is time consuming and a subjective process that requires extensive ground truth data collection. The huge amount of imagery information collected by high resolution remote sensing satellites such as LANDSAT and SPOT and especially by the medium resolution multi-sensor satellites TERRA (1999) and ADEOS-II (2001), which will complete global coverage every 4 days, requires development of a new classification technique which allows fast and automated analysis of both single date and multi-temporal data sets for land cover mapping. Recently, research into an automated classification of land cover was initiated by the author in the framework of the NASDA Research Announcement for the ADEOS-II satellite, which will carry a GLI sensor that has 6 spectral channels similar to those of LANDSAT TM. One of the issues of automated classification of GLI data is to develop a system for defining land cover class in the image domain. This means a land cover category should be described by values derived from image data. The author has discovered several image invariants based on graphical analysis of the spectral reflectance curve of a pixel. The invariant found include modulation of the spectral reflectance curve, total reflected radiance index (TRRI) and spectral angles. These invariants seem to be quite stable for image data generated by the same sensor. In this paper the author reports on efforts to develop a digital definition of land cover using image invariants for automated classification. The proposed legend has been applied to a data set simulated for the future GLI sensor and the obtained classification result has proved the chosen approach to be correct.

2 REMOTE SENSING BASED LAND COVER CLASSIFICATION SYSTEM

There are several land cover classification systems announced by IGBP-DIS, FAO, UNESCO, CORINE or LCWG/AARS (Land Cover Working Group / Asian Association on Remote Sensing) that were designed based on different concepts. In general, all these systems meet both scientific needs (global change studies) and social needs (global, continental and national land use planning). Most of these systems have well developed hierarchical structures so that classes of the same level will have similar characteristics. However, each land cover class is described by a terminology and descriptors which follow conventional land cover mapping concepts and they are mainly suitable for integrated visual image interpretation and rather difficult for application in digital image processing, especially when using multi-temporal remote sensing data. To allow automated classification of land cover, each land cover category should be organised to have three components as in Table 1.

Land cover definition components	Information sources
Static component	Single date remote sensing image
Dynamic (seasonal change) component	Multi-temporal remote sensing images
Information on broader biophysical and socio-economic circumstances	Auxiliary information (topography, soils, climate)

Table 1. Components of land cover category definition

The static component describes current physical status of the cover at the moment of observation. This type of information could be extracted from single date remote sensing data. Example of this type of data includes water, vegetation of different coverage densities, soil types (sandy, muddy, dry, wet etc.).

The dynamic component (seasonal change or variation) is extracted from a multi-temporal remote sensing data set. This type of information reflects change of leaf coverage, water level or dryness of certain land cover categories.

The information on broader biophysical and socio-economic circumstances can not be derived from remote sensing data and it should be extracted from other information sources or database such as topography, soils or climate.

Considering the above idea, a flowchart for land cover mapping is proposed in Figure 1.

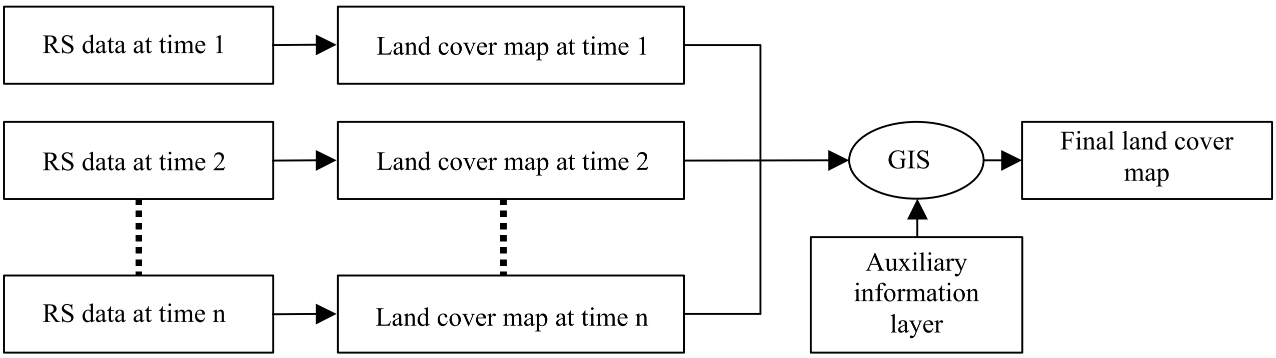


Figure 1. Flowchart of land cover mapping by multi-temporal remote sensing data

Based on the research undertaken in the framework of NASDA RA pre-launch algorithm development for land cover mapping by ADEOS-II GLI data, the author proposes the following land cover classification scheme. In this system land cover is divided into categories with different leaf coverage and water component percentages. Each class is coded by two digits, the first one indicates the dynamic component and the second one is linked to the static component. The dynamic component is ranked into 9 groups so that total number of classes for single date data analysis will be 100. In combination with multi-temporal data, the final land cover map can have 255 codes for different categories. Table 2 shows a proposal of the classification system for single date remote sensing data.

Class code	Dynamic component	Static component
91	**Vegetation with 70 – 100% coverage**	Broad leaf forest
92		Forest plantation
93		Needle leaf forest
94		Mangrove forest
95		Cropland
96		Other grass type vegetation
81	**Vegetation with 50 – 70% coverage**	Broad leaf forest
82		Forest plantation
83		Needle leaf forest
84		Mangrove forest
85		Cropland
86		Other grass type vegetation
71	**Vegetation with 30 – 50% coverage**	Broad leaf forest
72		Forest plantation
73		Needle leaf forest
74		Mangrove forest
75		Cropland
76		Other grass type vegetation
61	**Vegetation with 10 – 30% coverage**	Shrub land
62		Woodland
63		Wetland Shrub
64		Cropland
65		Other grass type vegetation
51	**Vegetation with 0 – 10% coverage**	Grassland
52		Wetland Shrub
53		Cropland
41	**Non-biotic cover**	Rock
42		Sand
43		Cloud
44		Construction
45		Dry soil
46		Cloud or topographic shadow

31	**Muddy surface**	Muddy land
21	**Turbid water**	Turbid water
11	**Water and other hydrographic bodies**	Clear water
12		Snow and ice

Table 2. Proposal of classification system for single date remote sensing data

3 DEFINITION OF LAND COVER CATEGORIES BY IMAGE INVARIANTS

Each land cover category features unique spectral absorption characteristics. For a remote sensing sensor with a fixed spectral channel composition, these characteristics should be unique and stable for a certain land cover class. There are many ways to extract a feature which is unique for a certain land cover class, the author has chosen a method called Graphical Analysis of Spectral reflectance Curve (GASC) to define image invariants. Each land cover class can be described by a set of invariants derived from normalized pixel vector (Nguyen Dinh Duong, 1997 and 1998). According to the latest research result, each land cover class could be described by some of the following invariants:

- Spectral curve modulation
- Total reflected radiance index TRRI
- Band ratios
- Hue angle
- Saturation angle
- Difference of normalized spectral values

The Hue and Saturation angles are computed based on a compression model of 6 spectral channel data (TM or GLI) into three components developed by the author using hexacone colour space.

Table 3 gives an example of the digital definition for some land cover categories based on proposed image invariants.

4 PRELIMINARY RESULT OF AUTOMATED CLASSIFICATION

The above proposed land cover classification has been used for classification of TM and GLI simulated data. A computer program for automated classification has been written by the author. The program runs on DOS prompt in command line mode which provides batch processing ability. The control file contains beside basic information about the data set such as number of lines, rows, file names for different spectral channel data file etc. also a table of classification rules for land cover categories. Structure of digital legend for a land cover category is as follows:

- Classification method (Graphical Analysis of Spectral Reflectance Curve)
- Data set name (GLI 250m channels)
- Number of classes (255 is maximum)

The following is repeated for the number of classes

- Class code (between 1 and 255)
- Full name of class (127 characters)
- Short name of class (6 characters)
- RGB Colour for visualization (example 1 213 255)
- M code - Modulation invariant 0 - 26
- Dij min max - Difference of channels i and j (example D15 15 60)
- T min max - TRRI (example T 2 10)
- H min max - Hue angle between 0 - 360 (example H 15 150)
- S min max - Saturation angle between 0 - 60 (example S 5 30)
- Pi min max - Normalized pixel value of channel i 0 - 100 (example P5 0 15)
- Aij min max - Absolute values of difference of channel i and j (example A15 0 20)
- Rij min max - Ratio of bands i and j (example R34 15.5 20.0)
- END the end of the description for one class

The threshold values for each invariant have been computed based on the normalized pixel vector. The normalization should be done so that it eliminates impact of the seasonal variability of solar radiation, sensor sensitivity and degradation and the quantization level of the sensor.

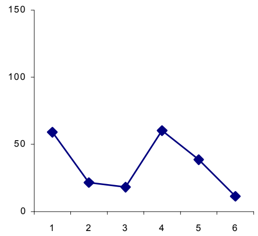
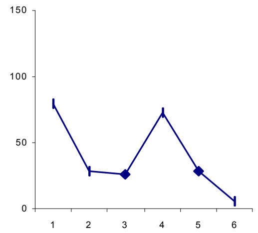
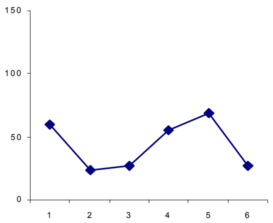
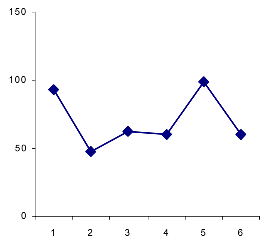
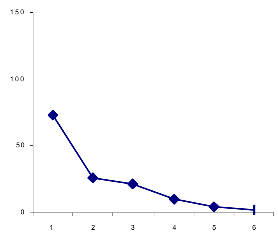
Class Code	Land Cover Category	Characteristic Spectral Curve (without thermal Infrared channel)	Invariants for Classification
91	*Vegetation with 70 – 100% coverage*: Broad leaf forest		M 8 R34 0 0.34 T 3 20 H 180 360
94	*Vegetation with 70 – 100% coverage*: Mangrove forest		M 8 D15 14.5 100 R34 0 0.42 T 3 20 H 30 180 S 0 20
73	*Vegetation with 30 – 50% coverage*: Needle leaf forest		M 26 R34 0 0.57 D45 -5 5 D16 0 16.5 T 3 18 H 180 360
44	*Non-biotic cover:* Construction		M 24 T 20 35 D16 12 24 H 285 328 S 8 15
11	*Water:* Clear water		M 0 T 3 16 H 0 60 S 28 60

Table 3. Example of digital definition for some land cover categories based on proposed image invariants

The actual number of invariants used for the description of a land cover class varies from class to class. An example of definition for a mangrove forest is:

Closed mangrove forest
MGR_F1
153,102,255
M 8
D15 14.5 100
T 3 20
H 30 180
S 0 20
END

The output is organized in two ways. One is the raw classified data file (binary) accompanied by a legend file (ASCII) and the other is a classified image data file in WinASEAN format (Nguyen Dinh Duong, 1997). The second option is convenient for those who are using the desktop public domain image analysis software WinASEAN 3.0 developed by the author. To demonstrate the practical application of the proposed concept, the author has used TM data of South Vietnam for analysis. Two methods of analysis have been compared: maximum likelihood classification and the proposed automated classification. The automated classification program was also applied to GLI data simulated by two TM scenes of South Vietnam. The results of this study is shown in Figures 2,3 and 4.

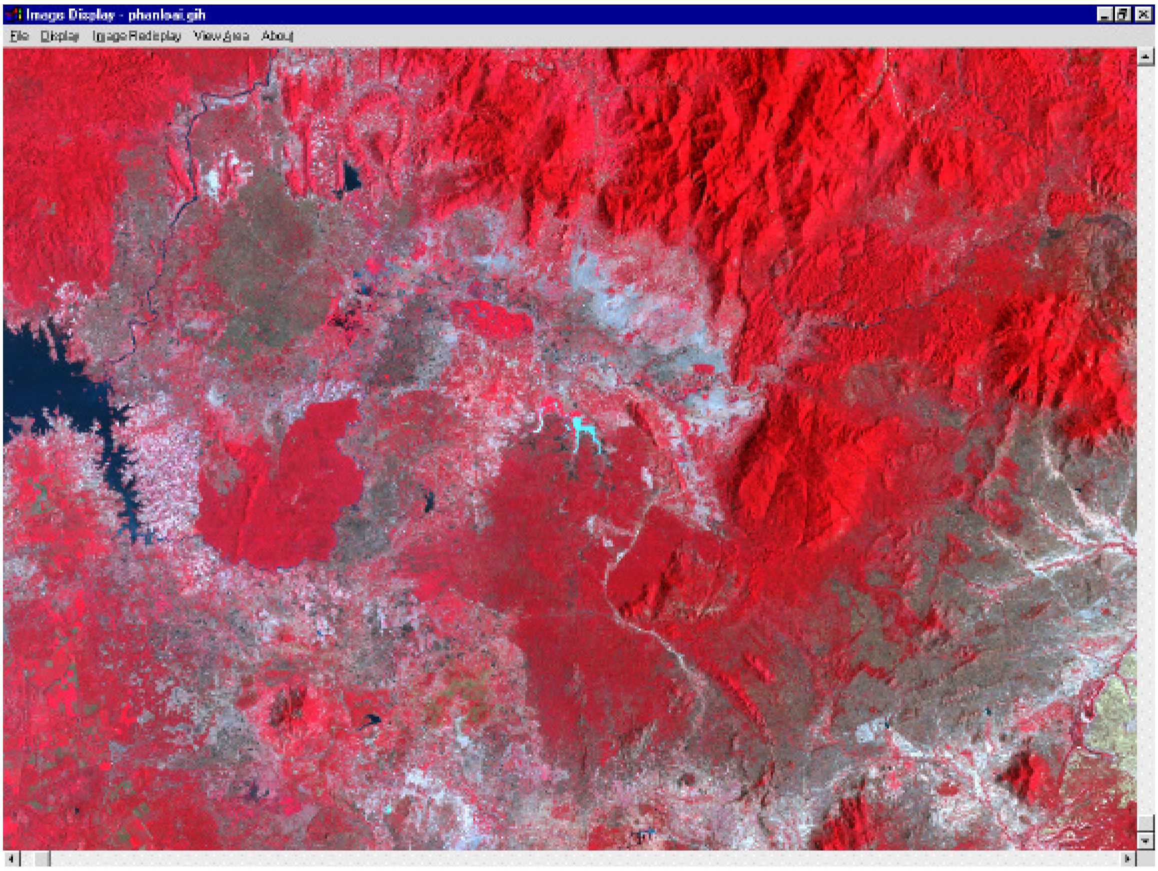


Figure 2. False color image of study area (TM data)

The maximum likelihood classification was executed by selecting 47 different classes representing major land cover categories of the study area. After classification, the classes were merged to create 29 classes including cloud, shadow and unknown which agree with the classification scheme of the GASC algorithm. Comparing the results of both methods, one could recognize their difference mainly in classification of pixel which lay in an overlap between two classes. While in maximum likelihood method the decision is made based on the Bayes solution which requires Gaussian distribution of sample population(that is in itself not always true), the GASC algorithm classifies the pixels according to their spectral reflectance curve and threshold values. In the former method, the feature space of a class is an ellipsoid-like surface and in the latter method the feature space is an arbitrary shape limited only by the modulation of the spectral reflectance curve and prescribed threshold values. Based on analysis using a field survey and local expert knowledge it was seen that the second method gives a more natural result then the first one.

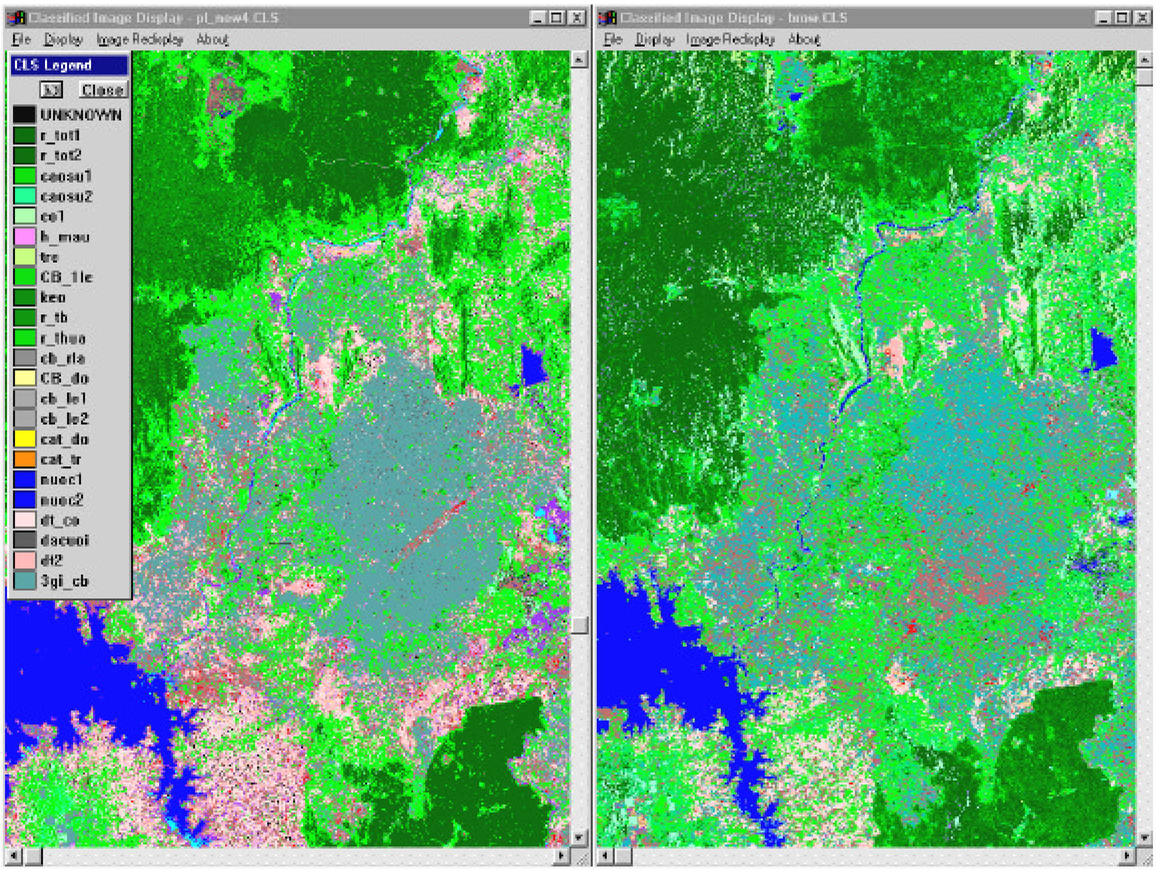


Figure 3. Classification result by Maximum likelihood (left) and GASC (right) methods

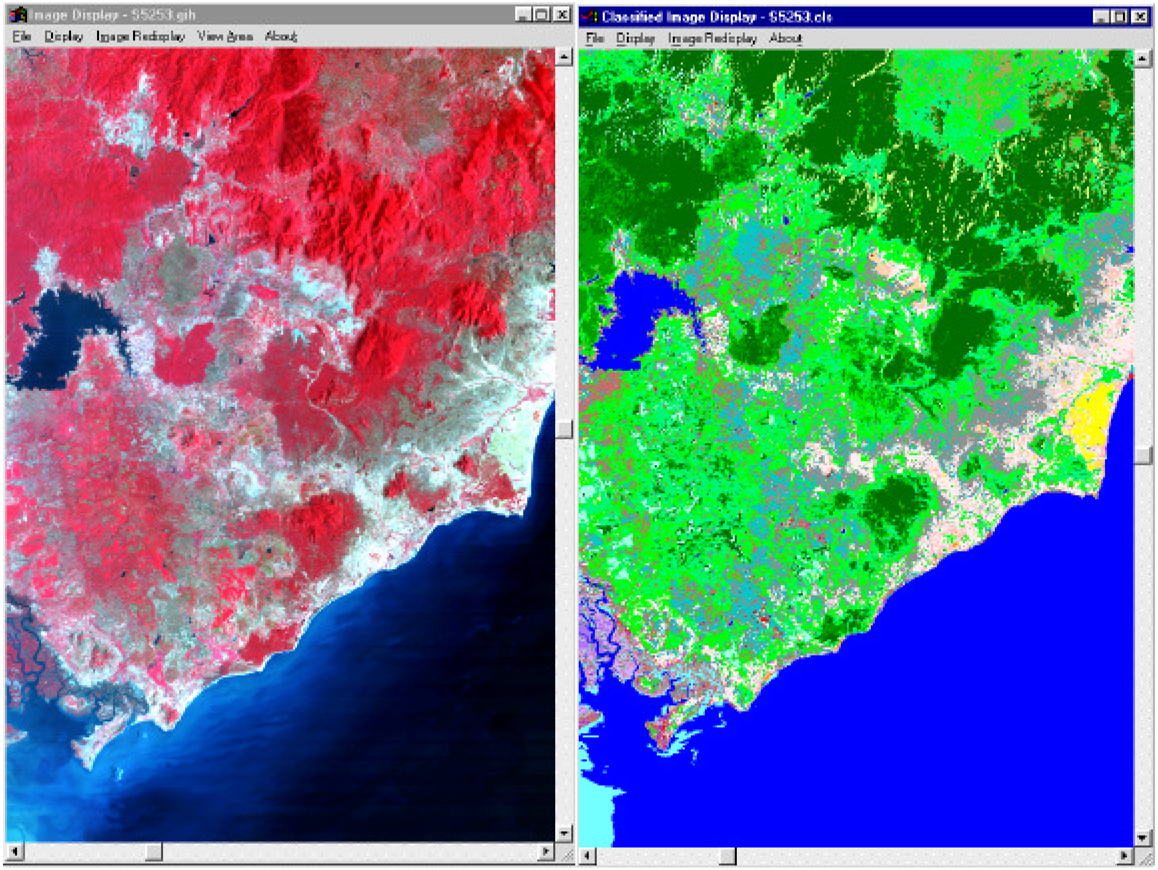


Figure 4. Simulated GLI data (left) and land cover classification by GASC method (right)

In the case of simulated GLI data, the classification result by the GASC algorithm gives a very clear picture of the land cover distribution of the study area. Even though spatial resolution is only 250 m, major ecological zones and important land cover categories, such as mangrove, evergreen broad leaf forest, scrub land, needle leaf forest, could be automatically extracted.

5 CONCLUSION

The proposed land cover classification system based on the digital definition of land cover categories and GASC algorithm is an attempt by the author to develop automated classification of multi-temporal remote sensing data set. The set of descriptors has been developed using TM data without the thermal infrared channel. The actual threshold values will be defined based on real GLI data which will be available after November 2001, when the ADEOS-II is launched. Such a system of classification can be applied for both land use and vegetation cover mapping depending on the availability of auxiliary information layers. An automated environmental monitoring system such as Remote Sensing Based Operational Forest Monitoring System (RSBOFM) could be formulated based on this idea. The concept of land cover definition and the GASC algorithm reported in this paper was specially developed for the classification of future ADEOS-II/GLI data. However, it can be modified and applied to many other remote sensing systems.

ACKNOWLEDGEMENTS

The author would like to thank the National Space Development Agency of Japan (NASDA) for funding this research through NASDA ADEOS-II RA program. The author also acknowledges the help of the research staff of the Environmental Remote Sensing Laboratory, Institute of Geography, Vietnam for their active involvement and support in this research.

REFERENCES

AARS Global 4-minute Land Cover Data Set with Ground Truth Information, Land Cover Working Group of Asian Association on Remote Sensing, CEReS, Chiba University, Japan.

IGBP Report No. 35: Land-Use and Land-Cover Change Science/Research Plan.

Nguyen Dinh Duong, Graphical Analysis of Spectral Reflectance Curve, Proceedings of the 18th Asian Conference on Remote Sensing. 20 –24 October, 1997, Kuala Lumpur, Malaysia.

Nguyen Dinh Duong, Semi-automatic Land Cover Classification using ADEOS/AVNIR Multi-spectral Data, Proceedings of the 18th Asian Conference on Remote Sensing. 20 –24 October, 1997, Kuala Lumpur, Malaysia.

Nguyen Dinh Duong, Total Reflected Radiance Index – An Index to Support Land Cover Classification, Proceedings of the 19th Asian Conference on Remote Sensing. 16 –20 November, 1998, Manila, Philippines.

Nguyen Dinh Duong, Semi-automated Land Cover Classification Using Satellite Data, Proceedings of the 8th Regional Seminar on Earth Observation for Tropical Ecosystem Management, October 25 – 30, 1999, Yangon, Myanmar.

Nguyen Dinh Duong, **Shoji Takeuchi**, WinASEAN for Remote Sensing Data Analysis, Photogrammetry and Remote Sensing, Vol. 52 No. 6, Elsevier 1997.