INTEGRATING TOPOGRAPHIC, SPECTRAL AND STRUCTURAL DATA FOR VEGETATION MAPPING IN COUNTRY PARKS OF HONG KONG

FUNG, TUNG,
Department of Geography,
The Chinese University of Hong Kong,
Shatin, N.T.,
Hong Kong.

ABSTRACT

An accurate inventory of vegetation resources is important for resource management in country parks. The use of spectral image classification of satellite images alone is able to map vegetation types but the resultant accuracy may not be always satisfactory. Topographic data, particularly elevation, slope and aspect, together with structural data extracted from the composition of spectral classes are added to improve classification result. Result shows that combining the spectral, topographical and structural data produces the most accurate classification.
Key words: spectral, DEM, structural, classification

INTRODUCTION

Hong Kong is a metropolitan city with a population of over six million. Out of her 1000 sq. km. of territory, over 10% is devoted to urban land uses. The rapid development of the city has imposed a severe demand of land for urban expansion. Much of these are carried out through sea reclamation and rural land consumption. Despite the urban nature of the city, a large bulk of land, many of them are located on hilly areas, is still covered with a wide range of flora. Since the 1970s, country parks were set up covering about 40% of the land area of Hong Kong (Catt, 1986). These parks are important for the preservation and management of vegetation as well as for the provision of recreational use in the countryside. For better management of resources within the parks, remote sensing data, particularly satellite imagery, offer viable tools for their wide spectral range, repetitive coverage and timeliness. With appropriate image analysis tools applied, it is not only possible to map and update vegetation resources in the parks, frequent stress such as forest fire can be monitored as well.

Satellite imagery has been widely used to map vegetation resources in different environments. To improve classification accuracy, information from digital elevation models (DEM) has been incorporated with spectral data for image classification (Cibula and Nyquist , 1987; Franklin, 1987). DEM data have been applied to alter the prior probability, to stratify the image for classification or to be incorporated as additional channels for classification (Hutchison, 1982). Textural as well as contextual data extracted from images have also been used to improve classification results (Franklin and Peddle, 1990; Gong and Howarth, 1990). Structural information such as the composition of spectral classes has also been used for map generalization (Wang et al., 1991). This information is also useful for the analysis of high frequency features (Fung, 1991).

The purpose of this paper is to map vegetation in the Tai Lam Country Park of Hong Kong. A SPOT HRV multispectral data together with DEM and structural data extracted from spectral classes are used for image classification. Accuracies of image classification are evaluated.

THE STUDY SITE

The Tai Lam Country Park is located at the western part of the territory of Hong Kong. It covers an area of 5,330 hectares. The Tai Lam Chung Reservoir is located at the centre of the park. Elevation ranges from 60 m to 500 m. Located on a granitic bedrock, this area was once notorious for the widespread distribution of badland. Since the 1960s, afforestation program has been implemented to control badland development as well as protect water catchments. Trees such as *Pinus massoniana* and *Acacia confusa* have been planted. Scattered badland can still be found along hill ridges at the western part of the Park. This part of the park has steep slope with rock outcrop scattered around where hill grassland is more commonly found.

DATA DESCRIPTION

A SPOT HRV multispectral data acquired on January 14, 1987 is available for the analysis. A rectangular image of 300 X 250 pixels is georeferenced with reference to the Hong Kong Metric Grid System at 20 X 20 m pixel size. This image was taken during the dry winter season. With a low humidity and moisture content, a distinctive contrast between afforested woodland and hill grassland is found where the former appears in a rich red hue whilst the latter is yellow in a false color composite. Badland appears as bright white scars on the hilltops.

For better identification of different vegetation classes, a normalized differenced vegetation index is calculated as:

$$NDVI = \frac{IR - R}{IR + R} \quad \text{or}$$

$$NDVI = \frac{XS3 - XS2}{XS3 + XS2}$$

This NDVI is incorporated with other spectral channels for image classification.

Digitized contour lines at a scale of 1:20,000 is available in ARC/INFO format. This vector file is created at a contour interval of 20 m. This vector file is rasterized from which spatial interpolation is performed. Based on the DEM, other topographic variables of slope and aspect are generated.

METHODOLOGY

Generation of spectral classes

An unsupervised classification is first performed using the three SPOT multispectral channels and the NDVI. Fifteen classes are generated. Nature of them is identified as follows:

1. water
2. coniferous woodland shadow/ water
3. water
4. coniferous woodland shadow/ water
5. wet bare soil
6. broadleaf woodland
7. wet bare soil/water
8. woodland shadow
9. wet bare soil/grassland
10. coniferous woodland/scrubland
11. scrubland/badland
12. grassland
13. badland
14. grassland
15. badland

Generation of Structural Information

In the study of urban land uses of Hong Kong, Fung (1991) noted that the composition of spectral classes can yield important information particularly for heterogeneous land uses. A land use can be described by its composition of spectral classes as

$$LU_i = f\{SC_1, SC_2, \ldots, SC_n\}$$

where LU_i is the land use type i
$SC_1, SC_2, \ldots, SC_n$ are the spectral classes derived from any classification.

This information is then generated by counting the proportion of each spectral class within a land parcel:

$$LU_i = f\{p(SC_1), p(SC_2), \ldots, p(SC_n)\}$$

where LU_i is the land use and land cover type i
$p(SC_j)$ is the proportion of the jth spectral class within the land parcel.

It was found that different land uses possess different composition of spectral classes. This information is useful for the study of complicated urban land uses. It is thus used to examine this structural information can be useful in a non-cultural environment as well.

To generate structural information from the spectral classes, a 3 x 3 window is used to count the frequency of occurrence of each spectral class. A total of 15 images is thus generated in which pixel values range from 0 to 9 depicting the likelihood of a particular spectral class to occur. A principal components transformation is performed on these 15 images to reduce data redundancy and to extract useful structural information. Table 1 shows the first three eigenvectors and their associated eigenvalues. The first three principal components, accounting for 73.97% of total variance are then used for further analysis. The principal components are basically area weighted by the corresponding spectral classes. The first principal component

Table 1: Principal Components of the Spectral Classes

Spectral class	PC1	PC2	PC3
1. water	-0.003	0.008	-0.001
2. coniferous woodland shadow/ water	-0.000	-0.000	0.001
3. water	-0.027	0.094	-0.017
4. coniferous woodland shadow/ water	-0.002	0.005	0.003
5. wet bare soil	-0.009	0.029	0.011
6. broadleaf woodland	0.888	-0.286	-0.021
7. wet bare soil/water	-0.001	0.006	-0.001
8. woodland shadow	-0.088	0.313	0.734
9. wet bare soil/grassland	-0.005	0.015	-0.001
10. coniferous woodland/scrubland	-0.069	0.127	0.074
11. scrubland/badland	-0.181	0.235	-0.660
12. grassland	-0.028	0.066	-0.038
13. badland	-0.031	0.086	-0.082
14. grassland	-0.401	-0.849	0.089
15. badland	-0.024	0.071	-0.051
eigenvalue	1804.565	1014.941	671.119
% variance	38.24	21.51	14.22

(PC1) explains the distribution of woodland mainly. PC2 explains the distribution of grassland and badland whereas PC3 accounts for the woodland shadow and grassland.

A total of 10 data channels are thus used for the analysis. They are:

spectral data	1. SPOT XS1
	2. SPOT XS2
	3. SPOT XS3
	4. NDVI
DEM	5. elevation
	6. slope
	7. aspect
structural	8. PC1
	9. PC2
	10. PC3

A maximum likelihood classifier is adopted for image classification. A supervised approach is used in which training sites are selected for 10 classes. They are:

1. water
2. badland
3. broadleaf woodland (mainly eucalyptus)
4. broadleaf woodland and scrubland
5. coniferous woodland (mainly pines)
6. coniferous woodland on shadowed slope
7. mixed grass and rock outcrops
8. grassland
9. scrubland
10. agricultural land

In order to compare the effectiveness of different types of data, four classifications are performed with different combinations of data. They are:

1. spectral data only
2. spectral data and DEM data
3. spectral data and structural data
4 spectral, DEM and structural data

Test sites are selected to assess the accuracies of classification. Error matrices together with the overall accuracies and Kappa coefficients of agreement are computed to compare different results.

Table 2: Classification Results

classifications	overall accuracy	Kappa coefficient of agreement
1. spectral	75.73	73.05
2. spectral + DEM	84.03	82.19
3. spectral + structural	80.11	77.58
4. spectral + DEM + structural	86.12	84.76

RESULTS AND DISCUSSION

Table 2 illustrates the accuracy levels of the four classifications. Using spectral data alone is able to yield an overall accuracy of 75.73% and a Kappa coefficient of 73.05. It is the lowest of the four classifications. Major confusion is found among (3) broadleaf woodland (4) broadleaf woodland and scrubland and (5) coniferous woodland. Similar confusion is found in other classification images as well.

Incorporating other data for classification does improve classification results. Combining the spectral and DEM data, the overall accuracy and Kappa coefficients produced are 84.03% and 82.19 respectively. Accuracy is lower if structural data is used. This means that the DEM data is relatively more important in the improvement of classification results. Combining all the data produces the most accurate result (Figure 1). The overall accuracy and Kappa coefficient are 86.12 and 84.76 respectively. They are slightly higher than that of using spectral and DEM data together.

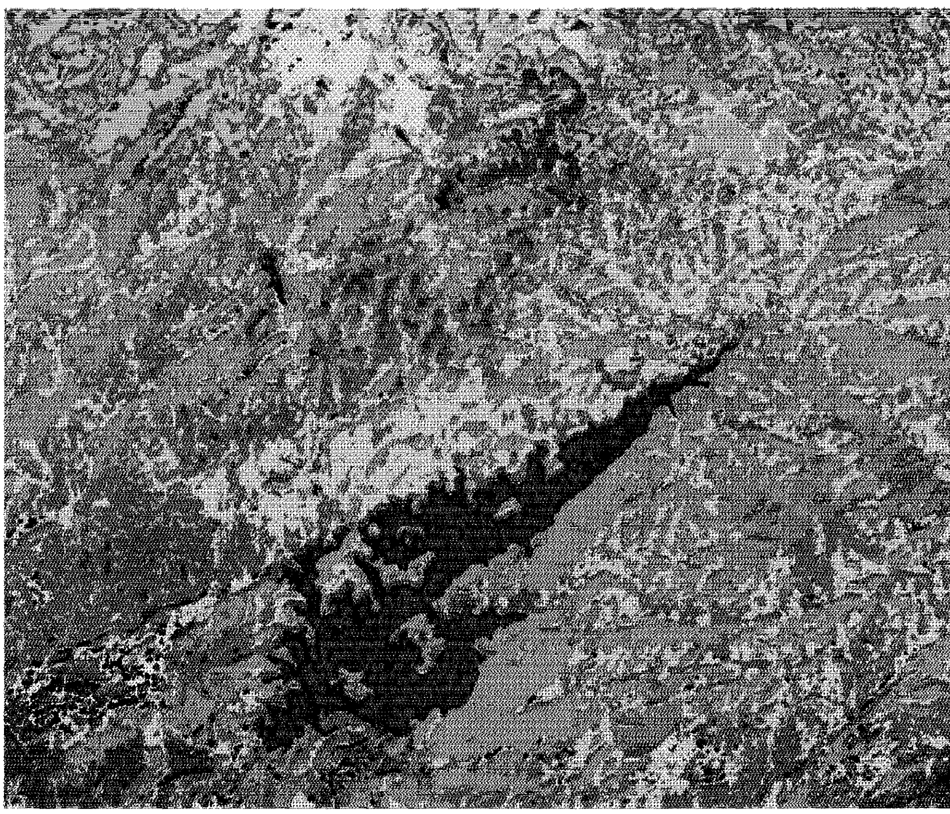
This result indicates that adding terrain variables and structural information can improve classification accuracy. In the study sites, the distribution of vegetation is associated with the terrain configuration. Along the valleys, scrubland and woodland are more commonly found where there are shelter and availability of water. In contrast, more grassland is located on hill tops and ridges and steep slopes. Topographic variables of aspect and slope are able to account for such a distribution.

The use of structural information is less useful as it can only slightly improve the classification result. The use of a small window size of 3 X 3 may not be able to capture essential variation of structural details. More importantly, structural information is found to be useful in the identification of heterogeneous land uses, particularly urban land uses. In the study area, the landscape is less heterogeneous than urban landscape. Therefore, the result becomes less promising.

CONCLUSION

In this paper, spectral data are integrated with topographic and structural data for image classification. It is concluded that incorporating ancillary data can improve remote sensing image classification. The use of topographic variables, elevation, aspect and slope is the most important as the distribution and growth of vegetation is much related to the terrain configuration. Using structural information can only slightly improve classification result. For further study, it is found that terrain features, such as ridges and valleys and their distribution are important in shaping the distribution of vegetation. Extraction of these features may offer even better results for classification in the study area.

Figure 1: Classification images generated from spectral, topographic and structural data



1. water
2. badland
3. broadleaf woodland (mainly eucalyptus)
4. broadleaf woodland and scrubland
5. coniferous woodland (mainly pines)
6. coniferous woodland on shadowed slope
7. mixed grass and rock outcrops
8. grassland
9. scrubland
10. agricultural land

REFERENCES

Catt, P. (1986), "Vegetation" in Chiu T.N. and C.L. So (eds.) A Geography of Hong Kong, second edition, Oxford, Hong Kong, 118-147.

Cibula, W.G. and M.D. Nyquist, "Use of topographic and climatological models in a geographical data base to improve Landsat MSS classification for Olympia Park", Photogrammetric Engineering and Remote Sensing, 53(1), 67-75.

Franklin, S.E., 1987, "Terrain analysis from digital patterns in geomorphometry and Landsat MSS spectral response", Photogrammetric Engineering and Remote Sensing, 53(1), 59-65.

Franklin, S.E. and D.R. Peddle, 1990, "Classification of SPOT HRV imagery and texture feature", International Journal of Remote Sensing, 11(3), 551-556.

Fung, T., 1991, Spectral composition: a structural approach for image analysis of high frequency features, paper presented at AAG Annual Conference, Miami, April, 1991.

Gong, P. and P.J. Howarth, 1990, "The use of structural information for improving land cover classification accuracies at rural-urban fringe", Photogrammetric Engineering and Remote Sensing, 56(1), 67-73.

Hutchison, C.F., 1982 "Techniques for combining Landsat and ancillary data for digital classification improvement", Photogrammetric Engineering and Remote Sensing, 49(1), 123-130.

Wang, E. P. Gong and P.J. Howarth, 1991, "Thematic Mapping from imagery: an aspect of automated map generalization", Technical papers 1991 ACSM-ASPRS Annual Convention/Auto-Carto 10, 123-132.