Land Use Classification of TM Imagery in Hilly Areas: Integration of Image Processing and Expert Knowledge

Feng Ding Wenhui Chen Daxian Zheng
Institute of Geography, Fujian Normal University
32 Shang San Road, Cang Shan District, Fuzhou, Fujian, 350007, China
fding@vip.sina.com whchen@sohu.com dxzheng@fjnu.edu.cn

Abstract: Improvement of the classification accuracy is one of the major concerns in the field of remote sensing application research in recent years. Previous research shows that the accuracy of the conventional classification methods based only on the original spectral information were usually unsatisfied and need to be refined by manual edit. This present paper describes a method of combining the image processing, ancillary data (such as digital elevation model) and expert knowledge (especially the knowledge of local professionals) to improve the efficiency and accuracy of the satellite image classification in hilly land. Firstly, the Landsat TM data were geo-referenced. Secondly, the individual bands of the image were intensity-normalized and the normalized difference vegetation index (NDVI) image was also generated. Thirdly, a set of sample pixels (collected from field survey) were utilized to discover their corresponding DN (digital number) ranges in the NDVI image, and to explore the relationships between land use type and its corresponding spectral features. Then, using the knowledge discovered from previous steps as well as knowledge from local professionals, with the support of GIS technology and the ancillary data, a set of conditional statements were applied to perform the TM imagery classification. The results showed that the integration of image processing and spatial analysis functions in GIS improved the overall classification result if compared with the conventional methods.
Keywords: Land use classification, TM, GIS, Expert knowledge.

1. Introduction

Earth surface system is the material base human being relies on. Its most remarkable landscape characteristic is land use / or land cover [1]. With the rapid development of computer science and spatial information technology, Remote Sensing, integrated with GIS and GPS technology, has become one of the most powerful tools for monitoring land use changes. Several methods and papers exist for land use classification using satellite Remote Sensing data [2] [3] [4]. Previous research shows that, improvement of the remotely sensed data classification accuracy is the major concern of many researchers [5] [6] [7] [8] [9] [10]. In conventional Remote Sensing classification methods, Maximum Likelihood (ML) classifier is most widely used since it takes into account the shape, size and orientation of a cluster thus it can usually provide the best results. But this classification method assumes the training samples to be normally distributed [11]. Such ideal situation may not occur in many cases especially in mountainous and hilly areas where the impacts of topographic shadows on the remotely sensed data were obvious and inevitable. In present paper, intensity normalization of the individual spectral bands of the TM imagery is introduced and implemented to minimize the topographic variations. Previous study also shows that spectral classification alone is very often not sufficient for extracting land use data, the conventional classification methods based only on the original spectral information were usually far from satisfaction and need further improvement by manual edit. By combining the ancillary data (such as the digital elevation model and the digital historical land use map) and expert knowledge of the study area, the classification accuracy can be apparently improved, as demonstrated below taking Langqi Island as an example.

2. Study Area and Data

The study area (Langqi Island, belongs to Fuzhou city, Fujian province, China) selected for this research is located in the southeast coast of China. As an agriculture-first and hilly island, quantity and quality of the cultivated land is always the major concern of land use related decision-makers and researchers in this area. Available Landsat TM data of this area was acquired on 23rd September 1999. A sub-image of size 391 pixels (columns) by 294 pixels (rows) covering the Langqi Island was extracted for processing. Data available also include: Langqi Island land use map (Year 1998, digital format, scale 1:10,000) and Langqi Island digital elevation model (DEM), at 30 m resolution, derived from the scanned topographic map of the study area (20 m contours, scale 1:10,000).

3. Data Processing

1) TM Image Processing and Maximum Likelihood Classification

The subset Landsat TM image covering the area of interest was firstly geo-referenced using control points from the 1:10,000 topographic map of the study area (In order to combine data from different sources for further analysis, all data should be converted or registered to the same coordinate system, in this research, the Gauss-

Krüger map projection with Beijing 54 coordinate system was chosen). All the spectral bands, excluding TM band 6, were used. To minimize the topographic effects, Shrestha *et al.* put forward the formula for spectral bands normalization [12] and in present study it was adopted and edited as follows:

$$NB_i = 255(OB_i / \Sigma OB_i + C) \quad i = 1 \text{ to } 6 \quad (1)$$

Where OB_i is the original spectral band_i and NB_i is the newly created band normalized by total intensity (C is a small enough number, e.g., 0.01, to make sure the numerator will not be divided by zero). The constant 255 is used to fit the data in a range of 0 to 255. The resulting bands have the characteristics that the sum of any pixel values is 255 due to normalization. After normalization of the original bands, it can be assumed that the TM data is nearly normally distributed. The ML algorithm is then applied to classify the TM image. The classification involved in the identification of paddy field, dry land, vegetable field, orchard, forest, residential areas, water body, and sea beach. The classification results were checked by a set of testing pixels collected from field survey, showing 88 percent accuracy for water body, 80 percent accuracy for forest, 76 percent for sea beach, 74 percent accuracy for orchard, 73 percent for residential area, 68 percent for vegetable field, 67 percent for paddy field and 62 percent for dry land (Table 1). As shown in Table 1, water body got the highest classification accuracy while the cultivated land (including paddy field, dry land and vegetable) got the lowest one. It was found that classification of cultivated land was only about 66 percent correct since some of the testing pixels were misclassified as orchard (mainly) and other land use types. The results clearly demonstrated the difficulty in classifying cultivated land and orchard.

Table 1. Accuracy of the ML classification results.

Land use type	Accuracy (%)
Paddy field	67
Dry land	62
Vegetable field	68
Orchard	74
Forest	80
Residential area	73
Water body	88
Sea beach	76

2) Improving Cultivated Land Classification

To explore the relationship between each land use type and its corresponding spectral features and to improve the classification accuracy of cultivated land, the ancillary data, i.e., the historical land use map of Langqi Island (Year 1998), the DEM and the NDVI were introduced and the following steps were taken:

Step 1: Firstly, the Langqi Island land use digital map (in vector format) was converted from coverage to grid by using the ARC/INFO command POLYGRID. Secondly, GRIDIMAGE command was used to convert the grid into the specified image format (in this research, ERDAS IMAGINE's img file format was the choice) [13]. Thirdly, the newly created land use image was georeferenced into the Gauss-Krüger map projection with Beijing 54 coordinate system. Fourthly, the cultivated land use image and the orchard land use image were extracted from the geo-referenced land use image by using the conditional statements such as "EITHER (land use map image pixel value) IF (land use map image pixel value = 10) OR 0 OTHERWISE" in the ERDAS IMAGINE environment. Fifthly, the extracted cultivated land use image and orchard land use image was used as mask respectively to subset the geo-referenced TM image (before normalization). Therefore, we got the multi-spectral images (containing all the 6 bands, i.e., TM band 1,2,3,4,5 and 7) for cultivated land and orchard and the spectral features of these two land use types could be explored and found. That could serve as spectral knowledge for further processing.

Step 2: Except for the spectral knowledge discovered from previous steps, the intensity normalized difference vegetation index (NDVI) was also generated from the spectral bands in the near-infrared and red portions of the spectrum (Landsat TM bands 4 and 3), using the following formulae which was incorporated in the ERDAS IMAGINE's graphical model library named "vge_ndvi.gmd" [14]:

$$N_i = (TM4 - TM3) / (TM4 + TM3) \quad (2)$$

$$NDVI = 255 [N_i - Min (N_i)] / [Max (N_i) - Min (N_i)] \quad (3)$$

Where formula (2) was used to calculate the original values of the intensity normalized difference vegetation indexes of each pixel of the image while the formula (3) was used to convert the fractional values returned from formula (2) to integers and make them fall into the range of 0 to 255 (i was the number of the pixels of the image).

By overlaying the land use map imageries (generated in Step 1) with the NDVI image, experiments were conducted to find the NDVI threshold value ranges for differentiation between vegetation and non-vegetation (K1 to K2) and between cultivated land and orchard land (K3 to K4). These threshold value ranges were also to be served as spectral knowledge.

During the field survey, it was noticed that cultivated land (especially the paddy field and vegetable field) of the Island generally did not develop at elevations higher than 20 m above mean sea level (confidence level = 0.85) while the orchard usually grow on the elevations higher than 20 m (confidence level = 0.75). The confidence level was decided by the local professionals and had been tested through random samples testing.

By combining all the knowledge discovered from the steps mentioned-above, some conditional ("IF, THEN, ELSE") and logical ("AND") statements were put forward as follows:

IF NDVI > K1 AND NDVI < K2
THEN Land use = “vegetation”
ELSE land use = “non-vegetation”

IF NDVI > K3 AND NDVI < K4 AND DEM < 20 m
THEN land use = “Cultivated land”

Where (K1, K2) was the NDVI threshold value range for differentiation between vegetation and non-vegetation while (K3, K4) was the threshold value range for “Cultivate land” and “Orchard”.

By using the above conditional statements, the overall classification accuracy of the cultivated land and orchard were improved to 73% and 78% respectively.

4. Conclusions

Maximum Likelihood (ML) classification could produce good results with the assumption that the distribution of its training samples was normally distributed. To minimize the topographic effects on the classification of TM imagery in hilly area, the method of normalization of the individual bands by total intensity was introduced and then the ML algorithm was applied. Integrating multiple approaches yields better results than just only one method. This research has advantages over the traditional methods of land use classification by integrating the ancillary data (e.g., historical land use digital map and DEM) and expert knowledge into the image classification process. Therefore, the overall classification accuracy was considerably improved.

Acknowledgement

This study was funded by the project “Fujian Environment and Resources Dynamic Monitoring System Based on 3S Technology” (approved by the Science and Technology Department of Fujian Province under the grants of 99-H-2) as well as by the “Digital Fujian” project “Fujian Eco-environment Dynamic Monitoring and Management System” (approved by the Development and Planning Committee of Fujian Province, 2002-147).

References

[1] Shi, P. J., P. Gong, X. B. Li, and X. B. Li, 2000. The Method and Practice of Land Use / Land Cover Change, Science and Technology Press, Beijing. (in Chinese).

[2] Bauer, M. E., T. E. Burk, A. R. Ek, P. R. Coppin, S. D. Lime, T. A. Walsh, D. K. Watlers, W. Befort and D. F. Heinzen, 1994. Satellite Inventory of Minnesota Forest Resources, *Photogrammetric Engineering and Remote Sensing*, 60: 287-298.

[3] Binaghi, E., P. Madella, M. G. Montesano and A. Rampini, 1997. Fuzzy Contextual Classification of Multisource Remote Sensing Images. *IEEE Transactions on Geoscience and Remote Sensing*, 35: 326-340.

[4] Goetz, S. and S. Prince, 1996. Remote Sensing of Net Primary Production in Boreal Forest Stands. Agraiculture and Forestry Meteorology, 78: 149-179.

[5] Yang, C. J., and C. H. Zhou, 2001. Extracting Residential Area from TM Image on the Basis of Knowledge Discovered, *Remote Sensing Technology and Application*, 16(1): 1-6. (in Chinese).

[6] Buheaosier, 1999. Classification of Saline Soil Based on Knowledge Discovery and Rule Base System Using Remote Sensing Data, *J. Image and Graphics*, 4A(11): 965-969. (in Chinese).

[7] Yang, C. J., and C. H. Zhou, 2001. Investigation on Classification of Remote Sensing Image on Basis of Knowledge, *Geography and Territorial Research*, 17(1): 72-77. (in Chinese).

[8] Hu, Y. Y., S. T. Cheng, H. F. Jia, H. X. Zhang, B. Tan, and J. Zhao, 2003. Extract Non-point Pollution Source Information from a TM Image by Hierarchical Classification, *Environmental Science*, 24 (1): 152-156. (in Chinese).

[9] Yang, C. J., and C. H. Zhou, 2000. Extracting Resid ential Areas on the TM Imagery, *J. Remote Sensing*, 4(2): 146-150. (in Chinese).

[10] Shu, H. L., and Z. Y. Mao, 1997. Knowledge Based Image Classification Approach Supported by a GIS, *ACTA GEODAETICA et CARTOGRAPHICA SINCA*, 26(4): 328-336. (in Chinese).

[11] Lillesand, T. M., and R. W. Kiefer, 2000. *Remote Sensing and Image Interpretation*, John Wiley & Sons, Inc., NY, USA, pp. 541-544.

[12] Shrestha D. P. and J. A. Zink, 2001. Land Use Classification in Mountainous Areas: Integration of Image Processing, Digital Elevation Data and Field Knowledge (Application to Nepal), *JAG*, 3(1): 78-85.

[13] ESRI Inc., 2000. *ArcDoc* (ARC/INFO Workstation Online Help), Redlands, CA, USA.

[14] ERDAS Inc., 2001. *ERDAS Tour Guide*, ERDAS Inc., Atlanta, USA.