Object-based land use/cover extraction from QuickBird image using Decision tree

Eltahir Mohamed Elhadi [12]and Nagi Zomrawi[2]
1-China University of Geosciences Faculty of Resources, Wuhan, 430074, China,
2-Sudan University of Science and Technology Faculty of Engineering, Khartoum,Sudan,
tahirco2006@yahoo.com, nagizomrawi@yahoo.com

Abstract The traditional pixel-wise statistical and mono-scale based classification approaches do not lead to satisfactory results for neglecting the shape and context aspects of the image information, which are among the main clues for information extraction at very-high spatial resolutions like QuickBird image. This paper extracts land use/cover information from occurrence filters texture features that were derived from the grey-level occurrence matrix from QuickBird image using CART Decision tree, because, this method have substantial advantages for remote sensing classification problems due to their nonparametric nature, simplicity, robustness with respect to non-linear and noisy relations among input features and class labels, and their computational efficiency. CART has a simple form which can be compactly stored and that efficiently classifies new data ,also it can recursively partitions a data set into smaller subdivisions on the basis of tests applied to one or more features at each node of the tree. Overall accuracy of texture features using CART Decision tree is higher than other methods. It concluded that texture features can be used to improve classification accuracy. [Journal of American Science 2010;6(2):176-180]. (ISSN: 1545-1003).

Keywords object-based, land use / cover, classification, decision tree,QuickBird

1. Introduction

The automatic analysis of remotely sensed data has become an increasingly important topic over the last decades. Especially land use/cover and land change information is useful for city development. The segmentation of satellite images into regions of different land cover is of major interest: given data from several spectral bands, one wants to determine for each pixel of the image which type of land cover is present at the corresponding area on the surface (Keuchel et al., 2003, Carlson and Arthur, 2000, Le Hegarat-Mascle et al., 2005, Fan et al., 2007). In land cover classification of remote sensing data, it is desirable to use multisource data in order to extract as much information as possible about the area being classified.

However, classification of multisource remote sensing and geographic data is a challenging problem, especially since a convenient multivariate statistical model is in general not available for such data (Gislason et al., 2006). The traditional pixel-wise statistical and mono-scale based classification approaches do not lead to satisfactory results for high spatial resolution remote sensing data like QuickBird image.

The main drawback of these methods is that they neglect the shape and context aspects of the image information, which are among the main clues for information extraction at very-high spatial resolutions. The successful launch of very-high spatial resolution panchromatic and multi-spectral satellites renders the potential to carry out thematic mapping at large scales in urban areas.

Unfortunately, the high spatial resolution of these advanced sensors increases the spectral within field variability and, therefore, may decrease the classification accuracy results.

This is because most classification techniques are based on spectral homogeneities only (Cushnie, 1987), and do not take into account the textural attributes of the mapped image's features. Due to the more heterogeneous spectral-radiometric characteristics within the land-use/cover units portrayed in high resolution images, applications of traditional single resolution classification methods have led to unsatisfactory results. This paper extracts land use/cover information from texture features that were derived from the grey-level occurrence matrix using CART Decision tree.

2. Study area

The study area covers Chenggong districts in Yunnan province in southwest of China (Fig. 1). The centre is latitude 24°55’43″N and longitude 102°50’10″E. The remote sensing data consisted of QuickBird multispectral and panchromatic images that were acquired simultaneously on 4 May, 2004. The QuickBird radiances were not atmospherically corrected as time series analysis of consecutive image data was not required for this study, and detailed information on the atmospheric conditions at the time of overpass was not available.

3. Methods

Within the last 10 years, there has been increasing interest in the use of classification and regression tree (CART) analysis. CART analysis is a tree–building technique which is unlike traditional data analysis methods .Because CART analysis is unlike other analysis methods it has been accepted relatively slowly .Furthermore, the vast majority of statisticians have little or no experience with the technique. Other factors which limit CART analysis general acceptability are the complexity of the analysis and, until recently, the software required to perform CART analysis was difficult to use. Luckily, it is now possible to perform a CART analysis without a deep understanding of each of the multiple steps being completed by the software .In a addition, CART is often able to uncover complex interactions between predictors which may be difficult or impossible to uncover using traditional multivariate techniques.

CART analysis has a number of advantages over other classification methods, including multivariate logistic regression, first, it is inherently non-parametric .In other words, no assumptions are made regarding the underlying distribution of values of the predictor variables. Thus, CART can handle numerical data that are highly skewed or multi-model, as well as categorical predictors with either ordinal structure (Quinlan, 1993). This is an important feature ,as it eliminates analyst time which would otherwise be spent determining whether variables are normally distributed, and making transformation if they are not.

As discussed bellow ,CART identifies "splitting" variables based on an exhaustive search of all possibilities .Since efficient algorithms are used, CART is able to search all possible variables as splitters ,even in problems with many hundreds of possible predictors.

Finally, another advantage of CART analysis is that it is a relatively automatic" machine learning" method .In other words, compare to the complexity of the analysis, relatively little input is required from the analyst. This is in marked contrast to other multivariate modeling methods, in which extensive input from the analyst, analysis of interim result, and subsequent modification of the method are required.

Despite its many advantages, there are a number of disadvantages of CART which should be kept in mind. First, CART analysis is relatively new and somewhat unknown. Thus, there may be some resistance to accept CART analysis by traditional statisticians. In addition, there is some well-founded skepticism regarding tree methodologies in general, based on unrealistic claims and poor performance of earlier techniques. Thus, some statisticians have a generalized distrust of this approach. Because of its relative novelty, it is difficult to find statisticians with significant expertise in CART. Thus, it may be difficult to find someone to help you use CART analysis at your own institution .Because CART is not a standard analysis technique, it is not included in many major statistical software packages (e.g., SAS).

This paper extracts land use/cover information using texture features that were derived from the grey-level occurrence matrix. Occurrence Measures can output five different texture filters. The occurrence filters available are data range, mean, variance, entropy, and skewness. Occurrence measures use the number of occurrences of each gray level within the processing window for the texture calculations. In this paper, 3×3, 5×5, 7×7, 9×9, 11×11 processing windows size were selected. In every processing window, all 4 bands can render 20 layers gray level images (one band has 5 layers). Adding the original 4 bands, Total 104 image layers were used in classification.

In this paper, CART (Classification and Regression Tree) algorithm was used. CART was suggested by Breiman et al. in 1984 (Brieman et al., 1984). The decision trees produced by CART are strictly binary, containing exactly two branches for each decision node. It recursively partitions the records in the training data set into subsets of records with similar values for the target (Steinberg et al., 1997, Manoj Kumar et al., 2002, Bittencourt et al., 2003). CART is able to search all possible variables as splitters, and it is inherently non-parametric ,the non-parametric property means that non-normal, non-homogenous and noisy data sets can be handled, s well as non-linear reations between features and classes.missing values and both numeric and categorical inputs (Friedl et. al., 1997). CART trees are relatively simple for nonstatisticians to interpret. Another advantage of CART analysis is that it is a relatively automatic "machine learning". It analysis has a number of advantages over other classification methods. In this paper, inputting all 104 layers into CART algorithm, the final decision tree is shown in Fig. 2.

4. Results and disscusion

The classification map constructed by CART Decision tree is shown in Fig. 3. In order to verify classification accuracy, the result classified by different classification methods and data were compared. Overall accuracy of original bands using Maximum likelihood, texture features using Maximum likelihood, original bands using CART Decision tree and texture features using CART Decision tree are 93.5%, 97.3%, 92.6% and 98.5% respectively. Furthermore, the CART algorithm is more transparent compared to the other algorithm, because in the former the classification sequence that is followed is controlled by the analyst. Classification and Regression Tree (CART) analysis is a powerful technique with significant potential classification utility. Nontheless, a substantial investment in time and effort is required to use the software, select the correct options, and interpret the result. Nontheless, the use of CART has been increasing and is likely to increase in the future, largely because of the substantial number of important problems for which it is the best available solution. From the Decision tree (Fig. 2), some main results can be concluded:

1) Overall accuracy of texture features that were derived from the grey-level occurrence matrix is higher than the original data. Texture features can be used to improve classification accuracy.

2) Among all occurrence filters included data range, mean, variance, entropy, and skewness, mean is more effective in classification than others.

3) Different processing windows size can enhance different land use/cover information. Band1 when processing windows size is 9×9 or 11×11 can distinguish different land use/cover type.

4) Due to low spatial resolution or other reasons, some band like band4 is not suitable for occurrence filters.

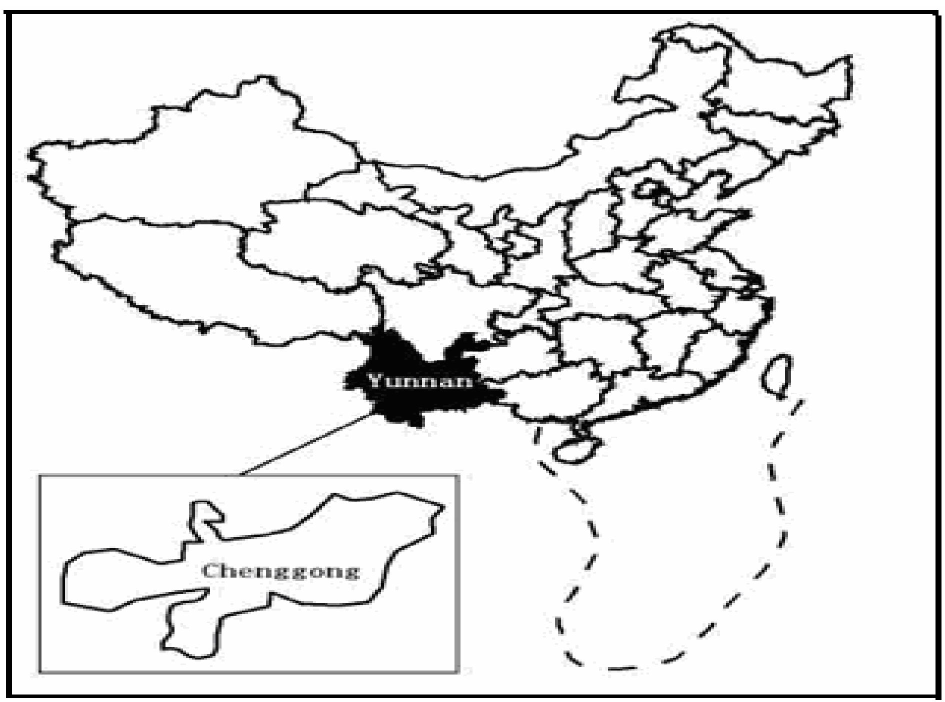

Figure1: The boundary map of China and Chenggong city

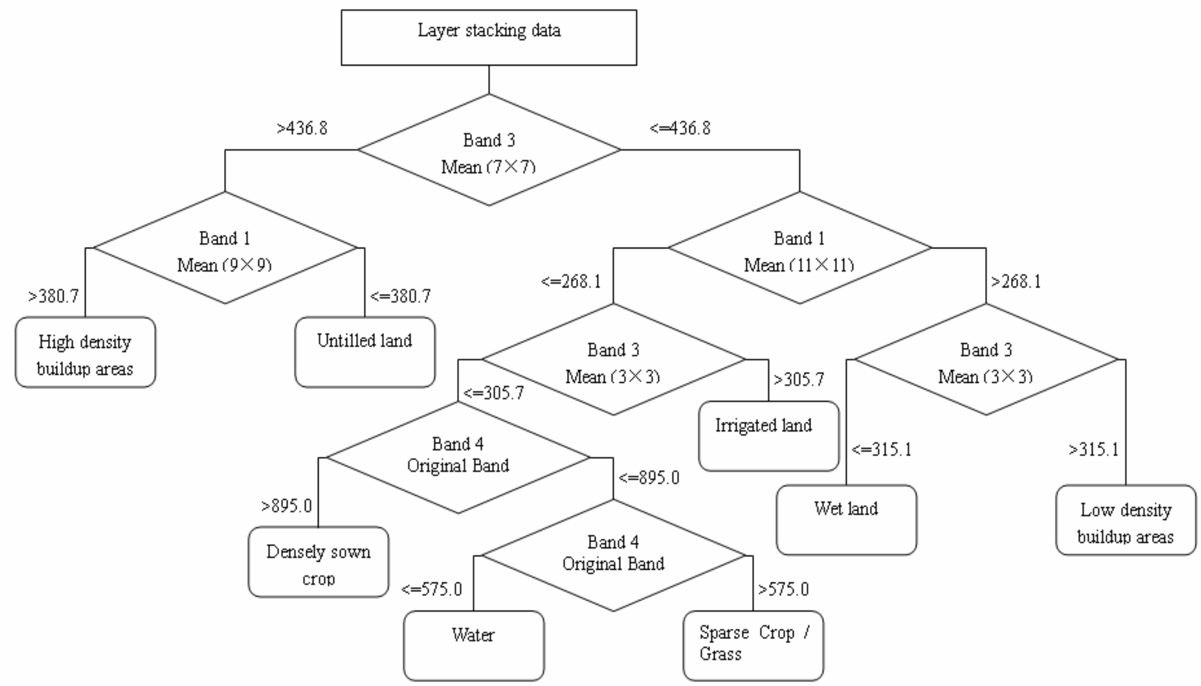

Figure 2: The decision tree constructed by CART algorithm.

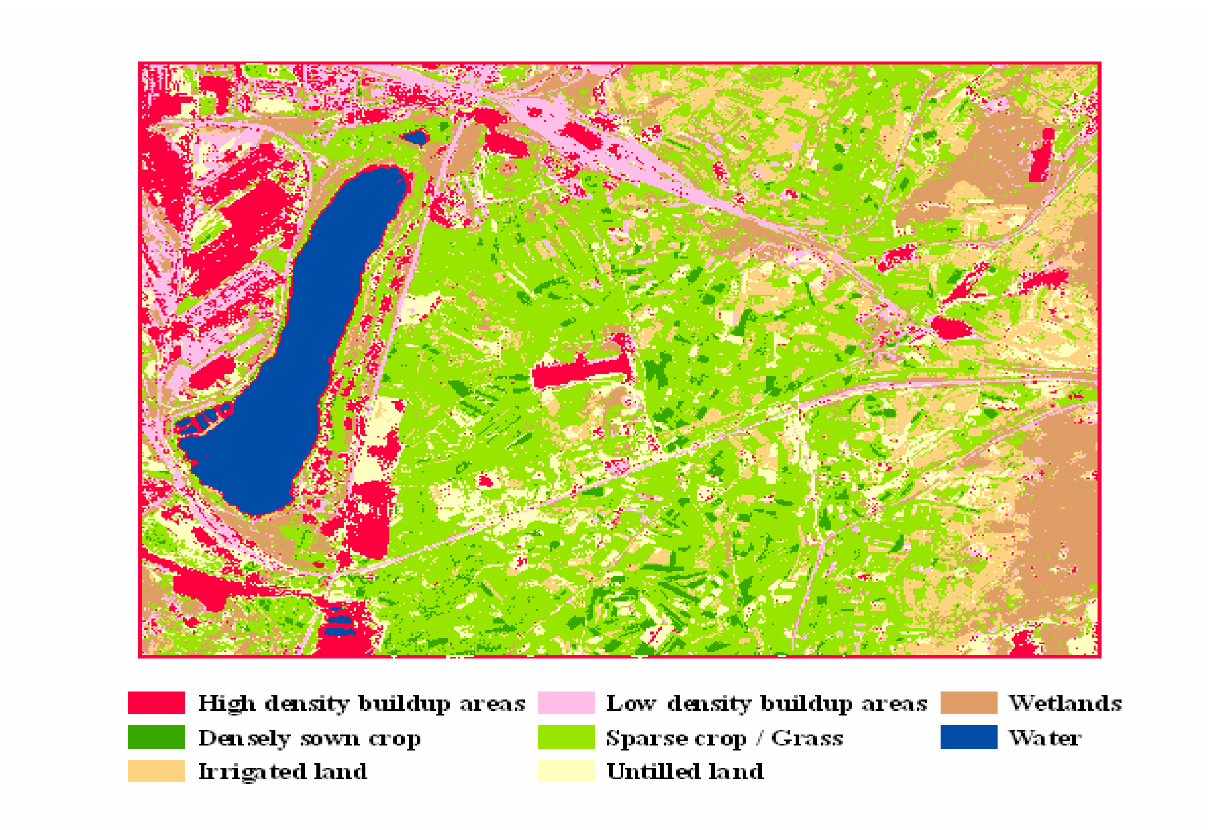

Figure 3: the classification map using CART decision tree.

Acknowledgement:

Many thanks to Professor Hu Guangdao at the Institute of geology and Remote Sensing, faculty of earth resources, China university of geosciences For providing the data .

Correspondence to:
Eltahir Mohamed Elhadi
China University of Geosciences, Faculty of Resources, Wuhan, 430074, China
Telephone: 00862763962740.
Emails: tahirco2006@yahoo.com,
nagizomrawi@yahoo.com

References

[1] Bittencourt, H. R. & Clarke, R. T., 2003. Use of classification and regression trees (CART) to classify remotely-sensed digital images. in *Geoscience and Remote Sensing Symposium, 2003. IGARSS '03. Proceedings. 2003 IEEE International*, 3751-3753 vol.6.

[2] Brieman, L., et al., 1984. *Classification and Regression Trees,* Boca Raton, FL, Chapman & Hall/CRC Press.

[3] Carlson, T. N. & Arthur, S. T., 2000. The impact of land use- land cover changes due to urbanization on surface microclimate and hydrology: A satellite perspective. *Global and Planetary Change,* 25(1): 49-65.

[4] Cushnie, J. L., 1987. The interactive effect of spatial resolution and degree of internal variability within land-cover types on classification accuracies. *International Journal of Remote Sensing,* 8(1): 15-29.

[5] Fan, F., et al., 2007. Land Use and Land Cover Change in Guangzhou, China, from 1998 to 2003, Based on Landsat TM/ETM+ Imagery. *Sensors,* 7: 1323-1342.

[6] Friedl, M.A., Brodley, C.E., 1997. Decision tree classification of land cover from remotely sensed data. Remote Sensing of Environment.61, 399-409.

[7] Gislason, P. O., et al., 2006. Random Forests for land cover classification. *Pattern Recognition Letters,* 27(4): 294-300.

[8]Keuchel, J., et al., 2003. Automatic land cover analysis for Tenerife by supervised classification using remotely sensed data. *Remote Sensing of Environment,* 86(4): 530-541.

[9] Le Hegarat-Mascle, S., et al., 2005. Land cover change detection at coarse spatial scales based on iterative estimation and previous state information. *Remote Sensing of Environment,* 95(4): 464-479.

[10] Manojkumar, P., et al., 2002. A rule-based classifier using Classification and Regression Tree (CART) approach for urban landscape dynamics. in *Geoscience and Remote Sensing Symposium, 2002. IGARSS '02. 2002 IEEE International*, 1193-1194 Vol.2.

[11] Steinberg, D. & COLLA, P., 1997. *CART—Classification and Regression Trees,* San Diego, CA, Salford Systems.

[12] Quinlan, J.R., 1993.C4.5: Programs for Machine Learning .Mogran Kaufmann, California.

9/22/2009