

AN OBJECT-ORIENTED APPROACH FOR AGRICULTURAL LAND CLASSIFICATION USING RAPIDEYE IMAGERY

Huiyong Sang^a, Liang Zhai^a, Jixian Zhang^a, Fang An^a

^a Chinese Academy of Surveying and Mapping, 28 Lianhuachi West Road, Haidian District, Beijing, China –
(huiyong.sang,zhailiang,zhangjx)@cas.m.ac.cn

KEY WORDS: Object-oriented, RapidEye, Red Edge, Agricultural Land, Decision Tree

ABSTRACT:

With the improvement of remote sensing technology, the spatial, structural and texture information of land covers are present clearly in high resolution imagery, which enhances the ability of crop mapping. Since the satellite RapidEye was launched in 2009, high resolution multispectral imagery together with wide red edge band has been utilized in vegetation monitoring. Broad red edge band related vegetation indices improved land use classification and vegetation studies. RapidEye high resolution imagery acquired on May 29 and August 9th of 2012 was used in this study to evaluate the potential of red edge band in agricultural land cover/use mapping using an object-oriented classification approach. A new object-oriented decision tree classifier was introduced in this study to map agricultural lands in the study area. Besides the five bands of RapidEye image, the vegetation indexes derived from spectral bands and the structural and texture features are utilized as inputs for agricultural land cover/use mapping in the study. The optimization of input features for classification by reducing redundant information improves the mapping precision over 9% for AdaTree. WL, and 5% for SVM, the accuracy is over 90% for both approaches. Time phase characteristic is much important in different agricultural lands, and it improves the classification accuracy 7% for AdaTree.WL and 6% for SVM.

1. INTRODUCTION

It is significant to map major agricultural producing areas for high-yield agriculture management and global and national geographical conditions monitoring. Remote sensing imagery is widely applied in land cover information extraction and agricultural production. Traditionally, pixel-based classifiers are dominant in land cover mapping by using low and medium resolution remote sensing images for regional and global monitoring. Among the pixel based algorithms, decision tree classifier is often utilized, which is a classification and prediction model using data mining techniques (Quinlan, 1986). With the development and application of high resolution remote sensing sensors, object-oriented classification algorithms are extensively utilized in regional land cover mapping and precision agriculture monitoring (Myint, 2011; Laliberte, 2004). One single classification algorithm cannot meet the need of quick land cover mapping with high efficiency, by utilizing high resolution remote sensing imagery with a large amount of information including spectral, structural and texture features.

Since the satellite RapidEye was launched in 2009, high resolution multispectral imagery together with wide red edge band has been utilized in vegetation monitoring (Herrmann, 2010; Viña, 2011). Broad red

edge band related vegetation indices improved land use classification and vegetation studies in drought pressure and crop lands mapping (Li, 2009; Schuster, 2012; Eitel, 2011). In this study, an object-oriented decision tree method is proposed for agricultural land cover mapping by utilizing RapidEye multi-spectral imagery.

2. METHODOLOGY

2.1 Study Area and Datasets

The study area is located in central Songnen Plain, north of Songhua River, and east of Harbin city, Heilongjiang province, China (Figure 1). Songnen Plain is one of the major agricultural regions in China, where long dry cold winter and warm rainy summer dominate the temperate continental monsoon climate. Single season corn and paddy rice are local major crops, and both are cultivated in May and harvested in October. Besides crops, the study area also grows cash crops and nursery seedlings. Other land cover and land use types in the study area include small towns, villages, lakes and wetlands. Field surveying was carried out in late July of 2012 to collect in-situ digital photos with GPS coordinates, phenology data and structural parameters of local major crops.

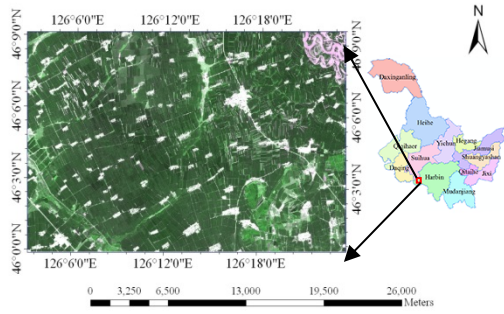


Figure 1. the study area is shown on the left image map, and the right map is Heilongjiang province, China.

RapidEye's Multi-Spectral Imager (MSI) acquires image data in five different spectral bands, including three visible bands, near-infrared (NIR) band, and red edge. Each band has a geometric pixel resolution of 6.5m. RapidEye imagery covering the study area was collected on May 29th and August 9th of 2012, respectively. The subset of RapidEye multi-spectral image acquired in August is shown in Figure 1.

2.2 AdaTree WL. Decision Tree Classifier

Decision tree algorithm is one of the most popular inductive reasoning approaches in data mining field. Any attributes and features conducive to classify different categories could be utilized to construct the decision tree or decision rules. Its interpretive structure and high efficiency in processing multi-dimensional data make this algorithm be widely applied in many research fields, such as medical science, financial analysis, and remote sensing. Decision tree classifier has been successfully applied in regional and national land cover mapping using remote sensing imagery.

A new combined decision tree classifier—AdaTree. Weight Leaf (AdaTree. WL) is proposed in this study. It uses AdaBoost. M1 algorithm to combine a few of single decision trees for the final decision (Freund, 1996). Every single decision tree in the combined decision tree AdaTree.WL has a binary tree structure. In order to construct one single decision tree, information gain ratio is utilized to measure the selection of features in the split node, a depth threshold is predefined, and the algorithm EBP (Error-Based Pruning) is adopted for pruning (Quilan, 1993). The final classification is carried out through utilizing a final hypothesis from the multiple single decision trees. A prediction distribution is defined by the number of training samples, the number of correctly classified samples, and the distribution of that single binary tree (Zhang, 2014). The accuracy of the leaf node l is defined as:

$$accuracy(l) = \frac{n_{correct}}{n_{covers}} \quad (1)$$

where n_{covers} = the number of training samples covered by the leaf node l
 $n_{correct}$ = the number of correctly classified samples.

The prediction weight of the leaf node l is defined:

$$prediction(l) = accuracy(l) \times \log \frac{1}{\beta_l} \quad (2)$$

Where $\beta_l = \varepsilon_l / (1 - \varepsilon_l)$

$$\varepsilon_t = \sum_{i: h_t(x_i) \neq y_i} D_t(i)$$

ε_t = the error of h_t

$D_t(i)$ = the distribution of the example i in the iteration t

h_t = the hypothesis in the iteration t (Freund & Schapire 1996)

The final hypothesis of AdaTree.WL is:

$$h_{fin}(x) = \arg \max_{y \in Y} \sum_{r: h_r(x)=y} prediction(l) \quad (3)$$

2.3 Feature Selection

The pixel based decision tree classification approach is not very suitable for land cover mapping using high resolution remote sensing imagery, especially when the ground objects with obvious spatial characteristics or affected by artificial activities need to be differentiated. An object-oriented decision tree approach is proposed in this paper to map agricultural lands in the study area. This approach has been successfully applied to map urban and rural land covers using high resolution satellite imagery (Boloorani, 2006). The major steps of this approach include image segmentation to get image objects and pixel-based classification based on the image objects. The multiple resolution segmentation algorithm is adopted in this study, which utilizes spatial, spectral and texture features to extract features of image objects. There are 60 feature outputs generated from image segmentation process, including 35 spectral features, 14 structural features, and 9 texture features.

The accuracy of land cover classification is determined by the characteristics of different land covers and the classification algorithm. The characteristics of land covers in remote sensing imagery mainly include spectral features and related vegetation/water/soil indexes from spectral bands. With the widely applications of high resolution remote sensing imagery in terrestrial monitoring, the structural and texture features are utilized in land cover classification, especially in object-oriented classification algorithms. However, excessive features not only slow down the learning process, but cause the classifier to over-fit the training data, as irrelevant or redundant features may confuse the learning algorithm. In this study, the input features are preliminarily selected according to the contribution of each input feature in constructing AdaTree. WL decision tree. The contribution of each input feature is calculated from the number it occurs and its importance in the

tree as a splitter (Yu, 2006). The contribution of all features are ranked from high to low, and the first 35 relevant features with higher contribution are selected for secondary feature filter.

A correlation based method is used as a secondary feature filter to reduce redundant features and get an optimal subset of features for decision tree development (Yu, 2004). Linear and non-linear correlation analysis are two broadly applied for measuring the relationship between two random variables. Linear correlation measures may not be able to capture correlations that are not linear in nature. Among non-linear correlation measures, the information-theoretical concept of entropy is widely utilized. The entropy of a variable X is defined as:

$$H(X) = -\sum P(x_i) \log_2(P(x_i)) \quad (4)$$

The entropy of X after observing values of another variable Y is defined as:

$$H(X|Y) = -\sum_j P(y_j) \sum_i P(x_i|y_j) \log_2(P(x_i|y_j)) \quad (5)$$

Where $P(x_i)$ = the prior probabilities for all values of X

$P(x_i|y_j)$ = the posterior probabilities of X given the values of Y

The amount by which the entropy of X decreases reflects additional information about X provided by Y , $IG(X|Y)$, and is called information gain (Quinlan, 1993).

$$IG(X|Y) = H(X) - H(X|Y) \quad (6)$$

The normalized information gain with the corresponding entropy of X and Y is defined as:

$$SU(X,Y) = 2 \left[\frac{IG(X|Y)}{H(X) + H(Y)} \right] \quad (7)$$

It compensates for information gain's bias toward features with more values and restricts its values to the range[0,1]. A value of 1 indicates that knowing the values of either feature completely predicts the values of the other; a value of 0 indicates that X and Y are independent.

The spectral, structural and texture features are filtered by the process of relevance and redundancy when comprehensively considering the contribution in decision tree development of each feature and the correlation between each other. An optimal feature set (26 features) is determined for building AdaTree WL.

Decision tree, including 19 spectral features, 4 structural features, and 3 texture features.

2.4 Data Analysis and Results

The selected optimal feature set including 26 features in section 2.3 are input to build AdaTree WL decision tree for agricultural land cover mapping. The classification map is shown in Figure 2. The Support Vector Machine (SVM) classifier is compared with AdaTree WL decision tree classifier in this study. The same training samples and feature inputs are utilized in SVM classifier. The overall accuracy is 91.7% with a Kappa coefficient of 0.907 for AdaTree WL decision tree, and 91.3% with a Kappa coefficient of 0.907 for SVM. The entropy-based correlation analysis improves the classification precision about 9% compared with the preliminary contribution-based feature selection for AdaTree WL decision tree, and 5.4% for SVM.



Figure 2. The classification map using AdaTree WL decision tree classifier

It is difficult to map different crop land covers from other vegetated lands due to their similar phenological phase changes in the study area. The red-edge spectral characteristics from various vegetation are distinct and utilized to classify vegetated lands. The potential of red edge band in agricultural lands mapping is evaluated in this study, through comparing the classification results from AdaTree WL Classifier with red edge band included (Option A) or not (Option B). The Option A improves the classification precision 2%, compared with Option B(81.7%). The red edge band make the accuracy increase 7.5% when using SVM classifier, in which the overall accuracy is about 75% for Option B. The classification accuracy of forest, rice paddies, corn, grasslands, fallow lands and bare lands increase 9.25%, 55.3%, 8.5%, 6.2%, 13.3%, 22.3% ,respectively, when red edge band join in building decision tree. The results of artificial lands, water bodies, melon lands and nurseries have no obvious improvement for either classifier. The red edge band is more sensitive to homogeneous vegetation. The heterogeneity of melon lands and nurseries growing diverse vegetation types causes no effect with the joining of red edge band.

3. CONCLUSION

The Object-oriented classification is the main method of high spatial resolution remote sensing classification. Image segmentation and image objects classification are its main steps. This paper proposed a fast method of object-oriented automatic classification and it was an integration of image segmentation and the classifier of AdaTree WL. A two-step feature selection approach is utilized in this study to find relevant feature and remove redundant features to get an optimal feature set for decision tree construction. The experiment with RapidEye image showed that the AdaTree WL classifier outperformed efficiently. Furthermore, the red edge band of RapidEye imagery shows its great potential in agricultural land cover mapping.

ACKNOWLEDGEMENTS

This work was funded by: National Natural Science Foundation of China (41371406), China Geographical Conditions Monitoring Project (B1512), Open research fund of Key Laboratory of Geo-Informatics of National Administration of Surveying, Mapping and Geoinformation (7771514), Basic Research Fund of Chinese Academy of Surveying and Mapping (7771507, 7771509).

REFERENCES

- Boloorani A D, Erasmi S, Kappas M. Urban land cover mapping using object/pixel-based data fusion and ikonos images[J]. Pan, 2006,450:900.
- Eitel, J.U.H., et al., Broadband, red-edge information from satellites improves early stress detection in a New Mexico conifer woodland. *Remote Sensing of Environment*, 2011. 115: p. 3640-3646.
- Freund Y. Schapire R. E. Experiments with a new Boosting Algorithm. *Machine Learning: Proceedings of the Thirteenth International Conference*, 1996
- Herrmann, I., et al. Utilizing the vensur red-edge bands for assessing LAI in crop fields. in *Core Spatial Databases - Updating, Maintenance and Services - from Theory to Practice*. 2010. Haifa, Israel.
- Laliberte A S, Rango, Havstad K M, et al. Object-oriented image analysis for mapping shrub encroachment from 1937 to 2003 in southern New Mexico. *Remote Sensing of Environment*, 2004, 93(1/2):198-210.
- Li, F., et al. Evaluating red edge vegetation indices for estimating winter wheat N status under high canopy coverage condition. in *the International Plant Nutrition Colloquium XVI*. 2009. UC Davis.
- Myint S W, Gober P. Per-pixel vs. object-based classification of urban land cover extraction using high spatial resolution imagery. *Remote Sensing of Environment*, 2011,115: 1145-1161.
- Quinlan J R. Induction of decision trees. *Machine Learning*, 1986(4):81-106.
- Quinlan J R. C4.5: Programs for machine learning. Morgan Kaufman, 1993:81-106.
- Schuster, C., M. Forester, and B. Kleinschmit, Testing the red edge channel for improving land-use classifications based on high-resolution multi-spectral satellite data. *international Journal of Remote Sensing*, 2012. 33(17): p. 5583–5599.
- Viña, A., et al., Comparison of different vegetation indices for the remote assessment of green leaf area index of crops. *Remote Sensing of Environment*, 2011. 115: p. 3648-3478.
- Yu Q, Gong P, Clinton N, et al. Object-based detailed vegetation classification with airborne high spatial resolution remote sensing imagery[J]. *Photogrammetric engineering and remote sensing*, 2006,72(7):799.
- Yu L, Liu H. Efficient feature selection via analysis of relevance and redundancy[J]. *The Journal of Machine Learning Research*, 2004,5:1205-1224.
- Zhang X. Implement of decision tree classifier and application in remote sensing image classification, 2014, Thesis of Master degree, Lanzhou Jiaotong University