CLASSIFICATION OF COASTAL ZONE BASED ON DECISION TREE AND PPI

Shanshan Li, Bing Zhang, Lianru Gao, Liang Zhang
E-mail: white_snowbird@163.com Tel: +8613720034690

Center for Earth Observation and Digital Earth, Chinese Academy of Sciences

1. INTRODUCTION

In recent years, hyperspectral remote sensing images have been widely used in applications of geology, agriculture, forest, ocean, and etc. This paper reports aimed at assessing the suitability of hyperspectral data for coastal zone classification. The coastal zone is a complex space where terrestrial environments and marine environments influence each other, including various coast flats and many artificial objects. Traditional statistical classification methods, such as k-means, minimum distance and maximum likelihood algorithm, cannot differentiate classes with similar spectral feature. In this paper, an integrated approach of decision tree and endmember extraction is proposed to classify the coastal areas。

2. METHOD

2.1 Decision Tree Classifier

Classifiers structured as decision trees are hierarchical classifiers. A decision tree provides a representation of feature space in which patterns x_i are allocated to classes w_j (j=1,2,...,k) according to the result obtained by following decisions made at a sequence of nodes at which branches of the tree diverge[1].

The characteristics of the data used as sample of supervised classification have a considerable influence on the quality of classification result. It is essential that the training data provide a representative description of each class. So choice of training data will affect the result and accuracy of the decision tree directly.

2.2 Endmember Extraction with PPI

Boardman developed Pixel Purity Index technique to extract image using convex geometry with Kruse and Green [2].It is designed to locate the most spectrally extreme pixels, which typically correspond to mixing endmembers. The extreme pixels in each projection are recorded and the total number of times each pixel is marked as extreme is noted. A Pure Pixel Index (PPI) image is created in which the digital number of each pixel corresponds to the number of times that pixel was recorded as extreme. A histogram of these images shows the distribution of “hits” by PPI. Pixels with higher PPI include more features of some class in comparison with other similar subsets.

2.3 Decision Tree Classifier Based on PPI

In the case of using decision tree to perform classification for hyperspectral image, it is always assumed that features are

extracted perfectly and efficiently. So we present a decision tree based on PPI. Firstly, sorting PPI of whole image, and then, choosing training samples according to their PPI. Compared with traditional ROI choice by visual interpreting, it is more efficient and statistical. Finally, we adopted C4.5 algorithm to classify hyperspectral image. C4.5 (Quinlan, 1993) is an extension of ID3 algorithm. To overcome Information Gain which always tends to select attributes that have a large number of values since the gain of such an attribute would be maximum. Quinlan (1993) suggested the use of Gain Ratio as a measure to select the splitting attribute instead of Information Gain [3].

3. EXPERIMENT RESULT AND DISCUSSION

By using the decision tree classification described above, the image was classified into 10 classes. From the result image we can see that water body was classified well with little miss classification. According to the reflectance of spectral features, though curve of rock shock is close to vegetable, and curve of sand beach is close to the artificial objects. After classification as Fig.2, a main road was clear in the image, with its boundary full of pixels of vegetation, buildings and others that were unclassified. The Kappa coefficient and accuracy of Confusion Matrix have been largely improved in comparison with traditional statistical method.

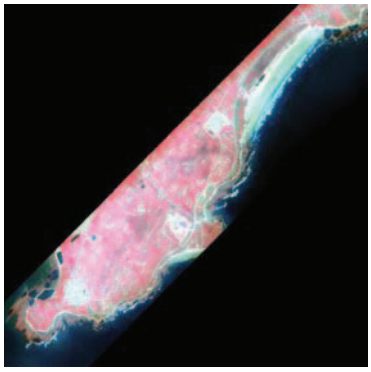

Fig.1 False color image of coastal zone

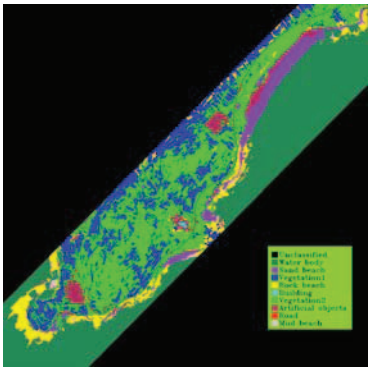

Fig.2 Classification result

4. CONCLUSION

We proffer a hybrid decision tree classification algorithm based on PPI, optimum spectral features of class pairs are used on every tree node.The coastal image is classified into several classes, and this algorithm can classify several general classes quickly without assumptions about distributions of input data. Furthermore, the classification flow shows the classification structure explicitly and easily interpretable, it is applicable for hyperspectral image of coastal zone classification.

REFERENCES

[1] Hélio Radke, etc, “A Binary Decision Tree Classifier Implementing Logistic Regression as a Feature Selection and Classification Method and its Comparison with Maximum Likelihood”, Geoscience and Remote Sensing Symposium, IGARSS 2007. IEEE International, pp.1755-1758, Sept.2007.

[2] Qinxi Tong, Bingzhang, etc, “Hyperspectral Remote Sensing”. High Education Press. Beijing. pp.262-263, Jun.2006.

[3] B. Chandra, “A Robust Algorithm for Classification Using Decision Trees”, Cybernetics and Intelligent Systems, 2006 IEEE Conference, Digital Object Identifier, pp.1-5, Jun.2006.