PROCEEDINGS OF SCIENCE

Decision Tree Algorithm Based on Regional Growth for the Automatic Oil Field Road Extraction

Wenhui Li[1]
Department of Computer Science and Technology Jilin University , Changchun, 130012, China

Yunfan Du, Huiying Li[2]**, Xuezhi Wang**[3]**, Jinlong Zhu**
Department of Computer Science and Technology Jilin University ,Changchun, 130012, China
E-mail: lihuiying@jlu.edu.cn

The road extraction of remote sensing image has always been a hot topic in the information extraction field. It is not only a great theoretical value, but also has a broad prospect. There are algorithms to extract hierarchical road features from high-resolution aerial images of oilfield; but these studies indicate that data extraction is harder to satisfy the needs of practical production in oilfield. In this paper, we proposed an oilfield road feature extraction method based on the regional growth algorithm and CART(Classification And Regression Tree) decision tree algorithm. The proposed method may reduce the complexity of the extracted sample data by feature classification in order to identify the characteristics of the road. The experimental results showed that the method worked properly.

CENet2015
12-13 September 2015
Shanghai, China

[1]Speaker
[2]Corresponding Author
[3]The work described in this paper was funded by Natural Science Research Foundation of Jilin Province of China (20140520071JH, 20120305). The Satellite project of National Development and Reform Commission- [2013] No.2140: "Demonstration of Integrated Application of Satellite Technology to the Oilfield Exploration and Development".

PoS(CENet2015)006

1. Introduction

The image feature extraction is the extraction of useful information from the image [1, 2, 3]. The road characteristics are different in road extraction [4, 5, 6]. In this paper, according to radiation characteristics, we proposed the regional growth algorithm to select feature points and CART algorithm to get the rule to extract road network from image. Compared with the extraction of the city road, the extraction from oil field is easier because there are fewer interventions including houses, bridges, trees and other vegetation etc. during extraction [7]. This advantage also ensures the accuracy of regional growth to collect the positive sample point.

The structure of this paper is shown as follows: the second part introduces the regional growth and the decision tree of the two main algorithms, and explains the procedure of algorithm as we proposed. The regional growth extracts positive samples as training samples. The decision tree learns training samples and produces a binary tree. The binary tree as we generated is to classify the image. Finally, we draw the conclusion.

2. Algorithm Description

2.1 Regional Growth

The regional growth refers to a region-based image segmentation method. It is also classified as a pixel-based image segmentation method since it involves the selection of initial seed points. This approach to segmentation examines the neighboring pixels of initial seed points and determines whether the pixel neighbors should be added to the region. The process is iterated on in the same manner as the general data clustering algorithms. Fig. 1 shows the flowchart of the regional growth.

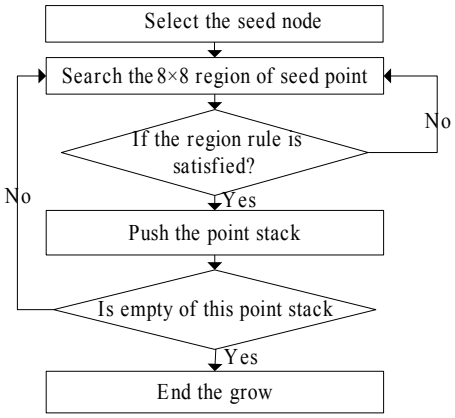


Figure 1: Flowchart of Regional Growth

2 .2 Decision Tree based on Regional Growth

The tree CART (Classification and Regression Tree) is a supervised classification method, which uses the training samples to construct a binary tree. CART algorithm uses the Gini index as a way to divide property. The Gini index is mainly to measure the impurity of divided data. The value of Gini is smaller and the purity of the samples is higher. When the value of Gini_Gain is the smallest, the classification is the best.

For a data set T, its Gini index is calculated as follows:

$$gini(T)=1-\sum_{j=1}^{n} p_j^2 \tag{2.1}$$

Where T: the one category of all samples, means the split node of CART, n: the number of decision type of category T, p: the decision probability of the same category T. P is obtained by using the CART algorithm. It is statistical results. In the process of training for sample in CART, n is set as 2.2.

When we count all the GINI index of a feature value, the Decision Tree can get the Gini Split Info.

$$Gini_{split}(T)=\sum \frac{N_i}{N} gini(T_i) \tag{2.2}$$

Here, I is the i-th value of feature

For CART, i = (1, 2). We get the Gini Split Info (Gini_gain) as follows:

$$Gini_{split}(T)=\frac{N_1}{N} gini(T_1)+\frac{N_2}{N} gini(T_2) \tag{2.3}$$

Where N: the number of T; N_i: the number of i-th category in T.

If we want a more precise classification, the more samples have to be collected; but the extraction of positive samples (it means the point representing road in image) manually is slowly and imprecisely. Using the road self-features to extraction can reduce greatly the complexity of the sample.

Road features are summarized as radiation characteristics, spectral characteristics, geometry, texture and other relevant characteristics [8, 9].

(1) Geometric features: the road shows double-edge which is parallel to each other. The width of the road is essentially unchanged. The degree of road curved is in the appropriate range. The shape of the intersection is usually cross or T-shaped.

(2) Radiation characteristics: the radiation characteristics of road mean it have two obvious edges, and there is a sharp contrast between the interior gray and the adjacent region gray. According to this feature, the road can be distinguished from other land features easily.

(3) Topological features: the topological features mainly refer to the network structure of road. As the roads are connected and will not be suddenly interrupted, they form the network. In the remote sensing image, the road network is very clear; therefore, we can extract the road network by means of this feature.

(4) Context features: the context refers to the image features of land object and background with respect to roads, such as buildings and street trees along the roads, whether the urban roads or rural roads. For example, the roads of wilderness, both rural and urban, will be of various widths, densities and bending extent. All of these we call the context features.[10].

Because of the radiation characteristics of the road, it is a reasonable way to use the regional growth to obtain points, which are similar to the seed point we set. When we get a sufficient number of the positive samples, we can use decision tree to train these positive sample data.

The training process is to calculate the gini gain of positive and negative samples, then, choose the smallest gini gain as the split criterion, and deduce the rules. The process stops when one of the following conditions is met:

1. The sample belongs to the same class
2. The decision tree height reaches the threshold value set by the user.

Then, when the decision tree is generated, we can get the whole classification rule. The process of algorithmic is shown in Fig. 2:

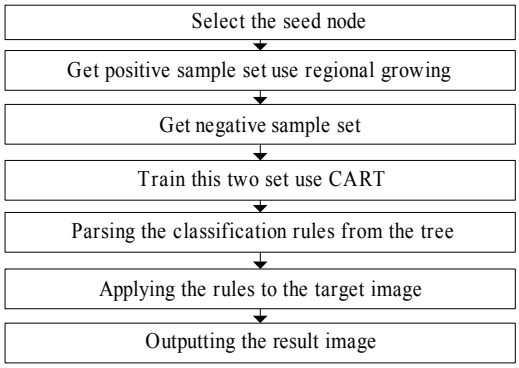
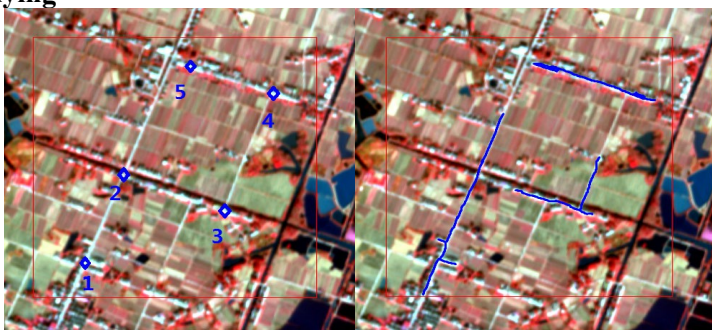


Figure 2: Flowchart of Road Extraction

3. Experiment Studying



(a) (b)

Figure 3: Seed Points and Result of Growth

In the experiment, the remote sensing image is the left of Fig. 3, the size of image is 500×500. This image has four bands. In different bands, the feature will have different reflectivities; therefore, we use the band value as the basis to classify the feature. As Table 3 shows, the features of this image have land, lake, road, grass, and other interferences and all of them are displayed as different colors.

Items	Band 1	Band 2	Band 3	Band 4
1	217	145	152	52
2	227	204	202	75
3	245	223	224	89
4	244	227	237	120
5	238	214	217	157

Table 1: Number of Seeds and Amount of Positive Samples

Items	Point number	Threshold
1	512	10
2	520	10
3	480	10
4	285	15
5	278	15

Table 2: Number of Seeds and Amount of Positive Samples

Items Color	Items
Red	land
Blue	lake
White	road
Green	grass
Others	interferences

Table 3: Image Specifications

Firstly, we use the regional growth to get positive samples. According to the image, we can find the road lines, then, try to select five points of road lines and mark them manually (you can choose more points as the seed point, the more the seed points you set, the more positive samples you'll get. Then, you can get more precise result). As shown in Fig. 3, the left is the image to be processed with five seed points. The band value in these five points is shown in Table 1. Use the regional growth algorithm to these five seed points and we will get a lot of positive sample data as shown in Table 2. The right column of Fig. 3 refers to the results of growth. When setting the point 1, 2 and 3 as seed point, it is found through experiments that we can get the best growth results when we set the threshold value as 10. And, the point 4 and 5, we set the threshold value as 15 can get the best result.

When get the positive sample set, and then select some negative samples randomly. Through the training of the positive and negative samples by Cart, we get the binary tree as shown in Fig. 4. This tree represents a rule which can extract the road from the remote sensing image. In this tree, the rule is that the fourth band value is greater than 117 and the first band value is less than 221, which means if these two bands belong to this range, it will be considered as road while others are will be treated as noise.

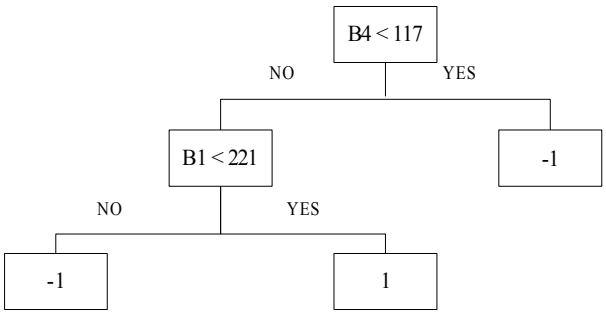
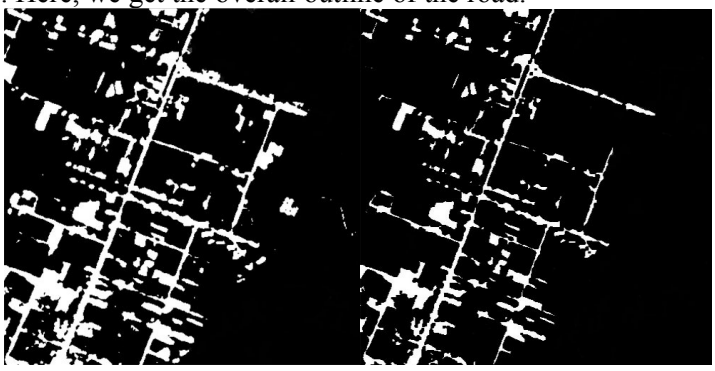


Figure 4 : Tree Generated by Training

Then what we do next is to apply this rule to each pixel of the image. If the band value of the pixel conformed to the rules, we marked this pixel with white color; otherwise, we marked it with black color. In Fig. 5, we found that through our rule, the basic framework of the road has been extracted in the entire image; however, there were small parts of the noise attendant. Then, we can do some processing, such as the corrosion operation of image. In this way, we get the right image of Fig. 5. Here, we get the overall outline of the road.



(a) (b)

Figure 5: Result as Regularized

Tree Depth	Extraction accuracy	Extraction noise ratio	Train time(s)	Extraction time(s)
2	82%	8%	2.7	18.6
3	89%	10%	3.1	23.7
4	91%	15%	3.8	30.5

Table 4: Evaluation Table of Road Extraction

Table 4 has proved that the deeper the decision tree is the higher accuracy of extraction will be; however, more noise will be extracted. In this sense, considering all indicators as to this image, under the premise of the same positive and negative sample sets, when the depth of tree is seated as 3, the extraction result is best.

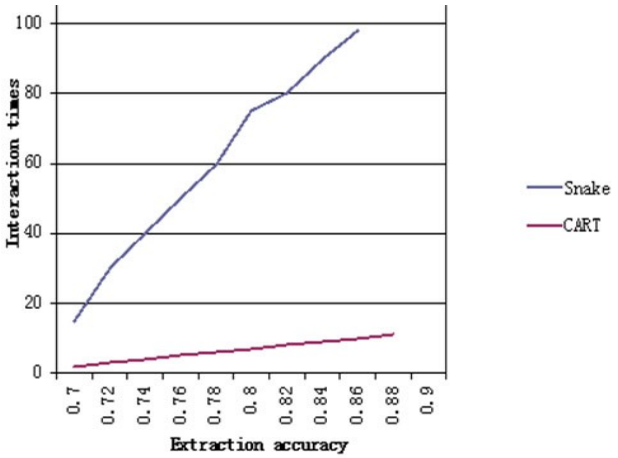
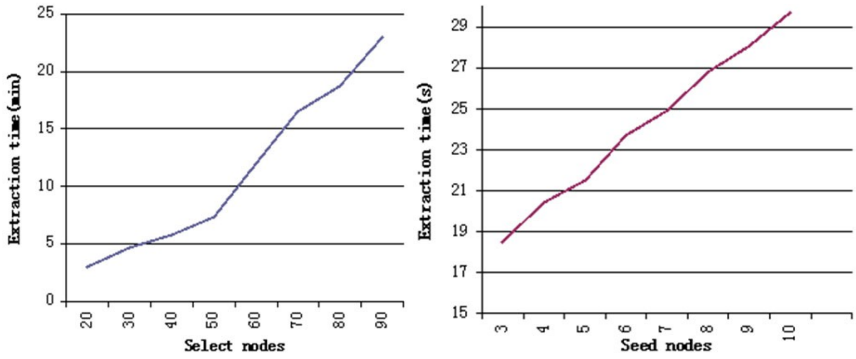


Figure 6: Relationship between Extraction Accuracy and Interaction Times of Snake and CART

At present, the snake algorithm has been widely recognized to extract the road; but some of artificial intervention is necessary because of the inherent limitations of the said snake algorithm. Fig. 6 and Fig. 7 show the comparison of Snake and CART (the depth of CART all is 3). In Fig. 6, the more points were chosen manually, the higher accuracy of the road extraction would be. Fig. 7 show both the number of points selected and road extraction time, the CART is better than Snake.



(a) Snake (b) CART

Figure 7: Relationship between Number of Nodes Selection and Time to the Road Extraction of Snake and CART

4. Conclusion

This article has developed an automatic algorithm to extract the road based on the decision tree. The proposed method can reduce the interactions while compared to the snake etc. Experiment shown in the fourth part can prove that this method may improve the performance of road extraction. In order to reduce the complexity and improve the accuracy of sample collection, a sample extraction method regional growth is hereby proposed to get the positive samples in this paper. We use Cart to train the samples gathered by regional growth and get an extraction rule. With this rule, the road network is generated. Through theoretical verification, the network extracted conforms to the actual road network generally.

References

[1] P. Doucette, P. Agouris, A. Stefanidis, and M. Musavi. *Self-organised clustering for road extraction in classified imagery*. ISPRS Journal of Photogrammetry and Remote Sensing. 55:347–358(2001).

[2] P.S. Tiwari, H. Pande and M.N. Aye. *Exploiting IKONOS and Hyperion data fusion for automated road extraction*. Geocarto International. 25(2):123-132(2010).

[3] M.-F. Auclair-Fortier, D. Ziou, C. Armenakis, S. Wang. *Survey of Work on Road Extraction in Aerial and Satellite Images*. Technical Report. 247(2000).

[4] L. Ma, J. Chen. *Classification and application of context information in road extraction.* Geomatics World. 6(4): 58-60(2008).(In Chinese)

[5] Q.P. Zhang and I. Couloigner. *Automated road network extraction from high resolution multi-spectral imagery.* ASPRS Annual Conf., Reno, Nevada, 10 p., CD(2006).

[6] W. Li. *Road extraction from remote sensing images.* Automation Panorama. 23(5): 20-23(2006). (In Chinese)

[7] J. B. Mena, J. A. Malpica. *An automatic method for road extraction in rural and semi-urban areas starting from high resolution satellite imagery.* Pattern Recogn. Lett. 26(9):1201-1220(2005).

[8] G.Y. Li, Y. Hu. *Road feature extraction from high resolution remote sensing images: review and prospects.* Remote Sensing Information, 2008(1): 91-95(2008). (In Chinese)

[9] W.Z. Shi, C.Q. Zhu, Y. Wang. *Road feature extraction from remotely sensed image: review and prospects.* Acta Geodaetica Et Cartographica Sinica. 30(3): 257-262(2001).(In Chinese)

[10]C. Heipke, H. Mayer, C. Wiedemann. *Evaluation of automatic road extraction.*International Archives of Photogrammetry and Remote Sensing. 32(1): 47-56(1997).(In Chinese)