An automatic method to extract Populus Euphratica forest in a large area using remote sensing

Y.Y. Wang

School of computer science and technology, Chongqing University of Posts And Telecommunications Chongqing 400065, China

State Key Laboratory of Remote Sensing Science, Jointly Sponsored by the Institute of Remote Sensing Applications of Chinese Academy of Sciences and Beijing Normal University, Beijing 100101, China

B. Zhong

State Key Laboratory of Remote Sensing Science, Jointly Sponsored by the Institute of Remote Sensing Applications of Chinese Academy of Sciences and Beijing Normal University, Beijing 100101, China

F.J. Shang

School of computer science and technology, Chongqing University of Posts And Telecommunications Chongqing 400065, China

ABSTRACT: Populus Euphratica is a haloduric desert vegetation growing in arid regions. It is a drought-enduring plant and it is also a wind barrier to fix sands in desert areas. The distribution of Populus Euphratica forest is required to carry out the management of water resources in the arid region, such characteristics make it play an important role in maintaining the hydrological ecological balance in desert. However, they are distributed discretely and the single tree is very small, so it is difficult or even impossible to be extracted at a large scale based by using moderate to high spatial resolution remote sensing data. Thus, the high spatial resolution remote sensing data are required for extracting Populus Euphratica. However, the utilization of high and very high spatial resolution remote sensing data in a large area is usually not implementable because of the high costs and low processing speed. In addition, the manual procedure is usually incorporated, which is further lower its implementalbilty. Therefore, there is hardly any related research for extracting Populus Euphratica in a large area. In this context, this paper proposes an automatic method to extract the populus euphratica forest in lower-stream of the Heihe River (total area of approximately 21646.6km^2) by using object-oriented classification method. Firstly, high spatial resolution data (higher than 2 m) are extracted from Google Earth (the data are used for free and the high level product based on Google Earth images don't have copyright issues). The extracted images are mosaicked automatically using the application programming interface provided by Google Earth. Thirdly, based on analyzing the characteristics of Populus Euphratica as image objects, a set of rules for extracting Populus Euphratica by employing object oriented method are constructed. Finally, the manual inspection method is employed to verify the accuracy of the extracting results and it shows an accuracy better than 87%. The proposed method is capable of extracting populus euphratica forest using Google Earth automatically with low-cost and high-precision and it will become a feasible technical solution to extract thematic information automatically with low-cost and high-precision. Moreover, it will lead a large amount of applications, which are able to provide high-precision and high-resolution thematic products at a very low cost for the sic-economic development in the future; therefore, it has great values on remote sensing applications.

KEYWORD: Google Earth; Populus Euphratica Forest; Object-oriented Classification; Automation; High spatial resolution

1 INTRODUCTION

Populus Euphratica is a desert vegetation growing in arid regions. It can withstand drier and is a saline tolerance plant. Such characteristics cause it to play a important roles in maintaining the ecological balance of Heihe River hydrological (Zhou et al, 2015). The Heihe River is the second largest inland river in the arid region of northwestern China where is dry that annual average precipitation is less than 45mm (Li et al, 2012), and that over-exploitation of river basin having been caused the natural oasis atrophy and expansion of downstream areas of land desertification (Zhang et al, 2004). So, studying distribution of populus euphratica in lower-stream of Heihe River is important in that it not only protect and restore fragile ecosystems downstream environment, but also is conducive to the development of hydrological and ecological construction in the basin.

"An integrated remote sensing experiment on hydrological and ecological processes in the Heihe River Basin" is one of the major projects supported by Chinese Academy of Sciences Action Plan for

West Development Project. The main goal of the test is getting to know the mechanism of aquatic ecosystems of Heihe River and its impacts on the ecology and climate change. The test can also provide basic theoretical and technical support for hydrological ecological security and improve application abilities of remote sensing in the integrated research of Eco-hydrology and water resources management to achieve sustainable development of inland rivers (Zhong et al., 2015). The downstream natural oasis of Heihe River is one of the test area, because downstream belongs to extreme arid climate zones, and has a natural oasis ecosystem. Scarce precipitation and large evaporation make the region structure is simple and fragile (Li et al, 2012). Therefore, researching the distribution of populous has important meaning in ecological protection, water resources management and regional sustainable development.

In 2005, Google launched a share of multi-source remote sensing image software, Google Earth. This platform not only provides licensed global satellite and aerial imagery, but also updates frequently. Google Earth has been applied in many fields, since its remote sensing images which contain major cities and most attractions in various scales, the accuracy even up to 1m (Yang et al., 2008). Populus, discretely distributed in the lower reaches of Heihe River, can be visually extracted by Google Earth. Therefore, this paper proposes an approach by combining free high-resolution images on Google Earth and object-oriented classification method to analyze the characteristics of Populus in the Heihe River lower-stream and extract it.

1.1 *Related Work*

Discrete distribution and small size are the spatial structure of Populus in the Heihe River lower-stream. Currently, the land cover classification of the lower reaches of Heihe River mainly base on remote sensing images in medium-resolution and low-resolution, such as HJ-1 / CCD, Landsat / TM, MODIS and so on. But, since there is a large number of mixed picture elements in the image, the classification which bases on picture elements will cause misclassification (Zhong et al, 2014). However, high-resolution images are suitable for the extraction of small surface features, such as Populus which is rarely extracted in classification products of Heihe River,because of a small amount of mixed pixels. Meanwhile, there is less work on the extraction of the populus in large area (Zhong et al., 2015).In the article" The Extraction of Populus-crown in Ejin based on Quick-bird",the researcher extracted the populus by QUICKBIRD, which is expensive. And it's not reasonable for extracting in large area (Shi et al, 2011).

1.2 *Contributions and Organizations*

The contributions of this paper are as follows:

(1) Our method achieves objectives of mission and radically reduces the cost of paying for data. (2) Our method can simple the processing in extracting, stitching and recognizing data of populus. And the above processing is automated and efficient. (3) We make experiment to compute precision of our method. Also, the experiment shows high precision and high efficient of our method.

In Section 2, we describe data acquisition and processing, and extracting method is described in Section 3. The experiment is shown in Section 4. Finally, we state the conclusions and acknowledgement in Section 5 and Section 6, respectively.

2 DATA ACQUISITION AND PROCESSING

2.1 *Selecting Training Area*

In this study, we select training area from lower-stream of Heihe River. Firstly, we need to determine a suitable spatial scale. The rules are as follows:

(1) Scale should not be too low in order to distinguish the Populus Euphratica in images.

(2) Scale should not be too high to increase unnecessary cost. According to these rules, we determined 1.5m scales for the researching.

Secondly, the requirement for selecting training are as follows:

(1) Most of Populus Euphratica grow in this area;

(2) In this area, Google Earth is able to provide high-resolution images;

(3) It contains more than one plant species. So it can represent the lowerstream of Heihe River.

Following these rules, we chose this part of the lower-stream. (41.0458^0-41.1508^0N, 100.229^0-100.439^0 E). The total area is about 1125.07 km^2.

2.2 *Experimental Data Acquisition*

Google earth provides interfaces for developers to clip image. The procedure can position to specified coordinate in screen windows according to the center longitude and latitude and altitud. Then it clip images and save screenshot. The interception preserved. 1.5m spatial scale corresponds to 2 km altitude by computing. In this study, according to the size of testing computer and screen resolution, the maximum size of each shot can be set as 1660 * 815. Due to the training area is larger than one image. A piece of Google earth image is difficult to cover the region completely. So, we need to get 527 images of training area. In order to improve the efficiency as well as to ensure the accuracy of shots, we used SetCamera-Params function provided by Google earth to clipped image automatically from the region's lower-left to the upper-right. In the end, sreenshots were saved as

BMP format. Each image includes R, G, B, Alpha band.

2.3 *Experiment Data Processing*

The picture clipped from the Google earth only include the information of RGB. In other word, the image doesn't contain geographic information. For stitching image and accuracy verification. We need to write affine transformation parameters and geographic projection for each image. The calculation process is as follows:

(1) The determination of projection. We used the same projection as the Google earth which is Geographic lat/lon and WGS84.

(2) The determination of lateral resolution and vertical resolution. Each image is known longitude, latitude of center point and image's size, so we can select respectively four points from each sides in the image. Then we get its latitude and longitude according to Google earth. Lateral resolution calculated by two points' longitude difference in the x direction and width of the image. And also, we can calculate the vertical resolution in the y direction.

(3) Calculating latitude and longitude of the upper-left of the image. Having known lateral and vertical resolution, we can calculate the latitude and longitude of the upper-left with the image size and longitude, latitude of center point.

And so on, all of the original image's corresponding affine parameters and image projection information can be obtained by this way .Then all of the images automatically contain these information and were transformed as GeoTiff pictures which contain 3 bands.

For the research, it is necessary to mosaicking the images for the whole area (Fig. 1).

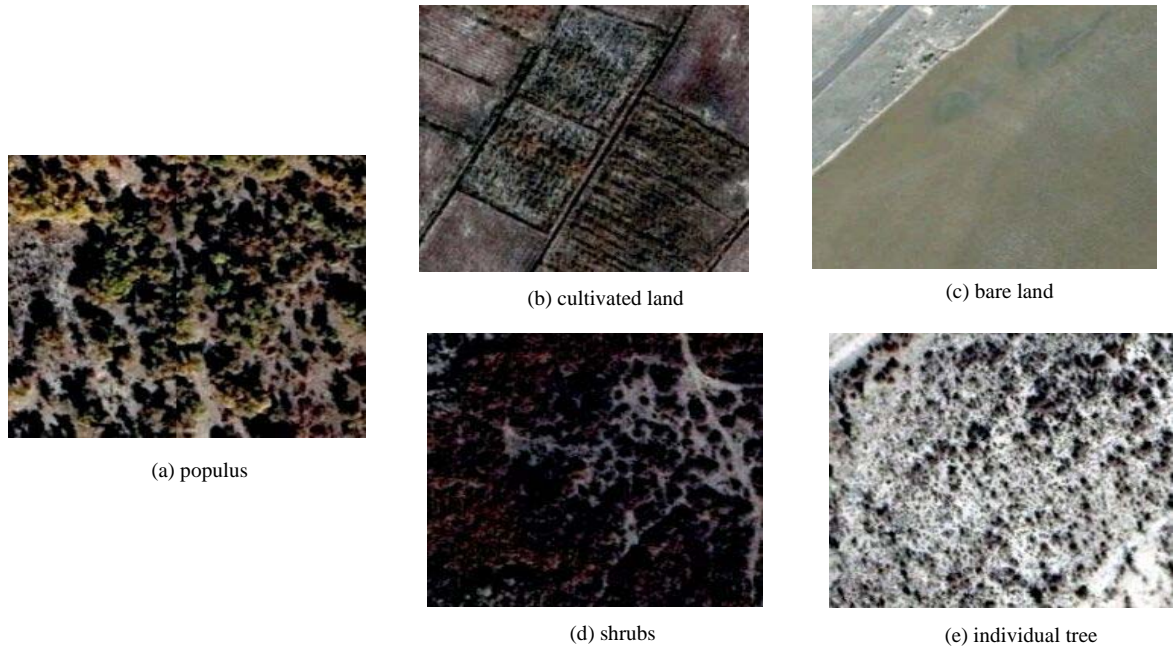Fig. 1   An example of mosaic images of the training area

(b) cultivated land

(c) bare land

(a) populus

(d) shrubs

(e) individual tree

Fig.2 The comparison diagram of euphratica in relation to farmland, bare land, shrubs and other vegetation

3 EXTRACTING METHOD

3.1 *The Analysis of Populus Euphratica in Heihe River Lowerstream*

Populus usually grows around river basin, and also grow alone in the drought places. Except populus, the cultivated land, lake and shrubs grow around these area (Tamarix mainly). Figure 2 shows the differences of spatial structure characteristic between these major feature and populus, which are summarized as follows:

(1) Arable land is different with populus (Fig. 2a). Its shape is regular, which is similar to rectangle (Fig. 2b). And also the cultivated land distributed closely;

(2) The populous' texture is very small and dense, But the texture of bare land and lake (Fig. 2c) is smooth in contrast;

(3) Poplar distribute discretely in the lower basin. But shrubs (Fig. 2d) are growth into pieces, which cover a wide area. And compared to the populus, their height is much lower. So, there are some differences in texture.

3.2 *Extracting Populus Euphratica Forest by Object-oriented Classification Method*

Object-oriented classification method is a new method of feature extraction, different with the conventional pixel classification methods, it also combines the structural characteristics of spatial objects textures, shapes and other surface features to identify object. And the target is composed of a plurality of pixels, rather than a single pixel (Xu et al., 2015). Image segmentation is the basis of object-oriented classification, which divide the image into different areas according to different features. So analysing the characteristic of an object is completed based on image segmentation. (Tan and Gao, 2011).
Further, since its precipitation is uneven in the lower reaches of Heihe River, resulting in complex and diverse types of vegetation and irregular distribution. Excluding all interferential objects in one-time will cause leaking most of populus points and misclassification. Therefore, the reasonable approach is to eliminate other objects step by step. Firstly, analysis the populus' characteristics with each other objects based on segmentation, and then set different rules to eliminate them. In this way, it can avoid interference between each object. Finally, we build a suitable extraction ruleset.

Currently, the image segmentation algorithm based on edge-based and region-based (Guan et al., 2010). In Heihe River lower-stream, the populus, shrubs, rivers and bare land have clear boundaries. So we decided to use the edge-based segmentation algorithm. During image segmentation, if the threshold is too low, some features of objects will be misclassified, we can merge images combination which is called Full Lambda-Schedule algorithm to solve this problem. If the threshold of segmentation and merging is reasonable, it can be better to express each objects. Through several sample training, we concluded the reasonable split and merge threshold.

Based on the image segmentation, each of the divided cells contains a variety of objects classification features, such as texture, shape and spectral characteristics. According to the training area, this paper selected nine features to analysis all the objects. Firstly, the texture features include the average gray value range within the scope of the convolution kernel (Texture Range) and the average gray value within the range of the convolution kernel (Texture Mean); Second, the shape of the main features include dividing unit area (Area), Rectangular Fit, the number of hole within the dividing cells(Number of Holes), directed axis length (Major length), circumference (Length) and Elongation; Finally, the spectrum features include the average gray value of each band (Spectral mean) (Deng, 2010).

3.2.1 *Arable land*

The first excluded object is arable land. Because of arable land belongs to one of the vegetation covering types, it is easily to be confused with other vegetation. However its shape characteristics is notable. For the shape of farmland and other features, the training threshold of image segmentation is 65, merge threshold is 98. First of all, since covering a large range of farmland, so the difference in segmentation cells is smaller than populus' (Xu and Yu, 2003). It also means the average gray value range (Texture Range) is lower. Also the its dividing area is larger than the populus'. Secondly, the shape of farmland is regular. But, due to other vegetation which grow in the arable land, it is not completely divided into rectangular units, but is similar with a rectangle. So its rectangular metric (Rectangular Fit) is larger than the populus'. The formula is below:

$$\begin{cases} Tx\_MEAN < 35 \\ Tx\_RANGE > 29 \\ \quad AREA > 50000 \\ RECT\_FIT > 0.4 \end{cases} \tag{1}$$

In the formula, Tx_MEAN is texture mean. Tx_RANGE is texture range, RECT_FIT is the measurement of the rectangular shape (Rectangular Fit).

3.2.2 *Bare land and Lake*

Bare land and lakes cover widely. Their texture is more smooth then the populus'.Therefore their Texture Range is smaller. And the training threshold of image segmentation is 55, merge threshold is 80. The formula is below:

$$Tx\_RANGE < 29 \tag{2}$$

3.2.3 *Shrubs*

After excluding large bare land and farmland, the shrubs become the most concentrated one in the picture. In Heihe River lower-stream, shrubs usually grows in the vicinity of the populus. Although both of Populus' and shrubs' texture are fine, but the populus distribute discretely, and shrubs' area is larger. Duiring the image segmentation, the shrubs are divided into a whole area and separated with the populus which is near it. So image segmentation threshold is 50, merging threshold is 85. Also the value of Area is greater than the populus, and the length of the major axis (Major Length) is larger. Due to fine texture, its value of the number of hole in the polygonse is greater. It means Texture Range and Number of Holes is larger. In addition to texture features, the color of shrubs is more similar to dark red, corresponding to the average gray value (Spectral Mean) and the average gray value within the range of the convolution kernel (Texture Mean) of Band 1 are lower than the populus. In accordance with these characteristics, shrubs can be described as follows:

$$\begin{cases} Tx\_RANGE > 55 \\ Tx\_MEAN < 25 \\ AVGBAND\_1 < 38 \\ NUMHOLES > 20 \\ AREA > 5000 \\ MAJAXILEN > 50 \end{cases} \tag{3}$$

In formula, AVGBAND_1 is the texture range of band 1. NUMHOLES is the number of holes within the divided cells. MAJAXILEN is the length of segmentation cells.

3.2.4 *Other Vegetation*

Firstly, other small desert vegetation is same to the populus. They are mostly distributed discretely, usually grow in bare ground. In the Pupulus National Nature Reserve of Heihe River lower-stream, there are more populus.So,the segmentation area (Area) is larger,and its perimeter (Length) is longer than the other small desert vegetation.

Secondly, the distribution of arable land and the trees along the road is regular. Therefore it's elongation (Elongation) and perimeter (Length) is larger.

Thirdly, the bare ground's luminance is higher then others'. It can be separated from populus by average gray value of Band 1 (Spectral Mean).

Based on the above analysis, image segmentation threshold is set to be 60, merging threshold is 93. And the formula is below:

$$\begin{cases} ELONGATION > 3 \\ AREA < 20 \\ LENGTH < 10\,or > 500 \\ RECT\_FIT > 0.66 \\ Tx\_MEAN > 135 \\ AVGBAND\_l < 30\,or > 120 \end{cases} \tag{4}$$

In formula, ELONGATION represents extensibility.

Thus, the features of main interference objects are analyzed and summarized. In Figure 3, the first one (Fig. 3a) is original image that intercepted from Google Earth. The second (Fig. 3b) is the result of segmentation and merging, the green line is split edges. It is clear that cultivated land were divided into a whole unit, separated from the populus. The third one is the Texture Range value map in the same area, the gray is darker the value is lower. It is obvious that average gray value of farmland is much lower than the populus'. The fourth one (Fig. 3d) is the overlay result of the extraction, which the green areas represent the extraction of the populus.
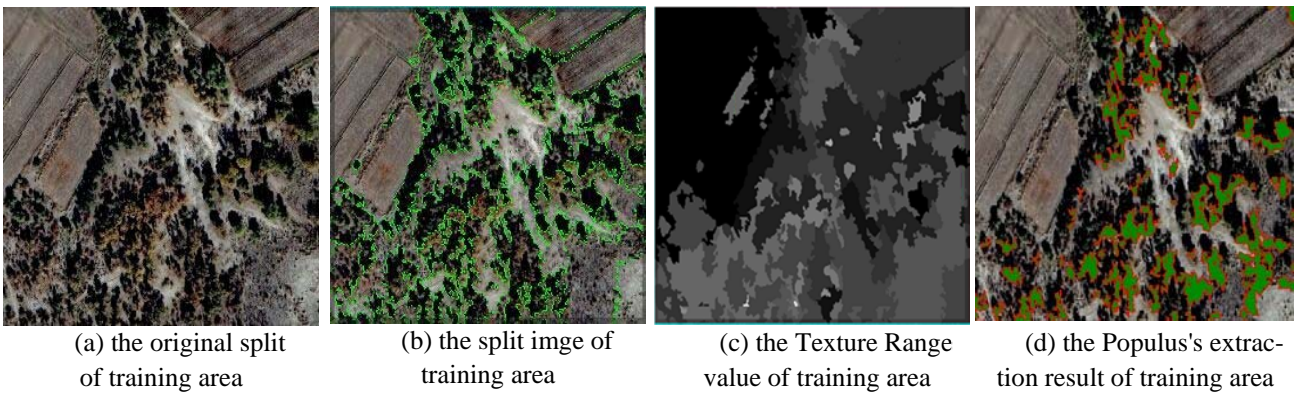

(a) the original split of training area   (b) the split imge of training area   (c) the Texture Range value of training area   (d) the Populus's extraction result of training area

Fig. 3 A piece of training area`s detail images

4 EXTRACTING POPULUS EUPHRATICA IN EXPERIMENTAL AREA

4.1 *experimental area*

In this research, the ruleset was applied to the Populus Euphratica Forest Natural Protection Zone in Heihe River lower-stream (41° 0'10.08"- 42°18'30.36" N,100°13'24.52"- 101° 6'40.59"E). In this area, the main species of vegetation are populus, shrubs, other small desert vegetation and arable land. Most of the populus grow along river, a small number of the populus grow in bare land.

4.2 *The extracting result*

In accordance with the process of research, we extracted the populus in the Populus Euphratica Forest Natural Protection Zone (Fig. 4), the extraction results have been indicated in green. In the Figure 4(a), the left one mainly contains rivers, bare land and populus; in Figure 4(b), the discrete populus distribute among other desert vegetation. It can be seen from the right picture that the populus is extracted, and other desert vegetation and shadow have been excludedt. In Figure 4(c), the populus are mixed with shrubs in the left one, however, in right one, the populus be retained and shrubs be excluded. In Figure 4(d), the populus distribution is scattered, that are alone the river with the shadow. We can conclude from extraction results in right picture that most of the shadow is removed and remained the single populus.
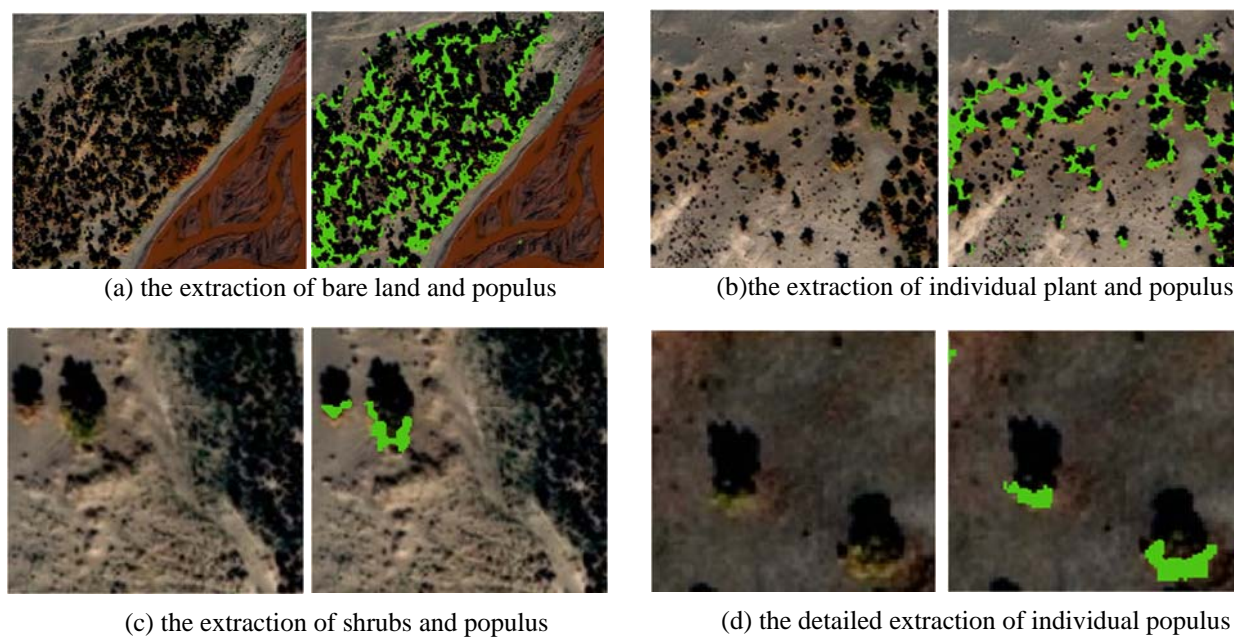

(a) the extraction of bare land and populus

(b)the extraction of individual plant and populus

(c) the extraction of shrubs and populus

(d) the detailed extraction of individual populus

Fig.4   The result details of euphratica in testing area

4.3 *Accuracy assessment*

This paper mainly extract the populus by visual interpretation through high-resolution images obtained from Google Earth. In order to objectively verify the accuracy of the extraction, we selected three researchers who is familiar with the populus to evaluate the accuracy of the results. The specific method is the confusion matrix. Compared with the Google Earth, we selected 680 rectangular ROI whose area is 3* 3, they occupied 6188 pixels. The evaluation results are shown in Table 1. Overall accuracy reached 87.4596%, Kappa factor is 0.7492. The leakage points are 386 pixels, the leakage rate is 12.51%, misclassification points are 390 pixels, error rate is 12.57%. Wherein the leakage points of the populus caused by shadow are 180, the leakage rate is 5.8%, error rate is 0%. It shows that the shadow is one of the main causes of the leakage points (Table 2)

Table 1 The confusion matrix of euphratica in testing area

Species	Others	Populus	Total
Others	2712	390	3102
Populus	386	2700	3086
Total	3098	3090	6188

Table 2 The leakage and error rate in testing area

	leakage points	leakage rate	error points	error rate
All populus	386/3086	12.51%	390/3120	12.57%
Shadow caused	180/3086	5.8%	0/3120	0%

5 CONCLUSION

Due to discrete distribution and small form of populus, traditional classification method and data cannot be automatically recognized and extracted, it also led to high cost and difficulty in automation. To resolve above problems, this paper propose an automatically extract method by remote sensing for populus forest in a large area. The proposed method is low-costed, automatic and highly precision, it is also superior in practical and economic application. Details are as follows:

(1) The image resolution of populus in Heihe River surpasses 2 m, and we need extract more than 3000 scenes of high-resolution remote sensing data which equals to 21646.6 sq.km. Nowadays, such as the high-resolution data of QUICKBIRD, and its price is 190 to 400 of each square kilometer. And it is impractical for a general case to takes about 650 million yuan at the above moderate prices. Since our method achieves objectives of mission and radically reduces the cost of paying for data, our method has high value in practical.

(2) Our method is automated in extracting, stitching and recognizing data of populus without any parts` joining. In large area, the number of high-resolution remote sensing data is large, and there is 58 GB of data in Heihe River. Our automated method can simple the processing in extracting, stitching and recognizing data of populus. Also, the above processing in our method is efficient.

(3) We also make experiment to compute precision of our method, and the precision is 87.46%, and our method is practical.

Although our method is for populus, it is easy to extend our method to the objects which also is small and discrete distribution. Moreover, due to the many times of full coverage for Google Earth, with high computing power of Cloud Computing, our method can also be applied in development of large application. Here, we make following scenarios as example:

(1) To precise agriculture and simplify the acquisition of national agricultural information, it is need to improve precision of agricultural data. We can achieve accurate extraction for the agricultural data within entire country and global using object-oriented approach and image data which resolution is about 2 m.

(2) To provide precision service for city plan, we can achieve function division for impermeable stratum in residential areas, commercial areas, industrial areas and mining areas using object-oriented approach and image data which resolution is about 4 m.

(3) We can achieve the recognize and extract of small and special object, such as airport, aircraft and power plants.

We will focus on achieving one or two of above applications with Cloud Computing to make more contributions. And we make a summary of the difficulties during achieving above objects as follows:

(1) Since the data of Google Earth do not have the spectral characteristics, so it only can use the texture characteristic of object to extract data of characteristic by object-oriented approach. But this processing is difficult for some special objects.

(2) With resolution increasing, the shadow effect of image will be more and more obvious. This would led to serious impact in the target recognition processing, which can reduce the resolution in recognition and extraction processing. Thus, it must need to consider how to reduce the shadow effect of image.

(3) Since the number of high-resolution data in large area is large, it is needed to apply Cloud Computing. And this is difficult in application level.

(4) Due to the different characteristics for different objects, it need specific analysis for specific objects. And there is some difficulties for general users.

Note that our method has a significant value which can be extended to a lot of applications in future. And this can contribute high-precision and high-resolution data for the development of national economic and society.

REFERENCES

Bo Z, Peng M, Nie A H, et al. Land cover mapping using time series HJ-1 / CCD data[J]. Science China Earth Sciences, 2014, 57(8):1790-1799.

Deng Shubin. ENVI Rmote Sensing Image Processing Method[M].Beijing: China Sience Publishing. 2010.

Guan Z, Cao G C and Yi J Z. The Classification Research of Object-oriented Classification[J]. Sience and Technology Innovation Herald., 00(34): p. 8-10.[DOI: 10.3969/j.issn.1674-098X.2010.34.007].

Li X, Liu S, Ma M, et al. HiWATER:An Integrated Remote Sensing Experiment on Hydrological and Ecological Processes in the Heihe River Basin[J]. Advances in Earth Science, 2012, 27(5):481-498.

Miaozhong X U. Automated Extraction of Shadows in Very-High Resolution Spatial Satellite Images[J]. Wtusm Bulletin of Science & Technology, 2003.

Shi C Y, Guo C L, Yang X T and Huang H G. The Extraction of Populus-crown in Ejin based on Quickbird[J]. Sience and Technology Innovation Herald., 2011(05): p. 13-15

Tan Q L, Gao J J. Building extraction from high resolution multispectral image using object-oriented classification method[J]. Engineering of Surveying & Mapping, 2010.

Xu F N, Qi Y, Wang J h and Zhang J L. Riparian Forest Vegetation Coverage Information Classification based on Object-

oriented Method in Heihe River[J]. Remote Sensing Technology and Application,2010. 30(5): p. 996-1005. [DOI:10.11873/j.issn.1004-0323.2015.5.0996]

Yang X,Zhen X Q,Zhao L and Zhang K Q. Rapid Acquisition of Elevation Data from Google Earth[EB/OL]. Beijing: Science Paper Online.2008. http://www.paper.edu.cn/release-paper/content/200810-506.

Zhang Z, Zhong X U, Wang J, et al. Value of the Ecosystem Services in the Heihe River Basin[J]. Journal of Glaciolgy & Geocryology, 2001.

Zhong B, Yang A, Nie A, et al. Finer Resolution Land-Cover Mapping Using Multiple Classifiers and Multisource Remotely Sensed Data in the Heihe River Basin[J]. IEEE Journal of Selected Topics in Applied Earth Observations & Remote Sensing, 2015, 8(10):4973-4992.

Zhou Y F, Zhang H F,Li X,Yang F and Ding C F. Extraction of Tree-crown of Populus Euphratica and Tamarix Ramousissima based on High Resolution Remote Sensing Data[J]. Remote Sensing Technology and Application, 2015. 30(3): p. 510-517 [DOI:10.11873/j.issn.1004-0323.2015.3.0510]