Study on Extraction Methods of Ocean Surface Oil Spill using HJ-CCD Data

Yingying Gai, Bin Zhou, Yuanfang Sun, Yan Zhou
Institute of Oceanographic Instrumentation
Shandong Academy of Sciences
Qingdao, China
gaiyingying@pku.edu.cn

Wenwen Li
GeoDa Center for Geospatial Analysis and Computation
School of Geographical Sciences and Urban Planning,
Arizona State University
Tempe, USA
wenwen@asu.edu

Abstract—Oil spill is one of the most common ocean environmental disasters. Rapid and accurate access to the oil spill information has great significances to dynamic monitoring, conservation and sustainable use of the ocean. HJ-1 is a new satellite platform designed for environmental disasters. However, multi-spectral images obtained from HJ-1 have fewer spectral bands and the accuracy of acquiring the oil spill coverage only by spectral information is low. In this paper, oil spill occurred in Gulf of Mexico was selected as the research object. The texture characteristics which affect oil spill identification were extracted by Gray Level Co-occurrence Matrix (GLCM) and a decision tree method combined spectral characteristics with texture characteristics was established to extract the oil spill. And comparative analysis with the extraction result of Maximum Likelihood Classification (MLC) was performed. Results showed that, compared with MLC, the decision tree classification raised the user accuracy and the producer accuracy of extracting oil spill by 11.85% and 4.28%, respectively.

Index Terms—Oil spill, HJ-1, Gulf of Mexico, texture characteristic, decision tree.

I. INTRODUCTION

Oil spills on the sea surface are seen relatively often [1]. Oil spills not only pose a serious threat to the marine ecological environment, but also bring serious impact on the social economy and human health to the coast. Therefore, taking a quick and effective way to monitor oil spills on the sea is of great significance.

As a real-time monitoring tool, remote sensing plays an important role for oil spill monitoring; as it can provide early detection of surface oil [2]. Nowadays, a lot of remote sensing satellites are being used for oil spill monitoring, such as Landsat, SPOT, and NOAA. As a new generation of small satellites, HJ-1 satellite has the advantages of a wide monitoring range, a short repeat cycle, and low cost. It has been developed into a principal tool for land and ocean monitoring. However, CCD multi-spectral images acquired from HJ-1 satellite have only four bands and the band width is so large that the amount of spectral information is insufficient to distinguish oil from other features. Especially, when the images are covered with clouds, the spectrums of oil and cloud are too close to make a distinction using only spectral information. In addition, coastal waters have similar spectral characteristics to oil slicks. It is also difficult to distinguish them only through spectral characteristics.

In this paper, a CCD image which shows the oil spill happened in the Gulf of Mexico was selected as the research object. On the basis of spectral analysis, texture information extracted by GLCM was combined, and a decision tree method was established to recognize oil spill.

II. DATA ACQUISITION AND PRE-PROCESSING

April 20th 2010, BP's Deep Water Horizon drilling platform exploded, resulting in a large number of crude oil leaking into the Gulf of Mexico. A CCD image which covered this area with a spatial resolution of 30 m was acquired on May 10th 2010 from HJ-1 satellite. After radiometric correction and geometric correction, the image was also atmospheric corrected using 6S model.

III. RESEARCH METHODS

A. Spectral Analysis of Typical Features

Generally, different features show different spectral characteristics in different bands. Therefore, it is possible to separate them based on the spectral characteristics [3]. In this study, 6 typical features including oil spill, land, pure seawater, coastal water, cloud and cloud shadow were sampled, and mean values of the reflectance in each band were counted to obtain the spectral curves as Fig. 1 shows.
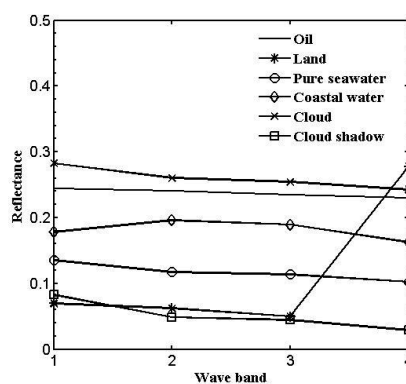

Fig. 1. Spectral curves of typical features

As can be seen from Fig. 1, the spectral curves between the land and the remaining five features are significantly different, because the land surface was almost covered by vegetation during this period. High reflectance of vegetation in near-infrared band makes reflectance of b4 (referring to the band 4, the same below) is much larger than that of b3. Therefore, the land can be separated by Normalized Difference Vegetation Index (NDVI). When oil is thick, the oil reflectance in near-infrared band is also larger than that in red band. However, compared with vegetation, the reflectance difference between two bands of the thick oil is relatively small. After repeated tests, the separation of land and the rest features works best when NDVI>0.1.

Besides land, there are also differences between the spectral curves of coastal water and other features. As coastal water is affected by the urban sewage or shallow beach, the reflectance of this area in each band is much higher than that of pure seawater, closing to the oil reflectance. It is found that most of the coastal waters in this area have a common point that sum of reflectance in blue and near-infrared bands is smaller than that in green and red bands, but other features not. After testing, the pixels which satisfy R (b2+b3-b1-b4)>0.03 are divided into coastal water. However, coastal water cannot be completely extracted by this characteristic and only the contaminated water far away from the edge of the land can be recognized, because coastal water close to the land, which we might reasonably call shoal water, does not have this characteristic. Therefore, it is necessary to combine texture characteristics to remove the interference of shoal water.

The spectral curves of oil, pure seawater, cloud and cloud shadow have similar trend. However, the reflectance of pure seawater and cloud shadow is much lower than that of oil and cloud. This difference is particularly evident in the visible bands. Therefore, when the sum of all the visible band reflectance is larger than 0.5, i.e. R (b1 + b2 + b3)>0.5, oil and cloud can be recognized, as well as complete coastal water. Combined with the above-obtained spatial distribution of contaminated water, an image contains only oil, cloud and shoal water is finally obtained.

The spectral curves of oil, cloud and shoal water have little difference, with a certain coincidence in the range of reflectance maximum and minimum, as shown in Table I. They cannot be distinguished based on the spectral characteristics and texture characteristics should be considered.

TABLE I. Spectral Reflectance Statistics of the Oil, Cloud and Shoal Water Sampling Points

Bands	Features	Mean	Maximum	Minimum
b1	oil	0.244	0.424	0.173
	cloud	0.283	0.401	0.215
	shoal water	0.194	0.397	0.162
b2	oil	0.241	0.451	0.161
	cloud	0.260	0.389	0.188
	shoal water	0.232	0.366	0.164
b3	oil	0.235	0.444	0.162
	cloud	0.254	0.380	0.183
	shoal water	0.218	0.397	0.168
b4	oil	0.230	0.473	0.149
	cloud	0.243	0.368	0.166
	shoal water	0.210	0.374	0.141

B. Texture Analysis of Typical Features

Texture is a very important factor in remote sensing image thematic information extraction. There are two types of texture analysis methods, statistical methods and structural methods. Statistical methods are often applied to images with fine and random texture, while structural methods regular and obvious texture [4]. The texture of oil spill is ruleless as it is influenced by winds and waves. Therefore, a statistical method called GLCM is used to extract the texture information.

GLCM describes texture by the spatial co-occurrence of pixel gray levels. The GLCM operates by computing a matrix that is based on quantifying the difference in gray-scale values between pixels at a predefined distance. This matrix is then used to compute a number of texture features to summarize and represent the structure of the matrix. According to GLCM, many measures which describe texture can be defined. Here we chose mean, variance, correlation, contrast, dissimilarity, homogeneity, angular second moment and entropy for analysis and comparison.

As the texture characteristics are structural, they have nothing to do with the band choice. Therefore, the blue band image was selected to calculate the measures. In order to accurately describe the relationship between target and surrounding environment, step was set to 1. In fact, research has shown that direction has little impact on classification result. Therefore, we took the mean value of four directions, i.e. 0°, 45°, 90°, 135°. Image histogram equalization was done before texture extraction to increase the dynamic range of gray values and image gray level was compressed to 16 to reduce the amount of computation.

An appropriate calculation window is essential for GLCM calculation. Texture would be portrayed too detailed to describe the internal pixels arrangement of the same feature if calculation window were too small. On the contrary, if calculation window were oversize, heterogeneous feature pixels information might be contained and texture boundaries would not be clear enough [5]. Considering the spatial resolution and the spatial distribution of oil spill, four windows of different sizes were chosen to calculate GLCM, i.e. 5×5, 9×9, 13×13, and 17×17. Figure 2 shows the correlation texture images under different window sizes. It can be seen that texture is too detailed under 5×5 and 9×9 windows, but texture images obtained under 13×13 and 17×17 windows fit well with the distribution rules. Considering the texture differences of oil, clouds and shoal water under different window sizes and the calculation time, 13×13 window was selected.
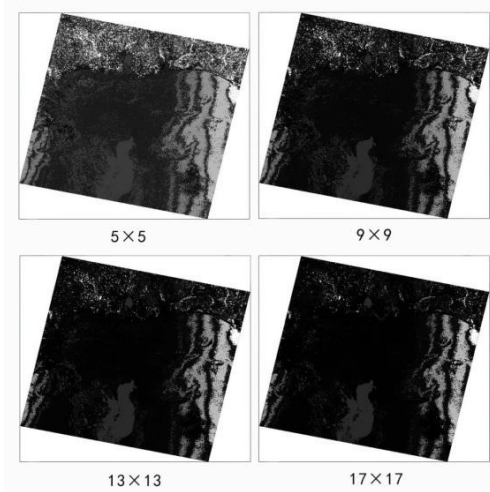
Fig. 2. Correlation texture images calculated with different windows

Due to the large difference among averages of the measures, normalization needs to be done before comparison. Figure 3 shows the normalized mean curves of eight measures under 13×13 window. Here shoal water was taken into account. From the figure, we can see the biggest difference of the curves among oil, shoal water and cloud is correlation. Therefore, correlation texture characteristics can be further used to distinguish them.
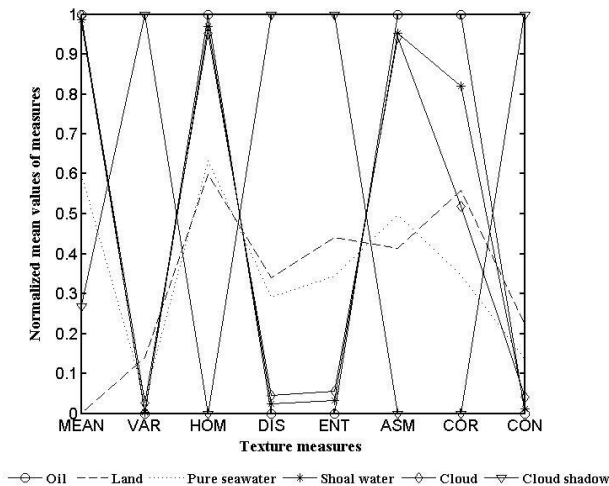

Fig. 3. Normalized mean texture measures curves of typical features

C. Decision Tree Model

Decision tree classification operates by dividing remote sensing datasets to get various sub-categories with different properties according to certain rules. Datasets which can be processed by decision tree include not only spectral information, but also texture, elevation and spatial characteristics [6].

A decision tree model which combined spectral and texture characteristics is shown in Fig. 4. Each segment threshold was selected after repeated tests. Since the research object is oil spill, no rules are established to distinguish pure seawater and cloud shadow, as well as cloud and shoal water. In this model, spectral and texture information are both important and indispensable.
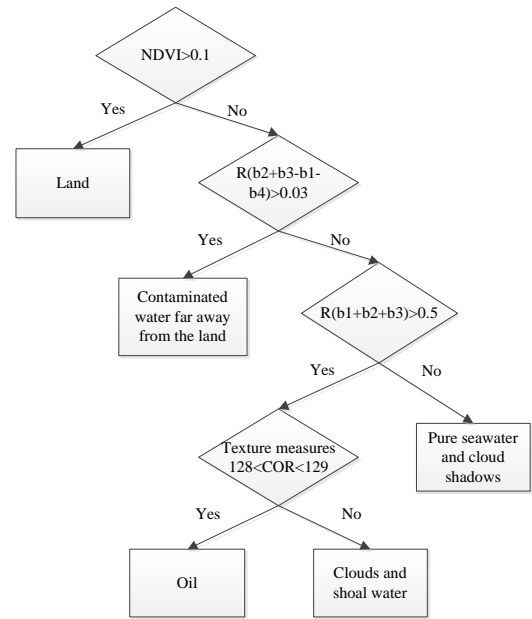

Fig. 4. Decision tree model for oil spill extraction

IV. EXTRACTION RESULTS AND ACCURACY ANALYSIS

The major oil spill area can be identified by decision tree as shown in Fig. 5, especially the thin oil in the edge. At the same time, it is able to remove most of the interference from the cloud and coastal water.
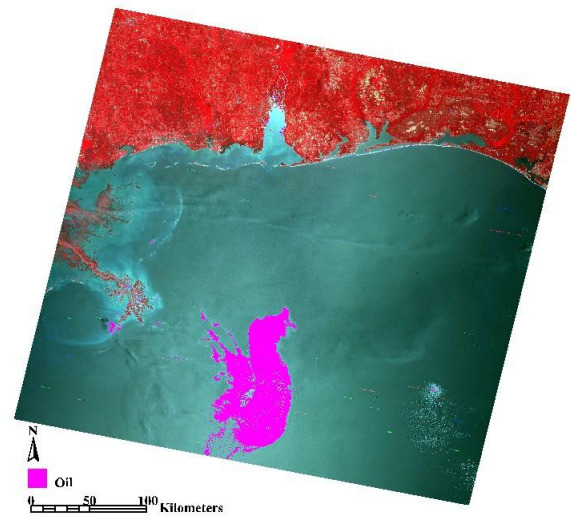

Fig. 5. Oil extraction result

Comparison of decision tree classification and MLC was carried out to illustrate the advantages of our method. First, some points of different features, which must be proportional to the classification, were randomly sampled from the image. Then, confusion matrix and Kappa coefficient can be obtained by comparing the extracted results with the sample data, as shown in Table II and Table III.

TABLE II. ACCURACY ANALYSIS OF DECISION TREE CLASSIFICATION METHOD

Features	Oil	Land	Coastal water	Clouds	Pure seawater and cloud shadows	User accuracy (%)
Oil	171852	361	11213	0	15223	86.51
Land	0	1168091	17	0	0	99.99
Coastal water	2411	58	24725	0	333	89.82
Clouds	11	0	0	2343	0	99.53
Pure seawater and cloud shadows	21224	6880	13117	2966	3260095	98.66
Producer accuracy (%)	87.90	99.38	50.39	44.13	99.53	overall accuracy =98.43% Kappa coefficient =0.9648

TABLE III. ACCURACY ANALYSIS OF MAXIMUM LIKELIHOOD CLASSIFICATION METHOD

Features	Oil	Land	Coastal water	Clouds	Pure seawater and cloud shadows	User accuracy (%)
Oil	153174	12011	9800	800	29379	74.66
Land	285	1163138	0	0	13	99.97
Coastal water	2	112	35768	7	11541	75.41
Clouds	11603	48	484	3695	6323	16.68
Pure seawater and cloud shadows	18113	81	3003	807	3249845	99.33
Producer accuracy (%)	83.62	98.96	72.91	69.60	98.57	overall accuracy =97.78% Kappa coefficient =0.9508

From the table, we can see both overall accuracy and Kappa coefficient are high for the two methods. However, the interference of other features with oil spill extraction is relatively great for MLC, especially cloud and pure seawater, while the interference of cloud and land can basically be ruled out for our method and the user accuracy and producer accuracy are up to 86.51% and 87.90%. Compared with MLC, they are increased by 11.85% and 4.28%, respectively.

V. CONCLUSIONS

Coastal water and cloud are the largest confounding factors for oil spill recognition in HJ-1 CCD images. Confusion would be likely to happen if relying only on spectral information. The shape, roughness and spatial position of the features are also important characteristics. The method which combines spectral characteristics with texture characteristics can effectively improve the accuracy of oil identification.

This method provides an effective way for oil spill extraction using HJ-CCD data. However, the thresholds of the model are based on the experience and repeated tests and do not apply to all images. Therefore, threshold acquisition should be further studied for the automatic monitoring of oil spills. In addition, texture extraction based on GLCM requires a lot of calculation and a long time. In order to realize real-time and high-precision monitoring of oil spills, we cannot reduce the amount of calculation by reducing the image gray levels, which requires further research on the extraction algorithms of texture information.

REFERENCES

[1] B. Camilla and H. S. Solberg Anne. "Oil spill detection by satellite remote sensing," Remote Sensing of Environment, 2005, 95(1): 1-13.

[2] F. Carnesecchi, V. Byfield, and P. Cipollini. "An optical model for the interpretation of remotely sensed multispectral images of oil spill," Proceeding SPIE 7105, Remote Sensing of the Ocean, Sea Ice, and Large Water Regions 2008. Wales, 2008.

[3] L. M. Jiang, F. Wang, Z. K. Xiao, and D. S. Zhang. "Study on extracting wetland information in Weishanhu Lake area based on texture feature," Journal of Hunan University of Science and Technology(Natural Science Edition), 2011, 26(4): 68-72.

[4] L. Zhang. "Wetland Information Extraction Combined with Texture Features," Remote Sensing Information, 2010(3): 30-34.

[5] X. Y. Tan. "A classification method by use of SAR image texture characteristics with combination of large and small windows," Systems Engineering and Electronics, 2000, 22(4): 15-17.

[6] S. Ba, Z. H. Liu, Z. J. Zhang, and Q. S. Li. "Decision tree and its application in remote sensing image classification," Plateau and Mountain Meteorology Research, 2011, 31(2): 31-34.