IOP Conf. Series: Materials Science and Engineering **392** (2018) 062199 doi:10.1088/1757-899X/392/6/062199

A Cloud Detection Algorithm for MODIS Images Combining Kmeans Clustering and Otsu Method

P S XIANG

College of Electrical & Information Engineering, Southwest Minzu University, Chengdu 610041 CHINA

sue_xiang@163.com

Abstract. This paper proposes a method to choose the initial clustering center for the kmeans algorithm and recognize the clustering results without artificial interpretation: Otsu method are used on the MODIS data to determine cloud pixels and not cloud pixels. Take the nearest pixels from each class's mean value as the initial clustering center of the kmeans algotrithm and determine the category of the clustering results according to the category of the initial clustering center. Theoretically, the method reduce the error caused by the random choice of the initial clustering center in traditional kmeans algorithm and realized automatic cloud detection without artificial interpretation. The experimental results verify the effectiveness of the proposed method.

1. Introduction

Due to cloud interference, the utilization ratio and accuracy of remote sensing image data are reduced. Therefore, cloud detection and cloud rejection have become one of the most important issues to be solved in the process of optical remote sensing image processing[1].

At present, most of the cloud detection methods are threshold method, but the threshold method has poor universality, and the selection of threshold depends on prior knowledge and is subjective. There are also researchers having cloud detection by using artificial neural networks, clustering algorithms, and so on. But it is not suitable for any situation.

Wang Wei from the University of Science and Technology of China has proposed clouds detection algorithm combining Kmeans with a fixed threshold in 2011, and got an ideal result. He adopted the maximum and minimum distance selection method[2][3] to determine the initial clustering center of the Kmeans algorithm, the selection of the first initial clustering center point is random and does not indicate that how to automatically determine which clustering result is the cloud.

Based on the research of Wang Wei, this paper used the Otsu method to determine the initial clustering center of the Kmeans algorithm and labeled clustering results obtained from the Kmeans algorithm. Finally, the fixed thresholds of other bands are combined to have cloud detection. And automatic cloud detection is realized with ideal results.

2. Cloud detection algorithm

The cloud detection algorithm in this paper is divided into three steps: first, the Otsu algorithm was used to perform initial cloud recognition on the MODIS data (generally select band 26), and one point being the closest to the mean values of the current set was respectively taken in the “cloud” set and “non-cloud” set. The points were used as the initial clustering center of the next Kmeans algorithm; then band 1 and band 2 were used as the horizontal and vertical axes on the feature plane, and the

IOP Conf. Series: Materials Science and Engineering **392** (2018) 062199 doi:10.1088/1757-899X/392/6/062199

Kmeans algorithm was used for clustering. The clustering formed by the points selected in the "cloud" set in the previous step as the initial clustering centers were cloud, smog, and snow. The clustering formed by the points in the "non-cloud" set as the initial clustering center were other land features; finally, the fixed threshold method in the literature[2] was used to eliminate the interference from the cloud, smog, and snow categories in the Kmeans clustering results, and the detection results of the band 26 were merged to obtain the final detection results. The flow diagram of this method is as follows:

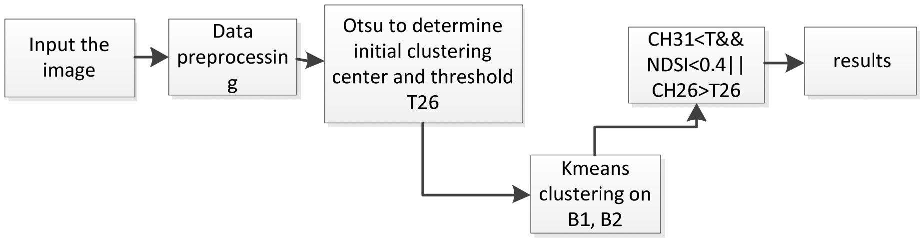


Figure 1. Flow diagram of cloud detection algorithm .

2.1. Initial cloud recognition of Otsu algorithm

In the mid-infrared channel of MODIS band 26 (1.38 μm), it is difficult for the radiation of the ground, medium and low cloud to reach the sensor due to the strong absorption of water vapor. The reflectance of the ground is almost 0, but cirrus belonging to the high cloud system has small humidity and large reflectivity [5]. Therefore, the band 26 data was selected to run the Otsu algorithm, the optimal threshold T26 was determined, and the "cloud" (CH26>T26) set and "not cloud" (CH26<T26) set in the image were also initially identified at the same time. Because the Kmeans algorithm is sensitive to the initial clustering center, when the selection of initial clustering center is close to the actual clustering, Kmeans has the best clustering effect, therefore the initial clustering center was selected according to the running result of the Otsu algorithm: in the "cloud" set and "non-cloud" set divided by the best threshold of the band 26, the point being the closest to the mean value of each set was selected as the initial clustering center of the next Kmeans algorithm.

2.2. Initial cloud classification of Kmeans algorithm

After determining the initial clustering center according to the running result of the Otsu algorithm, band 1 and band 2 were selected as clustering features [2] for Kmeans clustering. The clustering formed by the points in the "cloud" set in the running result of Otsu algorithm as initial clustering center is the "cloud, snow, smog" category, and this category of data was subjected to the next fixed threshold algorithm to eliminate interference.

2.3. Fixed threshold cloud detection

This part of the cloud detection method adds cloud detection result of band 26 based on the research results of Wang Wei:

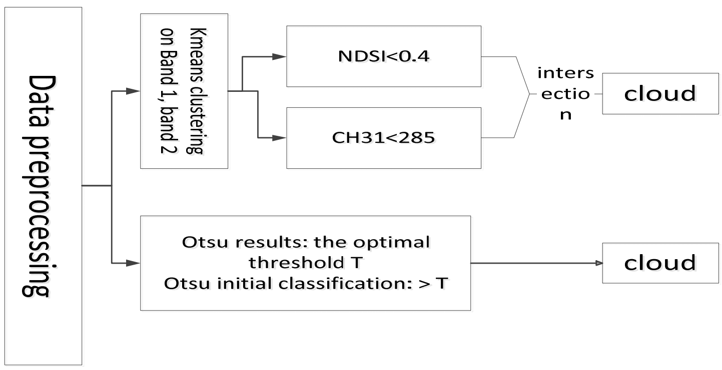


Figure 2. Flow diagram of fixed threshold cloud detection.

IOP Conf. Series: Materials Science and Engineering **392** (2018) 062199 doi:10.1088/1757-899X/392/6/062199

3. Experimental and analysis

In order to verify the applicability of this algorithm, the remote sensing images of four complex surface types and atmospheric conditions listed in [2][6] were selected for experiments. In the Fig. 3, the white area is the cloud area, and the black area is the underlying surface.

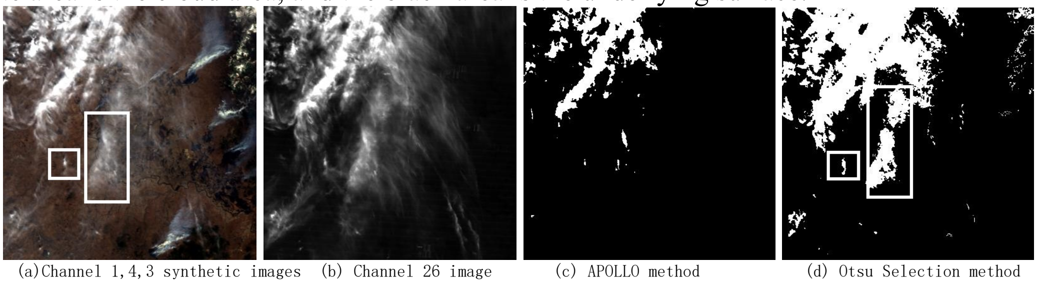


Figure 3. 400*400 Images captured from original image on April 27, 2009 and processing results of two cloud detection algorithms.

Repeating the experiment of Wang Wei: Otsu selection method can effectively identify the cloud area, since the Otsu classification result of band 26 was used to select the kmeans initial clustering center, it has higher reliability and adaptability and can automatically complete cloud detection without human involvement.

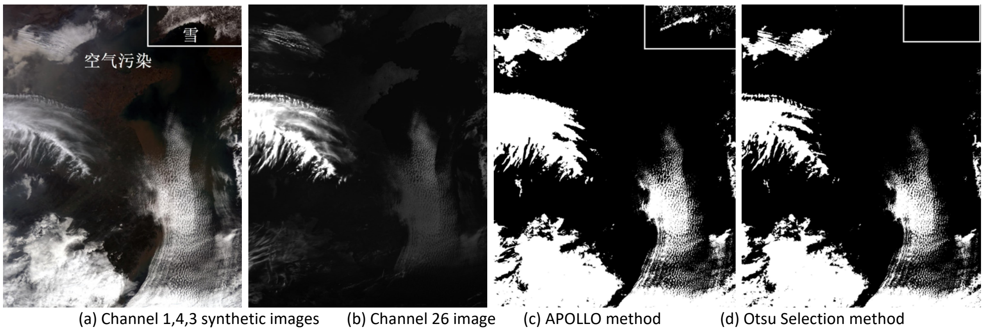


Figure 4. Original image on March 14, 2013 and processing results of two cloud detection algorithms.

Fig. 4 contains snow and air pollution. Comparing the effect of the two algorithms in Fig. 4, neither of the two algorithms misjudged the air pollution as clouds, and the Otsu selection method did not misjudge the snow in the image as clouds because it contains the snow filter.

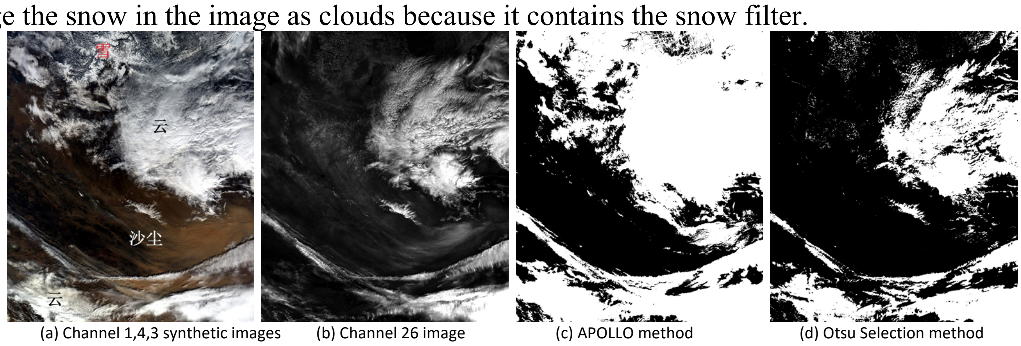
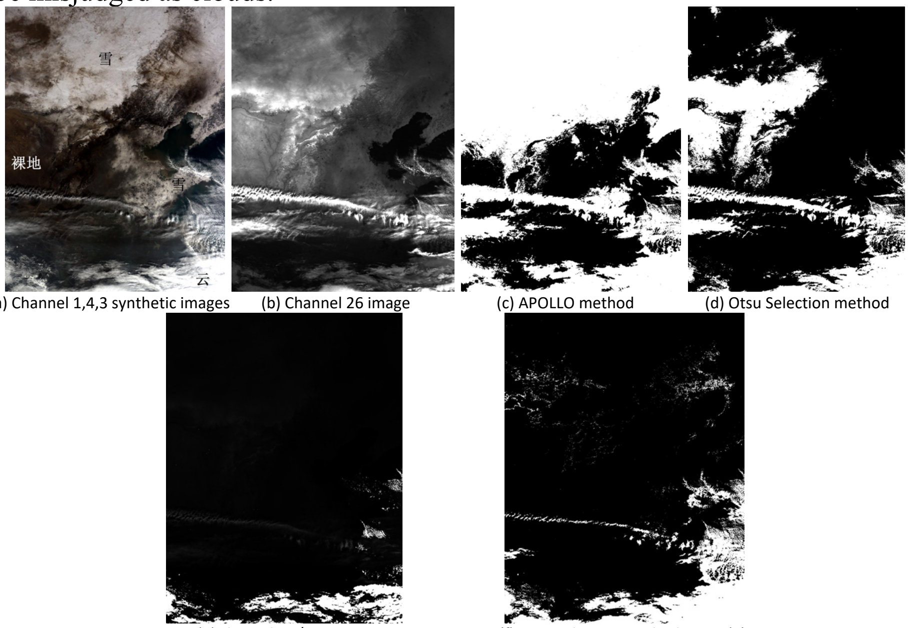


Figure 5. Original image on March 30, 2007 and processing results of two cloud detection algorithms.

Fig. 5 contains a lot of dust and a small amount of snow. Comparing the cloud detection effect of the two algorithms in Fig. 5, the APOLLO method did not take into account snow, thus it misjudged the snow as clouds, and also misjudged some of the dust as clouds. And the Otsu selection method can identify most of the dust and filter out the snow. The effect of cloud detection is ideal.

Since the Otsu selection method depends on the initial classification result of the band 26, when the image of the band 26 contains a bright ground, the initial classification misjudges bright ground as clouds, causing incorrect selection of initial clustering center of kmeans algorithm, and the naked ground will be misjudged as clouds.



(a) Channel 1,4,3 synthetic images (b) Channel 26 image (c) APOLLO method (d) Otsu Selection method

(e) ch26+ch20/10000 imag (f) Otsu Selection method to Fig.(e)

Figure 6. Original image on March 30, 2007 and processing results of two cloud detection algorithms.

In order to process the images in Fig. 6, the Otsu initial classification object was set to "CH26+CH20/10000", as shown in Fig. 6(e), and the detection result of Fig. 6(f) shows that when cloud clearly differs from other ground features in Otsu initial classification images, the cloud can be correctly identified. In general, the data of band 26 is sufficient, but in some cases, when the cloud and other features are not clearly distinguished in the band 26, the Otsu initial classification object still needs to be manually selected. Therefore, the algorithm still cannot completely automate all situations.

4. Conclusion

This paper proposed an Otsu-based Kmeans algorithm to achieve automatic cloud detection for remote sensing images. The initial clustering center of the Kmeans algorithm is determined by running Otsu operation on the Otsu initial classification object (generally select band 26), then the Kmeans clustering is performed by using the band 1 and the band 2 as clustering features, and finally the NDSI index and the bright temperature value of channel 31 are used to eliminate snow and smog interferences and merge Otsu initial classification results.

As seen in the above comparison experiments, the Kmeans algorithm is very sensitive to the selection of the initial clustering center point, resulting in the sensitivity of Otsu selection method to Otsu initial classification object: it requires a clear distinction between cloud and other ground features, then generates the correct initial clustering center of Kmeans algorithm. In general, the data of band 26 meets this requirement, but in some cases, Otsu initial classification object still needs to be manually selected. Otsu selection method can achieve automatic cloud detection for some complex surface and atmospheric conditions, but it still cannot achieve automatic cloud detection in all cases.

Acknowledgments

Authors wishing to acknowledge NASA for the data used in this study.

References

[1] Hou S wei,Sun W fang,Zheng X song. Overview of cloud detection methods in remote sensing images 2014 *J Space Electronic Technology* 3 68-76.

IOP Conf. Series: Materials Science and Engineering **392** (2018) 062199 doi:10.1088/1757-899X/392/6/062199

[2] Wang Wei , Song W guo, Liu S xing,ect. A cloud detection algorithm for MODIS images combining kmeans clustering and multi-spectral threshold method 2011 *J Spectroscopy and Spectral Analysis* 31 4 1601-04.

[3] Wei Wang 2011 The study of cloud and smoke plume detection using kmeans clustering and multithreshold approaches. (Hefei: University of Science and Technology of China)P38-40

[4] Ding Yuye 2013.Research on cloud detection algorithm for MODIS data. (Harbin: Harbin Institute of Technology)P26-27

[5] Li Wei, Fang S hui, Dian Y yong,ect.Cloud detection in MODIS data based on spectrum analysis 2005 *J Geomatics and Information Science of Wuhan University* 30 5 435-38.

[6] LYU Mingming, HAN Lijian, TIAN Shufang,ect.Cloud Detection under varied surfaces and atmospheric conditions with MODIS imagery 2016 *J Journal of Remote Sensing* 20 6 1371-80.