A New Method of Cloud Detection Based on Cascaded AdaBoost

C Ma[1], F Chen, J Liu and J Duan

[1] Center for Earth Observation and Digital Earth Chinese Academy of Sciences, No.9 Deng Zhuang Road, Haidian District, Beijing, China,100094

Email: abcmacai@163.com

Abstract. Cloud detection of remote sensing image is a critical step in the processing of the remote sensing images. How to quickly, accurately and effectively detect cloud on remote sensing images, is still a challenging issue in this area. In order to avoid disadvantages of the current algorithms, the cascaded AdaBoost classifier algorithm is successfully applied to the cloud detection. A new algorithm combined cascaded AdaBoost classifier and multi-features, is proposed in this paper. First, multi-features based on the color, texture and spectral features are extracted from the remote sensing image. Second, the automatic cloud detection model is obtained based on the cascaded AdaBoost algorithm. In this paper, the results show that the new algorithm can determine cloud detection model and threshold values adaptively for different resolution remote sensing training data. The accuracy of cloud detection is improved. So it is a new effective algorithm for the cloud detection of remote sensing images.

1. Introduction

Remote sensing satellite images have significant applications in different areas such as climate modelling, flood forecasting, water resources management, assessment of forest resources, examining marine environments etc [1, 2]. Cloud detection of remote sensing image is a critical step in the processing of the remote sensing images [2, 5]. How to fast, accurately and effectively detect cloud on remote sensing's data, is still a challenging issue in this area.

Most of the currently cloud detection methods were developed for moderate resolution sensors. Zhu [3] (2012), Hollingsworth [4] (1996) and Irish [6] (2006) proposed some methods for Landsat images. While, Wang [13](2006) and Li[16](2009) do research on the MODIS images. There are also some studies on other sensors [2, 15]. Most of the researches are based on finding better features to select the cloud from water, soil, shadow, stone and other objects on the earth's surface[3,4,6,7,8,9,10,11,12]. Those methods can do well for special remote sensing data. But it does not work well on the other images. How to select features, set the threshold values for features, and organize the model's order? It needs enough priori knowledge. In order to avoid disadvantages of the current algorithms, the cascaded AdaBoost classifier algorithm is successfully applied to the cloud detection. A new algorithm, i.e. the combined cascaded AdaBoost classifier and multi-features, is proposed in this paper. This new method solves the feature selection and the threshold value setting problems. The results show that the new algorithm can determine cloud detection model and threshold values adaptively for different resolution remote sensing training data. The accuracy of cloud detection is improved. So it is a new effective algorithm for the cloud detection of remote sensing images.

2. The AdaBoost Algorithm

In this section, we will give a briefly description of the AdaBoost algorithm. The original AdaBoost algorithm was introduced by Freundand Shapire [13, 14]. It is a supervised learning algorithm designed to find a binary classifier that discriminates between positive and negative examples. This

---

[1] To whom any correspondence should be addressed.

original algorithm can be seen as a weak classifier which is defined as a "rough and moderately inaccurate" predictor, but better than a chance [15]. Given (x_i, y_i), $1 \le i \le N$ denote a set of training data, where N represents the number of training data. In addition, $x_i \in R^d$ and $y_i \in \{0,1\}$, where d denotes the number of dimensions of input data and the value of y_i indicates whether x_i is a negative or positive example. Each weak classifier is given a function $h(x, f, p, \theta)$ returning defined by the Eq. (1). The $h(x, f, p, \theta)$ can be defined by the eq.1.

$$h_i(x, f, p, \theta) = \begin{cases} 1 & (pf(x) < p\theta) \\ 0 & (otherwise) \end{cases} \tag{1}$$

Where f denotes one dimension of input data x, θ is a threshold and p denotes the direction of the inequality. The function will return a boolean value {0, 1}. The output value of 1 represents that the input data x is classified as a positive example and 0 otherwise.

The AdaBoost algorithm boosts the classification performance of a simple learning algorithm by combining a collection of weak classifiers to a stronger classifier. The stronger classifier uses supervised learning with the wrapper method of feature selection. It selects the best weak classifier with respect to a given weighted error of the input data at each iteration. The data misclassified in previous classifiers get higher weights in the next iteration. Then the stronger classifier can be tuned in favour of previous classifiers. Assuming T denotes the number of the AdaBoost weak classifiers and the iterations times. The AdaBoost stronger classifier can be obtained according to the eq.2.

$$C(x) = \begin{cases} 1 & (\sum_{t=1}^{T} a_t h_t(x) \ge \frac{1}{2} \sum_{t=1}^{T} a_t) \\ 0 & (otherwise) \end{cases} \tag{2}$$

Where a_t is the weight value of the t th weak classifier, and it can be obtained by eq.3.

$$a_t = \log \frac{1}{\beta_t} = \log \frac{1-\varepsilon_t}{\varepsilon_t} = -\log \varepsilon_t \tag{3}$$

Where ε_t represents the classification errors of the t th weak classifiers.

3. The New Method of Cloud Detection Based on Cascaded AdaBoost
The stronger classifier can increase the classification precision. However, it needs to combine a great quantity of weak classifiers. Then, the resulting detector is very slow with mass weak classifiers. To overcome this problem, the cascaded AdaBoost algorithm is adopted in this paper. A new method of cloud detection based on cascaded AdaBoost is proposed. The cascaded AdaBoost model has two key properties that computation time and the detection rate can be adjusted.

3.1. Building the cascaded AdaBoost detector model
The cascaded AdaBoost classifier model is comprised of several stronger classifiers. Each stronger classifier gets a high precision rate. Then, good results will be obtained by a simple aiming for a fixed maximum false drop rate (f_{max}) and minimum detection rate (d_{min}) for each stage. Suppose, the detection rate of each stage classifier is d_i, false drop rate is f_i, and M is the number of the model stages. Then, the overall target precision rate D and the overall target false drop rate F can be obtained by the following equations:

$$D = \Pi_{i=1}^{M} d_i > (d_{max})^M \tag{4}$$

$$F = \Pi_{i=1}^{M} f_i < (f_{max})^M \tag{5}$$

More detail about the cascaded AdaBoost classifier model can be seen in the figure 1.

（1）Initialize the parameters: f_{max}, d_{min} and the overall target false drop rate F_{target}.

（2）Choose the training input data, and classify them to "cloudy pixels" and "clear-sky pixels". Then, specify the feature vector of the training input data as x_i.

（3）Initialize the false drop rate and detection rate as $F_0 = 1.0$, $D_0 = 1.0$. And set the number of current working stage as $i = 0$.

（4）While($F_i > F_{target}$)

{

    1) $i = i + 1$

    2) Initialize the number of weak classifiers on the i th stage, $n_i = 0$.

    3) Set $F_i = F_{i-1}$

    4) While ($F_i > f_{max} F_{i-1}$)

    {

        a) $n_i = n_i + 1$

        b) Use the training input data to generate n_i number of weak classifiers.

        c) Use the cascaded Adaboost model generated by before training to get the current the false drop rate F_i and detection rate D_i.

        d) Reduce the threshold of the stronger classifier on the i th stage. Until the current cascaded Adaboost model's detection rate $D_i > D_i.d_{min}$.

        e) Compute the new false drop rate F_i.

    }

    5) Till the number of "clear-sky pixels" is zero.

    6) If $F_i > F_{target}$, kick out the "clear-sky pixels" which is classified right as "clear-sky pixels" from the training data. Then update the training data.

}

Figure 1. The processing flow of building the cascaded Adaboost detector model in cloud detection

3.2. Feature Selection of input data

Cascaded AdaBoost is a degradation of the decision tree. Figure 2 shows the main procedures that the cascaded AdaBoost detector model was applied in the cloud detection. In the cloud detection processing for the target images, the feature vector of each pixel was specified first. Then those feature vectors were classified in the first classifier as the input data. If they are classified as "cloudy pixels", they follow into the next stage classifiers and so on. Once they are classified as "clear-sky pixels" by one stage classifier, they will finally be set as "clear-sky pixels". Only the pixels which were classified as "cloudy pixels" by every stage classifier can be set as "cloudy pixels".
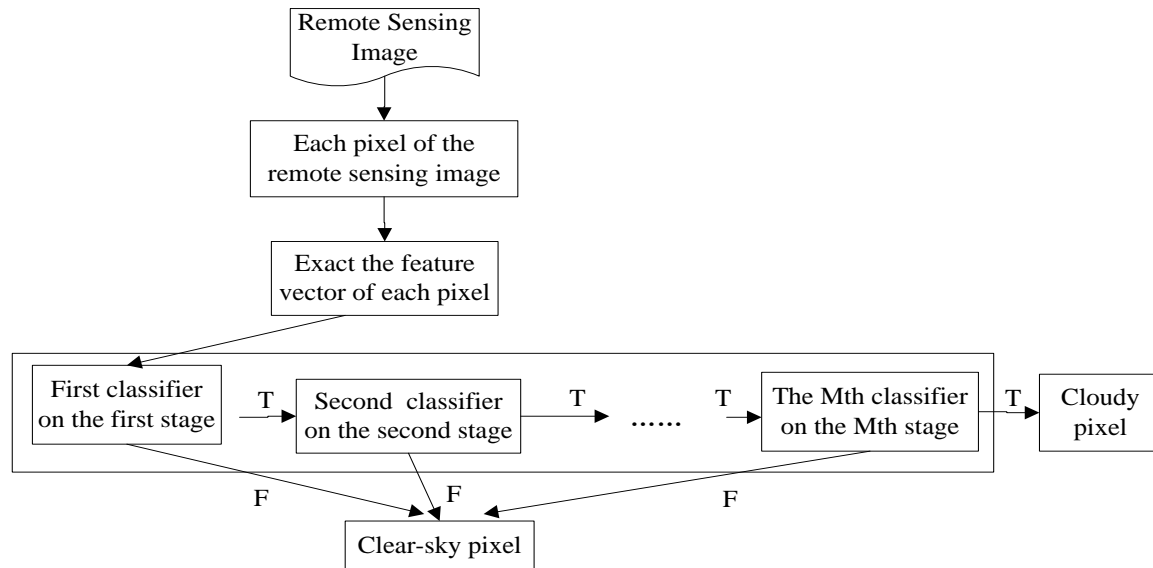
Figure 2. The main procedures of the cascaded AdaBoost detector model applied in the cloud detection

4. Experiments

4.1. Feature Selection of input data

As mentioned before, the training data are pixels which come from the Landsat-5 remote sensing images. In order to keep the universality of our model detector, the training data come from seven Landsat-5 remote sensing images of different districts (such as Japan, US and China) and various kinds of land cover(water, mountain, sea, city, vegetation and so on). Those pixels of the training data are label with "cloudy pixel" and "clear-sky pixel", corresponding to positive example data and negative example data. In this paper, the number of "cloudy pixel" and "clear-sky pixel" is 64,577 and 155,717. The total training data number is 220,254.

Feature extraction algorithm is the basis of cloud detection. To describe the visual content of the image, we implement the colour, texture and spectral descriptor. In this paper, digital number (DN) values are converted to Top of Atmosphere (TOA) reflectance for Band 1, 2, 3, 4, 5, 7 and Band6 Brightness Temperature (BT). In additional the TOA reflectance and BT values, the NDVI, NDSI, Whiteness, HOT and Ratio4_5 features are used to describe the input data [3]. The dimension of the feature vector is 12. The relevance description can be obtained by the following equation:

$$NDVI = (Band\,2 - Band\,5)/(Band\,2 + Band\,5)$$

$$NDSI = (Band\,4 - Band\,3)/(Band\,4 + Band\,3)$$

$$Whiteness = \sum_{i=1}^{3} |\,(Bandi - (Band1 + Band2 + band3)/3)/((Band1 + Band2 + band3)/3)\,|$$

$$HOT = Band1 - 0.5 \bullet Band3 - 0.08$$

$$Ratio\,4\_5 = Band\,4\,/\,Band\,5$$

(6)

4.2. System implementation details

Our implementation platform was carried out on the Matlab2010b environment. The empirical evaluation was performed on Dell3G memory Pc with Win7 operating system. The detail parameter setting in the cascaded AdaBoost detector is illustrated in the following: the fixed maximum false drop rate (f_{max}) and minimum detection rate (d_{min}) for each stage are set as 0.5 and 0.99. The overall target false drop rate (F_{target}) for the detector classifier model is set to 0.00001.

4.3. Experimental results

The results obtained by the new method of cloud detection based on cascaded AdaBoost, was compared with the stronger classifier of the AdaBoost algorithm and the object-based cloud detection

model [3]. The precision index was used as the performance measurement index for the different methods on the cloud detection. In order to argue the precision index on the test data, the input data was divided to training data and test data which are independent. The number of training data and test data is 110127 respectively. The precision index of test data with the object-based cloud detection model [3] is 92.03%.

Table 1 shows the precision index of AdaBoost algorithm and cascaded AdaBoost algorithm when the number of training data increases in the cloud detection processing. The result is the average precision index on 10 times experiments. It shows that the precision index using AdaBoost algorithm and cascaded AdaBoost algorithm is higher than the object-based cloud detection model result, especial with the increasing of the training data number. The precision index on the test data using cascaded AdaBoost algorithm is higher than using AdaBoost algorithm, when the number of training data is more than 60000. Above all, the precision index on the test data using cascaded AdaBoost algorithm is as high as 98.61%.

Table 1. Table of the Precision Index of Adaboost Algorithm and Cascaded Adaboost Algorithm in the Cloud Detection.

No.	Number of training data	the precision index of training data with Adaboost algorithm	the precision index of testing data with Adaboost algorithm	the precision index of training data with cascaded Adaboost algorithm	the precision index of testing data with cascaded Adaboost algorithm
1	50	1	0.94489948	1	0.9284846
2	100	1	0.95983237	1	0.9496586
3	200	1	0.96803752	1	0.9651045
4	300	0.999666667	0.97004159	0.9996667	0.9700852
5	500	0.9992	0.97375279	0.9948	0.974069
6	1000	0.9988	0.97919202	0.994	0.9779474
7	2000	0.99405	0.9818408	0.99105	0.9803094
8	3000	0.9905	0.98250005	0.9894	0.9808302
9	4000	0.98975	0.98303398	0.988325	0.9816387
10	5000	0.98844	0.98349073	0.98726	0.9817343
11	10000	0.98693	0.98456859	0.98582	0.9819735
12	20000	0.98616	0.98496631	0.983325	0.9809838
13	30000	0.985846667	0.98487278	0.9839367	0.9818647
14	40000	0.9856675	0.98518152	0.982725	0.981374
15	50000	0.985758	0.9851906	0.985574	0.9840268
16	60000	0.985811667	0.98525235	0.9865083	0.9854096
17	70000	0.985614286	0.98531864	0.98607	0.9848044
18	80000	0.9857075	0.98534315	0.9859375	0.984746
19	90000	0.985586667	0.98527323	0.9862567	0.9853626
20	100000	0.985694	0.9854612	0.986536	0.9857071

5. Conclusion
In this study, we argue that the cascaded AdaBoost algorithm was successfully applied on the cloud detection of the remote sensing images. In this paper, multi-features based on the color, texture and spectral features are extracted for the detection image pixels and the training data examples. Then, the cascaded AdaBoost detector model was trained by training data. Last, the "cloudy pixels" of the remote sensing image can be detected according the model. The new model for cloud detection is fast, accurate and robust. While the training phase is time consuming due to the large number of samples and the features, the resulting detector is very fast. Training with a large variety of training data from various kinds of clouds made the detector robust to some rotations, occlusions and also to the degradation of image quality to a significant extent. It also does not require any manual intervention. Thus, it is suitable for real-time surveillance and non-intrusive applications. However, there are some improvements for the model in the future. First, much more common features for the remote sensing data should be exacted in order to make the model applicable to different images. Second, the performance of the detector might further be enhanced by training with samples normalized to an

input pattern size common to the target application. This will reduce the amount of rescaling of the trained features that affects the generalization performance of AdaBoost.

References
[1]   Bandyopadhyay S, Maulik U, Mukhopadhyay A 2007 Multiobjective genetic clustering for pixel classification in remote sensing imagery *Geoscience and Remote Sensing, IEEE Transactions on,* 45(5) pp 506-15
[2]   Hagolle O, Huc M, Pascual D V, et al. 2010 A multi-temporal method for cloud detection, applied to FORMOSAT-2, VENμS, LANDSAT and SENTINEL-2 images *Remote Sensing of Environment* 114(8) pp 1747-1755
[3]   Zhu Zhe, Woodcock, Curtis E 2012 Object-based cloud and cloud shadow detection in Landsat imagery *Remote Sensing of Environment* vol 118 pp 83-94
[4]   Hollingsworth B V, Chen L, Reichenbach S E, et al. 1996 Automated cloud cover assessment for Landsat TM images *SPIE's 1996 International Symposium on Optical Science, Engineering, and Instrumentation. International Society for Optics and Photonics* pp 170-179
[5]   Ghonima M S, Urquhart B, Chow C W, et al. 2012 A method for cloud detection and opacity classification based on ground based sky imagery *Atmospheric Measurement Techniques Discussions* 5(4) pp 4535-4569
[6]   Irish R R, Barker J L, Goward S N, et al. 2006 Characterization of the Landsat-7 ETM+ automated cloud-cover assessment (ACCA) algorithm Photogrammetric *engineering and remote sensing* 72(10) pp 1179-1188.
[7]   He X Y, Hu J B, Chen W, et al. 2010 Haze removal based on advanced haze-optimized transformation (AHOT) for multispectral imagery *International Journal of Remote Sensing* 31(20) pp 5331-5348
[8]   Guo H T, Wang Y, Liu X P, Jiang Z H 2010 Integrated optimal Method of Cloud Detection with Meteorological Satellite Data *Journal of PLA University of Science and Technology(Natural Science Edition)* 11(2) pp 221-227
[9]   Wang J C, Yang S Z, Ma J J, Wang X B 2006 Automatic Cloud Detection Applied to MODIS Image in the Southeast of China *Geomatics and Information Science of Wuhan University* 31(3) pp 270-273
[10] Wen X F, Dong X Y, Liu L M 2009 Cloud Index Method for Cloud Detection *Geomatics and Information Science of Wuhan University* 34(7) pp 838- 841
[11] Yang L K, Xie Y J, Liang H Y 2013 Cloud Detection on HJ-1B Data Based on Multispectral Analysis *Science of Surveying and Mapping*
[12] Li Y, Zhang X F 2009 Comparisonand Daytime Cloud Detection from MODIS Data Using a Threshold Rule Based Approach *Acta Scientiarum Naturalium Universitatis Pekinensis* 45(6) pp 1003- 1011
[13] Bi Y 2007 The Combination of Pixel Wise Clustering and AdaBoost Algorithm for Face Tracking *National University of Defense Technology*
[14] Wu B, Liu J, Wang H Q, Wu Y R 2008 A Method for Automatic Object Extraction in High-resolution Remote Sensing Image *Journal of Electronics & Information Technology 30*(11) pp 2732- 2736
[15] Islam S M S, Bennamoun M, Davies R. 2008 Fast and fully automatic ear detection using cascaded adaboost *Applications of Computer Vision, 2008. WACV 2008. IEEE Workshop on. IEEE,* pp1-6.