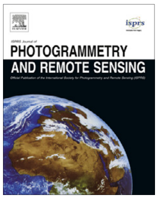
ISPRS Journal of Photogrammetry and Remote Sensing 124 (2017) 70–88

Contents lists available at ScienceDirect

ISPRS Journal of Photogrammetry and Remote Sensing

journal homepage: www.elsevier.com/locate/isprsjprs



A cloud detection algorithm-generating method for remote sensing data at visible to short-wave infrared wavelengths

Lin Sun [a,*], Xueting Mi [a,*], Jing Wei [a], Jian Wang [b], Xinpeng Tian [a], Huiyong Yu [a], Ping Gan [a]

[a] *Geomatics College, Shandong University of Science and Technology, Qingdao, Shandong 266590, China*
[b] *School of Geography, Beijing Normal University, Beijing 100875, China*

ARTICLE INFO

Article history:
Received 19 March 2016
Received in revised form 28 September 2016
Accepted 14 December 2016
Available online 31 December 2016

Keywords:
Hyperspectral sensor
Multispectral sensors
Pixel dataset
Data simulation
Cloud detection
Cloud possibility

ABSTRACT

To realize highly precise and automatic cloud detection from multi-sensors, this paper proposes a cloud detection algorithm-generating (CDAG) method for remote sensing data from visible to short-wave infrared (SWIR) bands. Hyperspectral remote sensing data with high spatial resolution were collected and used as a pixel dataset of cloudy and clear skies. In this paper, multi-temporal AVIRIS (Airborne Visible/Infrared Imaging Spectrometer) data with 224 bands at visible to SWIR wavelengths and a 20 m spatial resolution were used for the dataset. Based on the pixel dataset, pixels of different types of clouds and land cover were distinguished artificially and used for the simulation of multispectral sensors. Cloud detection algorithms for the multispectral remote sensing sensors were then generated based on the spectral differences between the cloudy and clear-sky pixels distinguished previously. The possibility of assigning a pixel as cloudy was calculated based on the reliability of each method. Landsat 8 OLI (Operational Land Imager), MODIS (Moderate Resolution Imaging Spectroradiometer) Terra and Suomi NPP VIIRS (Visible/Infrared Imaging Radiometer) were used for the cloud detection test with the CDAG method, and the results from each sensor were compared with the corresponding artificial results, demonstrating an accurate detection rate of more than 85%.

1. Introduction

Covering more than 50% of the global surface, clouds play an important role in the radiative balance of the planet because of their absorption and scattering of solar and infrared radiation (Harshvardhan et al., 1989; King et al., 2013). Clouds influence the energy transfer between the sun and earth's surface by blocking radiation between them, causing the information in a cloud-covered region lost or blurred and reducing the data utilization rate (Carslaw et al., 2002; Goodwin et al., 2013; Hagolle et al., 2010; Yang et al., 2012; Sun et al., 2016). Cloud cover greatly affects the accuracy and reliability of parameter inversion, blocking the combination of different detected images and the quantitative inversion of detailed surface parameters (Kazantzidis et al., 2011, 2012; Li et al., 2011; Sun et al., 2016). It is important to acquire cloudless images automatically because of the large amount required for multi-temporal and large-scale studies. However, this acquisition is difficult due to climatic factors and a large observation range (Greenhough et al., 2005; Li et al., 2016). To date, only some microwave sensors can eliminate cloud cover and obtain qualified images in the acquisition of geospatial information through remote sensing technology (Chang and Li, 2005; Yang et al., 2013). Thus, the elimination of cloud cover regions by an efficient detection method is an urgent priority in remote sensing data processing (Nakajima et al., 2011; Sun et al., 2010).

A variety of cloud detection methods have been developed over the years, of which the threshold, radiative transfer and statistical methods are the most notable. The threshold cloud detection method has been widely applied because of its simple algorithm, fast operation and high precision of detection and is the most widely used cloud detection method (Hagolle et al., 2010; Jedlovec et al., 2008; Zhu et al., 2015). The most representative methods include ISCCP (International Satellite Cloud Climatology Project), CLAVR (CLouds from the Advanced Very High Resolution Radiometer (AVHRR), NOAA) and APOLLO (AVHRR Processing scheme Over cLouds, Land and Ocean) (Kriebel et al., 2003; Murino et al., 2014).

The ISCCP algorithm described by Rossow et al. (1985), Rossow and Schiffer (1991), Rossow and Garder (1993) assumes that the observed visible and infrared radiances are caused by only two

---

* Corresponding authors.
E-mail addresses: sunlin6@126.com (L. Sun), mixueting@yeah.net (X. Mi).

http://dx.doi.org/10.1016/j.isprsjprs.2016.12.005
0924-2716/

types of conditions, cloudy and clear, and that the ranges of radiances and the variability associated with these two conditions do not overlap (Rossow and Garder, 1993). Comparing the observed pixel values with those of clear sky through a visible band of 0.6 μm and an infrared window of 11 μm, a cloud pixel can be identified if at least one radiance value is distinct from the inferred clear value by an amount larger than the uncertainty in the clear threshold value. Uncertainty can be caused both by measurement errors and natural variability. This algorithm was constructed to be cloud-conservative, minimizing false cloud detections but missing clouds that resemble clear conditions (Zhu et al., 2015; Sun et al., 2016).

The NOAA CLAVR algorithm (Stowe et al., 1991), adopting a pixel matrix of $2 * 2$ as the identification unit, can be used for global cloud detection. A cloud pixel is identified when all four pixels in the $2 * 2$ pixel matrix pass the cloud detection test; otherwise, clear sky is identified. When only 1–3 pixels pass the test, the pixel matrix is mixed. A pixel matrix can be identified as clear sky when the pixels in a cloud or mixed matrix meet the conditions for ice/snow, oceanic specular reflection or a bright desert background. The algorithm is divided into four classes according to the underlying surface properties and observation time: ocean and land during the day and night, respectively (Greenhough et al., 2005; Kriebel et al., 2003; Lin et al., 2015).

The APOLLO algorithm described by Saunders and Kriebel (1988), Kriebel et al. (1989) and Gesell (1989) uses the detection data from the five full-resolution AVHRR channels to set a threshold in each channel according to the object characteristics to identify cloudy pixels. Based on the data, a pixel is identified as a cloud pixel when the threshold set by the reflectance is higher or the threshold set by temperature is lower. A pixel can be identified as clear sky if the reflectance ratio of channels 2 and 1 is between 0.7 and 1.1, the temperature difference between channels 4 and 5 is above a set threshold, and the spatial uniformity over the ocean is greater than a set threshold (Saunders and Kriebel, 1988); otherwise, the pixel is identified as being polluted by clouds. Completely covered cloudy pixels can be distinguished from partially covered pixels if they are tested with different thresholds (Kriebel et al., 2003; Sun et al., 2016).

In addition, other methods that are also widely used for cloud detection from satellite data. Zhang et al. developed a haze optimized transformation (HOT) algorithm to detect hazy/cloud spatial distributions in Landsat images, and this transformation was designed to minimize terrestrial surface effects (Zhang et al., 2002; Zhang and Guindon, 2003). A scene-average Automated Cloud Cover Assessment (ACCA) algorithm has been used for Landsat-7 ETM+ data since its launch by NASA in 1999; by applying a number of spectral filters, and the ACCA algorithm can well estimate the overall percentage of clouds in each Landsat scene (Irish et al., 2006). Hégarat-Mascle and André (2009) proposed an automatic detection algorithm for cloud and cloud shadow on visible high resolution optical images based on the use of Markov Random Field formalism on graphs and performed well on the experiments with fast convergence and low false alarm rates. Li et al. (2012) developed a new variational gradient-based fusion method for cloud detection with visible and short-wave infrared (SWIR) bands that enables spatial enhancement and dehazing of visible imagery. A spatial-temporal reflectance fusion method was also proposed for cloud detection by using the difference between the spatial and temporal information. Sedano et al. (2011) presented a new cloud detection method for high resolution remote sensing data based on pixel-based seed identification and object-based region growing. Evaluation results showed that it can get a high detection accuracy for clouds of different types over various land surfaces. Lin et al. (2015) introduced a new method for radiometric normalization and cloud detection of optical satellite images based on the major axis of a scatterplot and the invariant pixels which can be determined by the proposed weighted PCA technology. Evaluation results indicated that it can perform well for multi-temporal images containing various landscapes. Sun et al. (2016) proposed a universal dynamic threshold cloud detection algorithm (UDTCDA) supported by a priori surface reflectance database and was applied to MODIS and Landsat 8 images for cloud test. Evaluation results showed that it can reduce the effects of mixed pixels and complex surface structures and obtain a high precision in cloud detection.

Clouds are generally characterized by a higher reflectance and lower temperature than the underlying earth surface. As such, simple visible and infrared window threshold approaches perform considerably well in cloud detection (Ackerman et al., 2010). In the threshold method, the determination of the threshold is a key factor that affects the classification accuracy. Traditional cloud detection thresholds are determined by the surface reflectance difference between the underlying surfaces and clouds, whereas the detection data are the apparent reflectance; however, the influence of the atmosphere in the radiative transfer process makes the surface reflectance different from the apparent reflectance. The difference in reflectivity and point-to-pixel expansion make the threshold unable to be effectively determined and affect the detection precision. In addition, we must study different cloud detection algorithms for different sensor data, which not only takes great effort but also reduces the application scope of the data, which is not conducive to future work. The automatic detection of clouds is not a simple task. A changing solar elevation angle and instrument viewing angles, limited spectral channels, instrument noise, and varying surface properties often limit the success of traditional cloud detection schemes when applied over a large area (Jedlovec et al., 2008). With the development and progress of science and technology, the realization of automated cloud detection has a wide range of applications and remote sensing requirements.

AVIRIS (Airborne Visible/Infrared Imaging Spectrometer) has measured spectral images for scientific research and applications every year since 1987 and was the first imaging spectrometer to measure the reflected solar spectrum from 400 nm to 2500 nm. AVIRIS measures the upwelling radiance through 224 contiguous spectral channels at 10 nm intervals across the spectrum. In the 400–2500 nm spectral range, molecules and particles from terrestrial, aquatic and atmospheric environments interact with solar energy through absorption, reflection, and scattering processes; therefore, this wavelength range is most commonly used for cloud detection (Green et al., 1998). AVIRIS has characteristics of high spatial resolution, spectral resolution and calibration accuracy. The spectral, radiometric, and spatial calibration of AVIRIS was determined in the laboratory and monitored inflight each year. More than 4 TB of AVIRIS data have been acquired since the initial flights that covered most parts of North America (Fig. 1), including vegetation, towns, lakes, bare soil, snow, glaciers and other land surface types (Green et al., 1998; Van der Meer, 1994). AVIRIS is currently the major source of high-spectral-resolution images. Fig. 2 shows the apparent reflectance differences between clouds and typical surfaces within the spectral range of 400–2500 nm of AVIRIS data.

Aiming at the above mentioned problems, this paper puts forward a cloud detection algorithm generating (CDAG) method for remote sensing data obtained at visible to SWIR wavelengths based on hyperspectral and high-spatial-resolution remote sensing data from AVIRIS. Most cloud detection algorithms judge the existence of clouds based on the results obtained from many cloud detection tests, using multispectral data acquired by sensors. In fact, an imager that has more bands can be expected to detect clouds efficiently and accurately (Nakajima et al., 2011). The CDAG algorithms make full use of every band in the visible to SWIR