ARTICLE IN PRESS

Remote Sensing of Environment xxx (xxxx) xxx–xxx

Contents lists available at ScienceDirect

Remote Sensing of Environment

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

Improvement of the Fmask algorithm for Sentinel-2 images: Separating clouds from bright surfaces based on parallax effects

David Frantz*,[1], Erik Haß[2], Andreas Uhl[3], Johannes Stoffels, Joachim Hill

Environmental Remote Sensing and Geoinformatics, Faculty of Regional and Environmental Sciences, Trier University, 54286 Trier, Germany

ARTICLE INFO

Keywords:
Cloud detection
Fmask
MSI
Parallax
Sentinel-2
View geometry

ABSTRACT

Reliable identification of clouds is necessary for any type of optical remote sensing image analysis, especially in operational and fully automatic setups. One of the most elaborated and widespread algorithms, namely Fmask, was initially developed for the Landsat suite of satellites. Despite their similarity, application to Sentinel-2 imagery is currently hampered by the unavailability of a thermal band, and although results can be improved when taking the cirrus band into account, Sentinel-2 cloud detections are unsatisfactory in two points. (1) Low altitude clouds can be undetectable in the cirrus band, and (2) bright land surfaces – especially built-up structures – are often misclassified as clouds when only considering spectral information. In this paper, we present the Cloud Displacement Index (CDI), which makes use of the three highly correlated near infrared bands that are observed with different view angles. Hence, elevated objects like clouds are observed under a parallax and can be reliably separated from bright ground objects. We compare CDI with the currently used cloud probabilities, and propose how to integrate this new functionality into the Fmask algorithm. We validate the approach using test images over metropolitan areas covering a wide variety of global environments and climates, indicating the successful separation of clouds and built-up structures (overall accuracy 95%, i.e. an improvement in overall accuracy of 0.29–0.39 compared to the previous Fmask versions over the 20 test sites), and hence a full compensation for a missing thermal band.

1. Introduction

Cloud detection is inevitably required for any earth surface-related usage of optical remote sensing imagery like Landsat and Sentinel-2 data. If not accounted for, clouds adversely influence virtually any image analysis like atmospheric correction or land cover classification (Zhu and Woodcock, 2012). Nevertheless, the fully automatic detection of clouds is not trivial, partly due to the high variability in reflectance and temperature of both land surfaces and clouds (Irish, 2000). As such, historically, cloud masks were often generated by hand, which is a very labor- and cost-intensive step, only feasible for few images. With the advent of increasing volumes of freely available satellite data (e.g. the opening of the Landsat archive; Woodcock et al., 2008), more and more automatic and accurate cloud detection codes evolved, which simultaneously paved the way for the automatic generation of higher-level earth observation products (e.g. Flood et al., 2013; Frantz et al., 2016; USGS, 2017) and an entirely new usage of the data for both large area and time series analyses simultaneously (Wulder et al., 2012).

In general, cloud detection codes for Landsat-like imagery can be grouped into mono- and multi-temporal approaches. Multi-temporal approaches are advantageous because they can isolate transient changes superimposed on a more stable background signal. Multi-temporal methods include bi-temporal change detection (e.g. Wang et al., 1999) and time series approaches (e.g. Frantz et al., 2015; Goodwin et al., 2013; Hagolle et al., 2010). While detection accuracies are often improved compared to mono-temporal methods (e.g. Goodwin et al., 2013), their inclusion in most Level 2 production systems is not feasible as these are mono-temporal in nature, thus there is still a pressing need for single-date cloud masks.

Mono-temporal cloud detection in Landsat images was initially performed with the automated cloud cover assessment system (ACCA, Irish, 2000; Irish et al., 2006). While the overall cloud contamination was well estimated, ACCA generally failed to identify the exact locations and boundaries of clouds needed for automatic analysis of the data (Zhu and Woodcock, 2012). As such, a number of other techniques were developed over the years (e.g. Choi and Bindschadler, 2004;

* Corresponding author.
E-mail address: david.frantz@geo.hu-berlin.de (D. Frantz).
[1] Present address: Geomatics Lab, Geography Department, Humboldt-Universität zu Berlin, Unter den Linden 6, 10099 Berlin, Germany.
[2] Present address: Robert Bosch GmbH, Future Systems Industrial Technology - Mobile Machinery Systems (CR/AEI1), Robert-Bosch-Campus 1, 71272 Renningen, Germany.
[3] Present address: Image Processing and Data Distribution Services, GAF AG, 80634 Munich, Germany.

https://doi.org/10.1016/j.rse.2018.04.046
Received 7 April 2017; Received in revised form 12 March 2018; Accepted 25 April 2018
0034-4257/

Hansen et al., 2008; Vermote et al., 2016). However, the probably most successful and elaborated algorithm for Landsat-like data has been the "Function of mask" algorithm (Fmask, Zhu and Woodcock, 2012), which marked an important game changer for the automatic processing and analysis of medium resolution optical imagery and was hence integrated into several Landsat Level 2 production environments (e.g. Flood et al., 2013; Frantz et al., 2016; USGS, 2017) and fully enabled automatic analysis of many data for a wide range of research questions (Bleyhl et al., 2017; Griffiths et al., 2014; Griffiths et al., 2013; Müller et al., 2016; Schmidt et al., 2016; Schneibel et al., 2017a, b; Senf et al., 2017; Zhu et al., 2012).

The accuracy of the Fmask results are generally good on Landsat data: cloud overall accuracy of 96.41%, cloud producer's accuracy of 92.1%, and cloud user's accuracy of 89.4% (Zhu and Woodcock, 2012). In Landsat images, Fmask is generally able to separate land surfaces from clouds. In a first step, a range of spectral tests are used to generate a potential cloud pixel (PCP) layer, which is anticipated to contain all clouds but also some bright clear-sky pixels – merely built-up objects. These false positives are eliminated by computing a cloud probability from all clear pixels in order to estimate a scene-based threshold. In the original Fmask (Zhu and Woodcock, 2012) – hereby defined as $Fmask_{2012}$ – the cloud probability over land is a combination of a temperature and a variability probability. The temperature probability is very effective because clouds are typically colder than the subjacent land surface. The variability probability combines spectral indices from the visible (VIS), near infrared (NIR) and shortwave infrared (SWIR) because clouds have fairly similar reflectance in this part of the spectrum. Since Sentinel-2 is not equipped with a thermal sensor, the cloud probability becomes the variability probability only, hereby defined as CP_{2012}.

Fmask was recently updated (Zhu et al., 2015) – hereby defined as $Fmask_{2015}$ – to also make use of the new cirrus band carried by Landsat 8, and it was demonstrated that the cirrus band can partially account for a missing thermal band. The new cloud probability over land (CP_{2015}) is defined as the sum of the variability probability and a cirrus probability, and can be readily applied to Sentinel-2 imagery.

However, Sentinel-2's cirrus band @ 1.375 μm is located in a strong water absorption band, which only observes the upper layer of the atmosphere (Hagolle et al., 2010). Consequently, the cirrus band is very helpful to detect high altitude cirrus clouds (Zhu et al., 2015), but low to mid altitude clouds are indistinctive and are susceptible to be removed in the cloud probability routine. In addition, many built-up structures – especially artificial materials – end up in the PCP layer and are inseparable because they appear indistinctive in the variability probability. Artificial materials can be very variable in the spectral range covered by Sentinel-2 and – like clouds – can be bright throughout the complete spectrum. In Fmask, bare soils are removed using a NIR to SWIR ratio as reflectance generally increases from NIR to SWIR. However, this is not necessarily the case for bright artificial materials, which results in many false positives in industrial and residential areas.

Hence, without a thermal band, low altitude clouds are susceptible to be omitted and built-up areas often remain as artifacts in the cloud mask. As Sentinel-2's spectral bands fail to succeed at separating artificial materials from clouds with high accuracy, we consequently propose to tackle this problem with an innovative approach that exploits Sentinel-2's unique sensor configuration. This approach is not solely reliant on spectral properties but specifically incorporates view angle effects.

1.1. Background: the S2A view geometry exploit

Sentinel-2A's Multi Spectral Instrument (MSI) is a push-broom sensor with 13 spectral bands that cover the VIS, NIR and SWIR domains (Drusch et al., 2012). Three of these bands have a spatial resolution of 60 m and are mainly intended for atmospheric characterization. The remaining 10 bands are provided at 10–20 m spatial resolution. Among the spectral domains, the NIR plateau is of special interest as there are three spectrally highly correlated bands available, which partially overlap (bands 7, 8 and 8A with central wavelength at 0.782 μm, 0.835 μm and 0.865 μm, respectively); Table 1 gives the correlation matrix between the NIR bands for a cloud-free acquisition (23 Aug 2016, relative orbit 108); note that the 10 m bands were reduced to 20 m using nearest neighbour resampling. All 15 tiles within the product were analysed in order to cover the complete field-of-view (FOV; upper-left: T31UGS, lower-right: T32UPV). The MSI is characterized by a complex sensor arrangement: for each band, twelve detectors are arranged in a staggered configuration to cover the wide FOV (Drusch et al., 2012); Fig. 1(*a*) displays the viewing vectors (average viewing geometry for each of the 12 detectors) for an across-track scanline; the along-track flight vector and Nadir line are superimposed in black. As a result of the push-broom concept, a parallax exists between odd and even detectors – and although less pronounced, there is also a parallax between bands (Gascon et al., 2017): the viewing vectors of different bands point to different locations on the ground (Fig. 1(*a*)). As this shift is systematic for all stationary objects with known altitude, it can be accounted for during systematic (including flight path adjustment) and geometric correction, as well as in the relative calibration of the focal planes when using a DEM (Gascon et al., 2017). Thus, in Level 1 products, displacement effects are small for objects on the land surface (< 0.3 pixels, Gascon et al., 2017), including mountainous areas. However, non-stationary objects with unknown altitude (like clouds) cannot be corrected this way, hence a displacement is still visible in the final Level 1 products. Table 2 gives the mean (below diagonal) and maximum (above diagonal) of view azimuth differences between the NIR bands. Most strikingly, highly correlated NIR bands 8 and 8A look in different directions (μ: 10.3°, max: 27.5°), whereas bands 8A and 7 are more similar (μ: 1.3°, max: 2.8°); see also Fig. 1(*b*,*c*). For most applications, this sensor design might affect the quantitative analysis of surface reflectance properties. However, we propose to exploit the Sentinel-2 detector arrangement, where three spectrally similar bands are observed under different viewing geometries, for enhanced cloud detection. While objects on the land surface are registered to the same position, objects above the land surface are projected onto slightly different locations in the focal plane – and remain in different positions after systematic and geometric correction.

Table 1
Mean correlation matrix for a cloud-free Sentinel-2 acquisition for the bands on the NIR plateau (23 Aug 2016, West-Germany); individual correlation matrices were computed for all 15 tiles, then averaged.

	λ	7	8[b]	8A
7	0.782		0.948	0.991
8[b]	0.835	0.948		0.949
8A	0.865	0.991	0.949	

[b] 10 m band.

1.2. Objectives

In this study, we propose to

- exploit the Sentinel-2 NIR parallax to separate clouds from artificial surfaces,
- demonstrate the superiority of this approach against the probabilistic approach currently used in Fmask (in absence of a thermal band),
- and propose how to integrate the parallax approach into Fmask.

The next section will outline the data used (Section 2), followed by theoretical considerations on how elevated objects are projected to the