Temporal Repetition Detector for Time Series of Spectrally Limited Satellite Imagers

Tristan Dagobert[1], Rafael Grompone von Gioi[1], Jean-Michel Morel[1], Carlo de Franchis[1,2]

[1] CMLA, ENS Paris-Saclay, France (`tdagobert@hotmail.com`, {`grompone`, `jean-michel.morel`, `carlo.de-franchis`}`@ens-paris-saclay.fr`)
[2] Kayrros, France

Communicated by Jean-Michel Morel   *Demo edited by* Tristan Dagobert

Abstract

This article addresses the problem of estimating scene visibility in time series of satellite images. It focuses on satellites with few spectral bands and high revisit frequency. Our approach exploits the redundancy of information acquired during these revisits. It is based on an unsupervised algorithm that tracks local ground textures across time and detects ruptures caused mainly by opaque clouds and in some cases by haze, cirrus and shadows. Experiments have been carried out on 18 PlanetScope image times series of various locations. These time series come with hand-made ground truth labels that are published together with this paper. We compare our results with the Unusable Data Masks (UDM) that Planet provides together with the images, and demonstrate the effectiveness of the proposed method: success rates of 97.78% and 89.36% are reached for the visible and occluded regions classification. This article is related to the following publication: [Tristan Dagobert, Jean-Michel Morel, Carlo de Franchis and Rafael Grompone von Gioi, Visibility detection in time series of Planetscope images, IEEE International Geoscience And Remote Sensing Symposium, 2019].

Keywords: satellite; cloud; shadow; PlanetScope; multi-temporal; SIFT

Source Code

The C source code implementing the algorithm described in the paper, an online demo, and the ground truth data used in the experiments are accessible at the associated web page[1].

---

[1] https://doi.org/10.5201/ipol.2020.245

1 Introduction

Evaluating visibility is a crucial preliminary step for optical satellite images analysis. Ground visibility is very often hindered by clouds and shadows. Haze can also cause blur and considerable signal attenuation. At any time, clouds and haze hide most of the surface of the earth (around 67% according to King et al. [15]), hence their detection is critical to avoid errors in automatic image analysis. All zones with no – or low – visibility have to be detected accurately in every image prior to further automatic analysis. Very often, this problem is presented as a cloud detection problem or as a classification of image pixels into ground, clouds, cirrus, snow, haze, etc. But what matters the most for remote sensing applications is to know where in the image the ground is visible and unaltered by atmospheric perturbations. Thus in this work we drop the classification aspect of the problem and focus on detecting the visible parts of the ground as accurately as possible.

Over the past decade, an increasing number of satellites for earth observation have been launched. This growing number of satellites allows to reduce the revisit time over the same region and to study accurately the temporal changes of the earth surface. Furthermore, the images are often taken repeatedly from the same viewpoint. This opens the way to reliable change detection and therefore to automatic earth monitoring, provided that the effects of atmospheric changes are handled. While some sensors have dedicated spectral bands to detect clouds on single images (e.g. the cirrus band on Sentinel-2), many multi-spectral sensors are limited to the usual visible bands (red, green, blue, near infra-red) where cloud detection can be very challenging.

In this study, we propose to detect the ground visibility stability based on the temporal information accumulated through repeated satellite passes to compensate for the lack of spectral information. Our main applications are the PlanetScope[2], WorldView[3] and Pléiades[4] missions which provide recurrent imagery. These satellites are characterized by a high spatial resolution but a limited number of spectral bands (see Table 1).

Constellation	PlanetScope	Rapideye	WorldView-{1,2,3}	Pléiades
Satellites	120+	5	1, 1, 1	2
Resolution	3 m	5 m	0.46 m, 0.46 m, 0.31 m	0.7 m
Spectral bands	B $(440-510$ nm) G $(520-590$ nm) R $(630-685$ nm) NIR $(760-850$ nm)	B $(440-510$ nm) G $(520-590$ nm) R $(630-685$ nm) R Edge $(690-730$ nm) NIR $(760-850$ nm)	Pan $(400-900$ nm) B $(450-510$ nm) G $(510-580$ nm) R $(655-690$ nm) NIR $(780-920$ nm)	Pan $(400-900$ nm) B $(430-550$ nm) G $(490-610$ nm) R $(600-720$ nm) NIR $(750-950$ nm)
Revisit time	1 day	1 day	1.7 day, 1.1 day, 1 day	1 day

Table 1: Imaging characteristics of some of the main optical satellite constellations. The revisit time is the shortest possible delay that separates two acquisitions over the same geographic location.

Our objective can be reformulated as a problem of abrupt temporal disturbances detection over the ground, which are mainly induced by clouds and in some cases by haze and shadows. This approach excludes consequently two cases: the season variations and the sea changes. On the first topic, a lasting state is not considered as a visibility disturbance *e.g.* a valley observed in the shadow of a mountain during spring. A lower SNR may hinder fine change detection, but change detection remains possible. In the same way, the temporary presence of snow cover is a seasonal event that is not considered as hindering visibility. It is simply a visible change of the ground aspect. On the second topic, seas, whatever their visibility state, are in continual change. This advocates for

---

[2]http://www.planet.com
[3]http://www.digitalglobe.com
[4]http://pleiades.cnes.fr

a specific definition of the task of sea change observation whose resolution could be found by the construction of particular descriptors. This problem is out of the scope of our study.

Our proposition is to calculate ground visibility by an unsupervised multi-temporal ground matching based algorithm able to learn the "normal" aspects of the ground from long enough time series of registered images. This high dimensional data would not be easily prone to efficient machine learning, even if reliable annotation were available. In addition, our hypothesis about the cloud cover frequency at each pixel is more flexible than that of Zhu and Woodcock [37, § 3.2.2] who assume that each pixel is cloud-free at least 15 times per sequence, while we only suppose that this happens twice per sequence.

The remainder of this paper is organized as follows: in Section 2 we give an overview of the existing cloud detection algorithms. In Section 3 we describe and justify our algorithm to discriminate temporal disturbance from stable ground. Section 4 provides an evaluation of our algorithm on ground truth datasets and a comparison with the Unusable Data Masks (UDM) distributed by Planet with PlanetScope images (see Section 2.3).

2   Related Work

Cloud detection has long been limited to the spatial analysis of a limited number of spectral bands from a single acquisition. In recent years, the use of sensors with additional spectral bands and the analysis of time series have led to notable improvements (see [3]).

2.1   Spatial Methods

Spatial methods rely on features based on spectral or texture information. Spectrum based analyzers are simple functions applied independently to the spectral bands of each pixel [11]. Texture analyzers use classic texture features such as the gray level co-occurrence matrix [34] or a gray level difference vector. They have been developed to discriminate cloud categories from each other [16]. These features are then passed to an appropriate classifier. There are mainly two types of classifiers. The traditional classifiers include Support Vector Machines [26], Bayesian classification [11] and maximum likelihood [35]. The neural-network classifiers are multilayer back-propagation neural networks [30], self organizing maps [28] and probabilistic neural networks [36].

Two algorithms deserve a special mention. The first one is the so-called "Cloud Cover Assessment" (CCA) [12, 13, 28]. It is part of the processing that produces Level-1 Landsat-8 images. The CCA uses several independent algorithms to detect clouds, then merges the several masks into the final Level-1 quality band *via* a weighted voting mechanism. The second mainstream method is the one implemented in the `sen2cor` software [20] for Sentinel-2 images. The algorithm is based on a series of threshold tests on the top-of-atmosphere reflectance from the Sentinel-2 multispectral bands. For each of these tests a confidence level is associated. The algorithm outputs clouds and snow masks with probabilistic quality maps. The author notes that the search for cloud shadows is a more complex problem than cloud detection itself. Indeed, low luminance can happen because of shadows from hills or buildings, or even because of water. This explains why `sen2cor` relies on the estimated projection of cloud shadows, which requires a precise knowledge of the position and altitude of the clouds and of the sun.

2.2   Spatio-Temporal Methods

These methods combine spatial processing over the several images of a times series. The most robust algorithms use spatially and temporally varying thresholds, which better capture local atmospheric

and surface effects [14]. The idea of these algorithms is that clouds, snow or cloud shadows cause sudden changes to the reflectance. Thus, by comparison with a reference image without clouds, snow and cloud shadows, the ones present in the observed image are more easily detected. These algorithms are reported to have higher accuracies than single image methods [37]. However, as pointed out by Goodwin et al. [5], the main limitation of multi-temporal methods is that these algorithms assume that there is no significant land cover change between the reference and the current image, so that differences in reflectance can only result from clouds, snow or cloud shadows. A range of non-cloud related variations in brightness over time may easily be confused with changes due to clouds or shadows. Nevertheless, several approaches have been proposed for limiting these undesirable effects. These methods use a hierarchical and iterative approach where four passes are necessary to discriminate clouds and shadows. At each iteration, and for each pixel, spectral bands are mixed together and thresholded with adequate coefficients. Temporal windows with rank filters are then used to highlight the outliers. Regions considered as clouds are then dilated by a morphological filter. Zhu and Woodcock [37] proposed an algorithm called Tmask (multi-Temporal mask) for automated masking of clouds, snow and cloud shadows in Landsat-7 image time series. This algorithm consists of two steps. The first step is based on a single-image algorithm that initially screens, for each image, most of the clouds, snow and cloud shadows. The second step selects temporal information from the remaining pixels and further improves the masks. For each pixel, three reflectance bands (bands 2, 4, and 5 in Landsat-7 band numbering) are used to estimate a time series model which takes into account the seasonal variations luminance. This regression model corresponds to the first two terms of the Fourier series where coefficients are computed iteratively by the robust iteratively reweighed least squares (IRLS) method. Outliers are identified by their difference to this model according to a fixed threshold.

Goodwin et al. [5] as well as Zhu and Woodcock [37], process images from Landsat-7 whose revisit time is 16 days. The next algorithm we examine has a much shorter periodicity. According to [3], the cloud detector of Vivone et al. [33] using *maximum a posteriori* and a Markov random field statistical framework for both spatial information and cloud motion estimation is one of the best and most efficient methods for cloud classification. Cloud motion is estimated from time series acquired by the satellite SEVIRI of the Meteosat-8 project. Images are recorded every 15 minutes hence land cover changes are not a problem.

Hagolle et al. [7, 9, 8, 10] have developed over the past 10 years the MACCS pipeline, which is now part of the MAJA processor (MACCS ATCOR Joint Algorithm). This processor for cloud detection and atmospheric correction was specifically designed to process time series of high resolution optical images acquired under quasi constant viewing angles. It allows to process time series of Landsat-8 or Sentinel-2 images. Since 2016, it is progressively including methods taken from DLRs ATCOR processor. It is now a collaboration between CNES, DLR and CESBIO, and benefits from ESA funding. Its main feature is to use the multi-temporal information contained in time series to detect the clouds and their shadows, and to estimate the aerosol optical thickness and correct the atmospheric effects (taking into account the adjacency effect and the illumination variations due to topography).

2.3 Methods for Band Limited Satellites

Concerning PlanetScope images, Muller et al. [21] authored an open source project based on the work of Zhu and Woodcock [37]. Guan and Khadka [6] patented an algorithm for RapidEye images. It consists first in identifying cloud seeds using a high-precision low-recall classifier, then clustering the seeds into full clouds, finally computing cloud heights by exploiting the parallax between the bands. The cloud shadow computation is made using sun illumination and sensor viewing geometry.

A patented ML-based algorithm [2] for WorldView images trains cloud and ground dictionaries,

in order to determine whether a given pixel is best represented by words from the cloud dictionary or words from the ground dictionary to make an initial determination. Then a max-flow min-cut segmentation with edge growing is applied to obtain the final cloud/clear masks.

Working with Pléiades-HR images, Latry et al. [17] proposed to merge the information given by the reflectance values of the spectral bands and the parallax estimated thanks to the panchromatic and visible bands shifts, to end up discriminating pixels with an SVM classifier. This parallax cue is not available in PlanetScope images as they are acquired with frame cameras.

Gómez-Chova et al. [4] detect abrupt changes on SPOT-4 time series with linear and nonlinear least squares regression algorithms. Some authors restrict the algorithm to the visible bands of multi-spectral imagers: the RGB channels of Gokturk-2 and RASAT satellites are processed by Özkan et al. [22] with a deep pyramid networks approach; the RGB Sentinel-2 channels are used by Singh et Komodakis [29] to train cyclic consistent generative adversarial networks.

3 Algorithm Description

We consider an input time series (u_n) of N gray level images, defined on the same discrete domain $\Omega = [\![1, W]\!] \times [\![1, H]\!]$ where W (resp. H) is the image width (resp. height). Each image can be a linear combination of the available bands (visible and near-infrared), or simply one of them. Pixels positions are denoted by $\mathbf{x}$. The output of the algorithm is a time series (B_n) of Boolean masks where $B_n(\mathbf{x})$ equals 1 if the ground is hidden by cloud, snow or shadow, and 0 if not.

We call (B_n) a "change series", as these masks indicate with the value 1, the unstable visible parts (masked by clouds, snow or shadows) of the images. Algorithm 1 summarizes the method.

3.1 Hypotheses

Our method relies on two assumptions:

(H_1) The frames (u_n) have been registered up to sub-pixel accuracy.

(H_2) For each pixel $\mathbf{x} \in \Omega$ there are at least two dates n and $m \neq n$ such that the respective neighborhoods $\mathcal{V}_n(\mathbf{x})$ and $\mathcal{V}_m(\mathbf{x})$ are cloudless, where $\mathcal{V}(\mathbf{x})$ denotes a given square neighborhood (e.g. a square patch of size 15×15 pixels centered at $\mathbf{x}$).

Hypothesis (H_1) is essential for robust texture matching. But we often observed that the images supplied by satellite image providers may suffer mis-registrations of up to several pixels. We enforce registration by pre-processing the time series with a fine sub-pixel registration method, such as the method proposed by Rais [25]. This method only fails when an image is fully cloudy, in which case registering accurately is irrelevant anyway.

Hypothesis (H_2) allows the algorithm to detect that an image patch is visible (i.e. not masked by clouds, snow, or shadows) by recognizing it in two different images. The underlying assumption is that temporary phenomena such as clouds, snow and shadows quickly change aspect through time, thus they cannot appear twice in the time series. In an extremely unfavorable situation, a completely clear image contained in a time series of totally cloudy images will be considered as not visible, for lack of match. However on the long run it is safe to assume that any point is seen at least twice under clear conditions.

3.2 Cloud Detection

The texture descriptor. We use the SIFT descriptor [18] because it is robust to a wide family of image perturbations, such as changes of lighting, noise, blur, contrast changes, moderate scene

---

Algorithm 1: Compute_Visibility_Maps$[(u_n), \alpha, \theta, \kappa, a]$

   Input:
   (u_n) the registered image time series,
   $\alpha \in [0, 1]$ the correlation threshold,
   $\theta \leq |\Omega|$ the surface threshold,
   κ the sub-sampling factor,
   a the length of the side of the neighborhood $\mathcal{V}(\mathbf{x})$.
   Output:
   $(\tilde{B}_n)$ the series of Boolean visibility masks.

1 /* Computation of the SIFT descriptors
2 **for** $n \in [\![1, N]\!]$ **do**
3     $\tilde{u}_n \leftarrow$ sub-sample image u_n by a factor κ
4     $\mathbf{s}_n \leftarrow$ compute SIFT descriptors over image $\tilde{u}_n$
5 /* Computation of the distances
6 **for** $\mathbf{x} \in \Omega'$ **do**
7   **for** $n \in [\![1, N-1]\!]$ **do**
8   **for** $m \in [\![n+1, N]\!]$ **do**
9           $d(\mathbf{s}_n(\mathbf{x}), \mathbf{s}_m(\mathbf{x})) \leftarrow$ compute the pair distance using (6)
10   **for** $n \in [\![1, N]\!]$ **do**
11      $C_n(\mathbf{x}) \leftarrow$ find the minimal distance using (3)
12 /* Computation of the visibility maps
13 **for** $n \in [\![1, N]\!]$ **do**
14    $C_n \leftarrow$ upsample C_n by a factor κ with bilinear interpolation
15    $B_n \leftarrow$ binarize C_n with threshold α as in (7)
16    $\tilde{B}_n \leftarrow$ remove regions smaller than θ pixels as in (8)
17 **return** $(\tilde{B}_n)$

---

deformation, while remaining discriminative enough for matching purposes. The SIFT descriptor encodes the spatial gradient distribution in the neighborhood of a key point (x, y) at a given scale σ with a 128-dimensional vector.

Since we look for a local description in a small neighborhood, SIFT descriptors are computed only at their first scale and octave. Moreover, while SIFT descriptors are usually computed only at extrema in the image scale space, we compute them at every vertex of a spatially regular sub-sample grid Ω'. We then define a SIFT descriptors series $(\mathbf{s}_n)$ associated to (u_n) defined spatially on the sub-sample discrete domain $\Omega' = [\![1, W/\kappa]\!] \times [\![1, H/\kappa]\!]$, where κ is the sub-sample factor in both directions. Sub-sampling is used to save computation time. We have formally

$$\mathbf{s}_n : \Omega' \quad \rightarrow \quad [\![0, 255]\!]^{128} \tag{1}$$
$$\mathbf{x} \quad \mapsto \quad \mathbf{s}_n(\mathbf{x}) = [\mathbf{s}_n(\mathbf{x})_1, \ldots, \mathbf{s}_n(\mathbf{x})_{128}]^T. \tag{2}$$

The texture distance. We consider for each pixel $\mathbf{x}$ and date n the lowest distance between the SIFT descriptor $\mathbf{s}_n(\mathbf{x})$ and all the other $\mathbf{s}_m(\mathbf{x})$ ($m \neq n$). We obtain the distance series (C_n) defined by

$$C_n(\mathbf{x}) = \min_{\substack{1 \leq m \leq N \\ m \neq n}} d(\mathbf{s}_n(\mathbf{x}), \mathbf{s}_m(\mathbf{x})), \tag{3}$$

where d denotes the distance between SIFT descriptors.

Several variants for the SIFT descriptor distance d can be considered. In addition to usual distances like the L_1 norm, the correlation or the χ_1 distances, distances more specific to SIFT descriptors have been proposed. Pele and Werman [23] have introduced a distance which takes into account the cross-bin relationships. It represents a thresholded variant of the Earth Mover's Distance by Rubner et al. [27] which requires the application of a max-flow-min-cut algorithm to be solved. Pele and Werman [24] have also presented a quadratic-χ distance which is a variation of the χ_2 distance and which goal is to reduce the impact of large bins having undue influence. This distance is computed after solving a linear system. A simpler method, called RootSIFT, was proposed by Arandjelovi et al. [1]

$$d_{\text{RootSIFT}}(\mathbf{s}_n(\mathbf{x}), \mathbf{s}_m(\mathbf{x})) := \sqrt{(\mathbf{s}_n(\mathbf{x}) - \bar{\mathbf{s}}_n(\mathbf{x}))^\top (\mathbf{s}_m(\mathbf{x}) - \bar{\mathbf{s}}_m(\mathbf{x}))}, \qquad (4)$$

where

$$\bar{\mathbf{s}}_n(\mathbf{x}) = \frac{1}{128} \sum_{i=1}^{128} \mathbf{s}_n(\mathbf{x})_i, \quad \bar{\mathbf{s}}_m(\mathbf{x}) = \frac{1}{128} \sum_{i=1}^{128} \mathbf{s}_m(\mathbf{x})_i, \qquad (5)$$

the values $\mathbf{s}_n(\mathbf{x})_i$ (resp. $\mathbf{s}_m(\mathbf{x})_i$) being those of the ith bin of the descriptor $\mathbf{s}_n(\mathbf{x})$ (resp. $\mathbf{s}_m(\mathbf{x})$). However, our experiments show that this last variant fails when the two SIFT descriptors to be compared represent flat unstructured regions where noise dominates. In that case the descriptors are also flat and their distance (5) is also small, while they are in fact uncorrelated. Thus we found more reliable to compare SIFT descriptors through their Normalised Cross Correlation (NCC), with the distance defined by

$$d(\mathbf{s}_n(\mathbf{x}), \mathbf{s}_m(\mathbf{x})) = \frac{1}{2}\left(1 - \frac{(\mathbf{s}_n(\mathbf{x}) - \bar{\mathbf{s}}_n(\mathbf{x}))^\top (\mathbf{s}_m(\mathbf{x}) - \bar{\mathbf{s}}_m(\mathbf{x}))}{\|\mathbf{s}_n(\mathbf{x}) - \bar{\mathbf{s}}_n(\mathbf{x})\|_2 \cdot \|\mathbf{s}_m(\mathbf{x}) - \bar{\mathbf{s}}_m(\mathbf{x})\|_2}\right), \qquad (6)$$

which takes its values in $[0, 1]$.

Once the minimal distance series (C_n) has been computed at each location, we zoom in the resulting images to retrieve the original dimensions (W, H) by bilinear interpolation (see Algorithm 1 line 11). Then, we binarize the previous series with a threshold $\alpha \in [0, 1]$ leading to the change maps

$$B_n(\mathbf{x}) = \mathbb{1}_{\{C_n > \alpha\}}(\mathbf{x}). \qquad (7)$$

3.3 Morphological Post-processing

A number of false or no detection areas may appear spatially isolated in the Boolean masks (B_n). These are generally detection errors that can be eliminated by considering a surface threshold $\theta \leq |\Omega|$ consistent with the image resolution and the minimal acceptable size of clouds. Each connected region ω of the mask B_n is removed if smaller than θ. This leads to the post-processed map defined by, for all connected region ω of B_n and all $\mathbf{x}$ in ω

$$B_n(\mathbf{x}) = \begin{cases} 1 - B_n(\mathbf{x}) & \text{if } |\omega| < \theta, \\ B_n(\mathbf{x}) & \text{otherwise.} \end{cases} \qquad (8)$$

This is done with the Boolean image partitioning approach of Tarjan and Endre [31].

3.4 Time Complexity Analysis

The Boolean maps computation has three steps. The first step is the computation of the SIFT descriptors. From the work of Vinukonda [32], the time complexity is $O(a^2 |\Omega|)$ (see [32] Table 2.1)

when the SIFT descriptors, each defined over a square tile of side a, are computed at every point of Ω. In our case we have $O(16^2|\Omega|/\kappa^2)$. The second step is the computation of the (C_n) series. For each location $\mathbf{x}$, all the pair comparisons are performed in $O(N(N+1)/2)$. The computation of the series $C_n(\mathbf{x})$ is $O(N^2)$. In total, the second step is performed in $O(|\Omega|/\kappa^2(N(N+1)/2 + N^2)) = O(|\Omega|N^2/\kappa^2)$. The third step is the Boolean masks computation. The bilinear resampling has complexity $O(|\Omega|/\kappa^2)$, the binarization $O(|\Omega|)$ and the region removal takes constant time (see [31]). Finally, the overall Boolean masks creation has complexity $O(|\Omega|N^2)$.

4 Experiments

4.1 Experimental Setup

A test set of 18 series of 15 PlanetScope images (ground sampling distance of 3 meters per pixel) of size 500×500 pixels was gathered, together with the Unusable Data Mask (UDM) files that Planet provides with every image (see Figure 1). UDM files give information on areas of unusable data within an image, such as cloudy and non-imaged areas. Cloud detection is performed by Planet on a decimated version of the image with a simple threshold based method, hence small clouds may be missed, snow is assessed as clouds, and haze and cloud shadows are missed. Note that these UDM files are not produced with the Tmask approach [21]. Hence to evaluate the results of our algorithm we prepared hand-made ground truth maps. They consist in grey level masks (B_n^{truth}) labelling opaque cloud, haze, shadow and clear ground regions. Figure 2 shows an example of such maps. One of the difficulties encountered during the ground truths creation was the classification of the haze. As haze is semi-transparent, the interpreters were instructed to mark pixels as cloudy only when the ground texture faded out. To avoid the influence of the boundary imprecision in ground truths, boundary neighborhoods were discarded (see Figure 2, right).

Figure 1: The assessment dataset contains 18 series of 15 frames of size 500×500 pixels. Eleven series show urban areas, four series show mountains, two series show agricultural fields and one serie shows a desertic region.

Figure 2: Given an image (left), a ground truth mask (center) is labelled with the following convention: opaque clouds white, cirrus light grey, shadow dark grey, clear ground black. To avoid the influence of the boundary imprecision in ground truths, one discards the boundary neighborhood (right, red contours). The assessment metric between estimated maps and ground truth is performed over the remaining regions.

4.2   Experimental Results

The full algorithm depends on four parameters, namely: the SIFT correlation threshold α, the spatial sub-sampling factor κ, the side a of the square patch used to compute a SIFT descriptor, and finally the surface filling area threshold θ. The input series (u_n) used in the following experiments is the series of the means of the three visible bands i.e. red, green and blue such as depicted in Table 1.

We set the parameters to $\alpha = 0.2$, $a = 15$, $\theta = 277$ and $\kappa = 4$. The value of θ corresponds to the ratio between a cloud minimal acceptable surface (i.e. 2500 m^2) and the image resolution (i.e. 9 m^2 per pixel). We apply the algorithm to the 18 series of our PlanetScope dataset.

Qualitative analysis.   We present some results in Figures 3, 4 and 5. Figure 3 shows the complementarity of the texture and contrast features to estimate at best the Boolean maps. Figure 4 illustrates the effectiveness of the method in different environments. Figure 5 shows several limitations of the texture matching approach when the ground varies at each date or when there is little texture.

The method does not necessarily detect shadowed regions. Such regions are actually visible, but their low contrast and brightness may lead to misinterpretations of the time series. Thus, temporary shadows due to clouds should be detected if possible. However, being robust to contrast changes, SIFT doesn't allow to reliably discriminate shadows. It actually detects well the transitions between shadowed and lit areas because of their apparent structure change, caused by the shadow boundaries.

Another visibility hazard is caused by cirrus and other atmospheric attenuation factors such as haze, smog, fumes, aerosols, etc. As we said, this attenuation of contrast does not prevent positive SIFT matching.

Comparative analysis.   Figure 6 shows some examples of cloud detections from PlanetScope UDM masks and our method. As the PlanetScope method relies mostly on the high luminance of clouds, it greatly underestimates dark clouds. This fact is confirmed by the confusion matrices of Table 2 which show the classification rates for the PlanetScope UDM masks and ours. Our method performs 28% (resp. 4%) better than PlanetScope UDM masks regarding the clouds (resp. the visibility) classification. Notice that, because our algorithm can detect shadow boundaries, it could label as occluded some regions that are visible in the sense of the PlanetScope UDM masks. That's why we don't take into account the annotated shadow regions of the ground truth for the computation of

the confusion matrices.

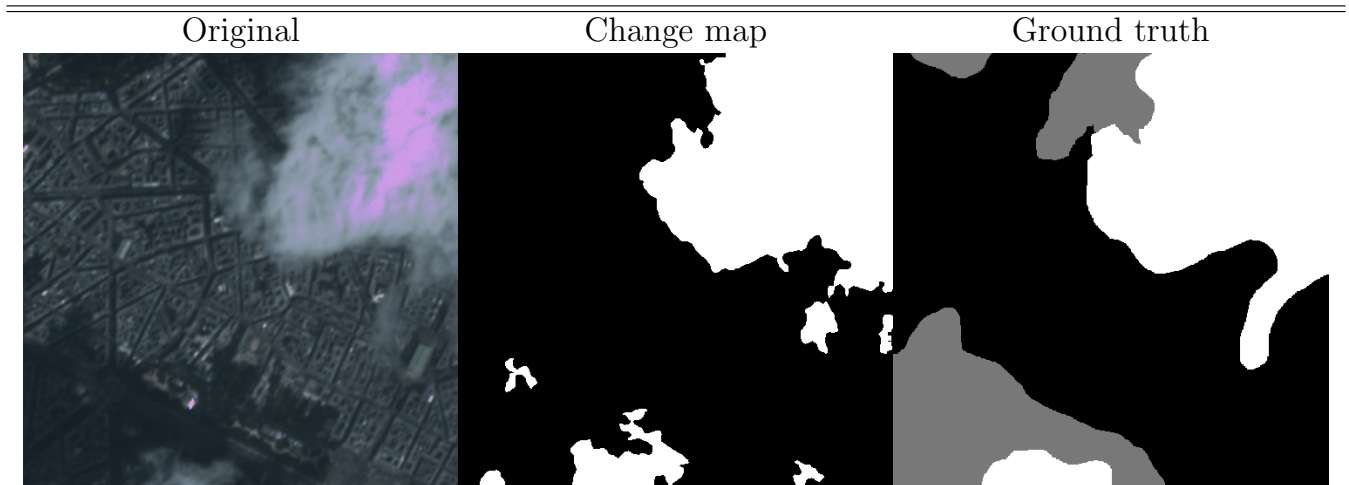| Original	Change map	Ground truth

Figure 3: Example of visibility map obtained with the method: the SIFT feature is robust to contrast changes and therefore recognizes the ground even in shadow. In some locations the boundaries are more accurate than those of the ground truth.

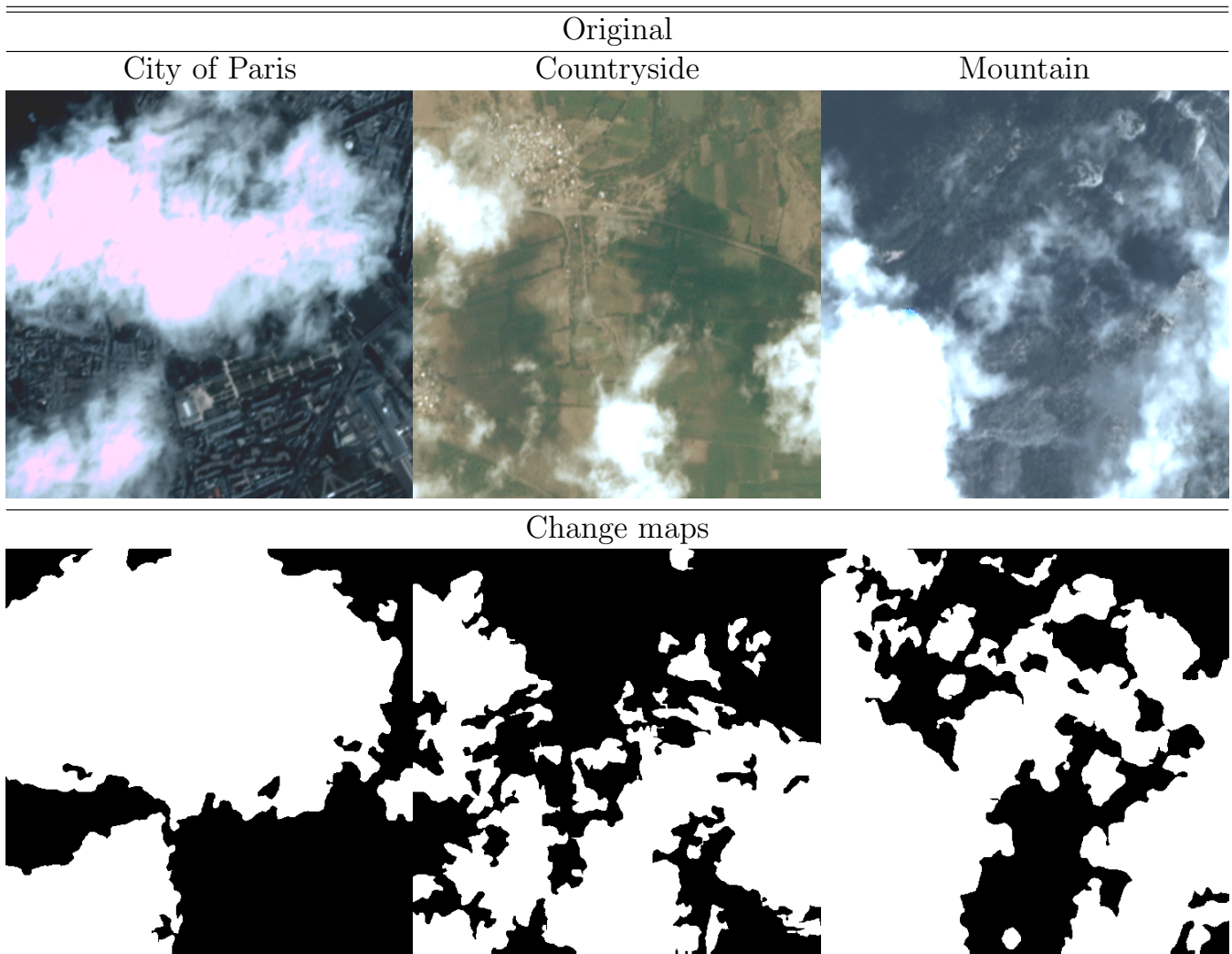|	Original	
City of Paris	Countryside	Mountain

Change maps

Figure 4: These results show the performance of the texture comparison approach for different kinds of scenes. Clouds boundaries are better estimated when the texture of the ground is stable over time.

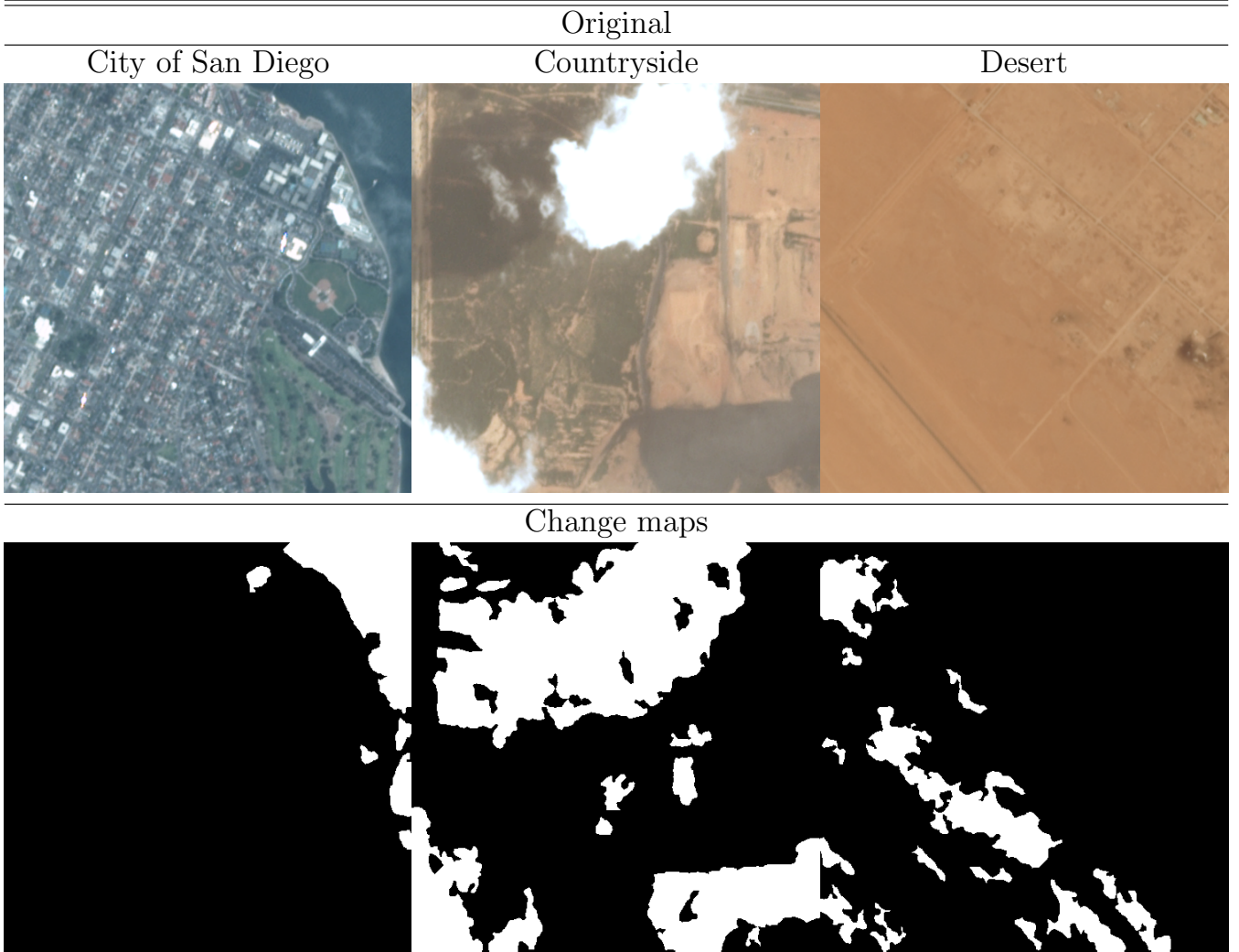| Original		
City of San Diego	Countryside	Desert

Change maps

Figure 5: The above results show some limitations of the texture comparison approach. On the left column, the map contains very few false detections except in the constantly changing sea area. However, this problem can be mitigated if the image has a green visible band ($\simeq 550$ nm) together with a near infra-red band ($\simeq 800$ nm): the NDWI (Normalized Difference Water Index) [19] can be computed to detect water bodies. On the middle column, the variations of the agricultural fields boundaries hinder texture matching: some false detections appear at the image center. On the right column, the non-textured ground is dominated by noise. Hence the normalized cross correlation (Equation (6)) of the SIFT descriptor fails and leads to false alarms.

		Planet UDM		Our method	
		Visible	Occluded	Visible	Occluded
Ground Truth	Visible	93.68	06.32	**97.78**	02.22
	Occluded	38.76	61.24	10.64	**89.36**

Table 2: Normalized confusion matrices computed from the 18 PlanetScope ground truth series compared to the cloud series extracted from the UDM masks by Planet (left matrix), and our visibility masks series (B_n) (right matrix). Our method outperforms Planet UDM masks for the cloud (occluded) category by about 28%.
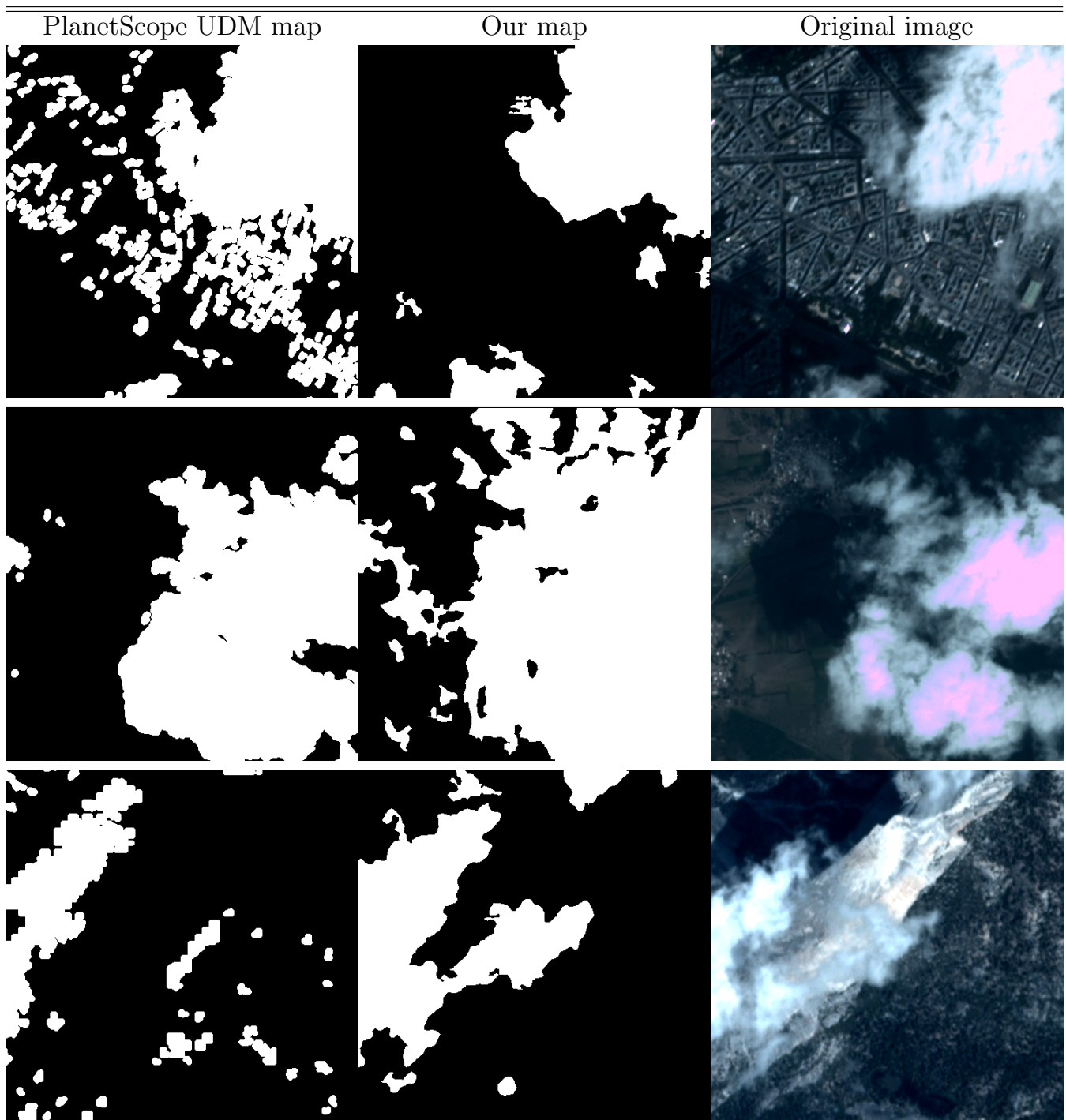
Figure 6: Examples of cloud maps estimated by the PlanetScope algorithm and our method. Results show that the luminance based method of Planet is very sensitive to the intensity gaps. Whereas it may over detect buildings which are very clear (top frame), it may under estimate clouds whose intensity is close to that of the ground (bottom frame). It precisely delimits the clear region of the cloud if the ground is quite dark (middle frame, bottom) but fails to detect clouds embedded in shadow regions (middle frame, top).

5   Conclusion

In this paper we have presented an algorithm to automatically compute a time series of visibility change masks, given an input times series of registered satellite images. Our method is based on uncorrelated texture detection thanks to SIFT descriptors. In addition, hand-made ground truth maps were created for 18 PlanetScope time series and are made available. Experiments have been carried out to demonstrate the effectiveness of the proposed algorithm: our method outperforms the PlanetScope UDM masks by 28% for the cloud classification.

Image Credits

All images in this manuscript were produced by the authors (license CC-BY-SA).

References

[1] R. Arandjelovic and A. Zisserman, *Three things everyone should know to improve object retrieval*, in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), CVPR '12, Washington, DC, USA, 2012, IEEE Computer Society, pp. 2911–2918. https://doi.org/10.1109/CVPR.2012.6248018.

[2] M. Aschenbeck, *Advanced cloud detection using machine learning and optimization techniques*, 2018.

[3] G. Chandran and C. Jojy, *A survey of cloud detection techniques for satellite images*, International Research Journal of Engineering and Technology (IRJET), 2 (2015), pp. 2485–2490. https://www.irjet.net/archives/V2/i9/IRJET-V2I9305.pdf.

[4] L. Gómez-Chova, J. Amorós-López, G. Mateo-García, J. Muñoz Marí, and G. Camps-Valls, *Cloud masking and removal in remote sensing image time series*, Journal of Applied Remote Sensing, 11 (2017), pp. 11 – 11 – 15. http://dx.doi.org/10.1117/1.JRS.11.015005.

[5] N. R. Goodwin, L. J. Collett, R. J. Denham, N. Flood, and D. Tindall, *Cloud and cloud shadow screening across Queensland, Australia: An automated method for Landsat TM/ETM + time series*, Remote Sensing of Environment, 134 (2013), pp. 50 – 65. http://dx.doi.org/10.1016/j.rse.2013.02.019.

[6] W. Guan, P. Khadka, and J. Gerar, *Cloud detection on remote sensing imagery*, 2017.

[7] O. Hagolle, G. Dedieu, B. Mougenot, V. Debaecker, B. Duchemin, and A. Meygret, *Correction of aerosol effects on multi-temporal images acquired with constant viewing angles: Application to Formosat-2 images*, Remote Sensing of Environment, 112 (2008), pp. 1689–1701. http://dx.doi.org/10.1016/j.rse.2007.08.016.

[8] O. Hagolle, M. Huc, D. Pascual, and G. Dedieu, *A Multi-Temporal and Multi-Spectral Method to Estimate Aerosol Optical Thickness over Land, for the Atmospheric Correction of FormoSat-2, LandSat, VENS and Sentinel-2 Images*, Remote Sensing, 7 (2015), pp. 2668–2691. http://dx.doi.org/10.3390/rs70302668.

[9]   O. Hagolle, M. Huc, D. Villa Pascual, and G. Dedieu, *A multi-temporal method for cloud detection, applied to FORMOSAT-2, VENuS, LANDSAT and SENTINEL-2 images*, Remote Sensing of Environment, 114 (2010), pp. 1747–1755. http://dx.doi.org/10.1016/j.rse.2010.03.002.

[10] O. Hagolle, S. Sylvander, M. Huc, M. Claverie, D. Clesse, C. Dechoz, V. Lonjou, and V. Poulain, *SPOT-4 (Take 5): Simulation of Sentinel-2 Time Series on 45 Large Sites*, Remote Sensing, 7 (2015), pp. 12242–12264. http://dx.doi.org/10.3390/rs70912242.

[11] A. Hollstein, K. Segl, L. Guanter, M. Brell, and M. Enesco, *Ready-to-use methods for the detection of clouds, cirrus, snow, shadow, water and clear sky pixels in sentinel-2 msi images*, Remote Sensing, 8 (2016). http://dx.doi.org/10.3390/rs8080666.

[12] R. R. Irish, *Landsat 7 automatic cloud cover assessment*, 2000. http://dx.doi.org/10.1117/12.410358.

[13] R. R. Irish, J. L. Barker, S. N. Goward, and T. Arvidson, *Characterization of the Landsat-7 ETM+ Automated Cloud-Cover Assessment (ACCA) Algorithm*, Photogrammetric Engineering & Remote Sensing, 72 (2006), pp. 1179–1188. http://dx.doi.org/10.14358/PERS.72.10.1179.

[14] G. Jedlovec and S. Haines, *Spatial and temporal varying thresholds for cloud detection in satellite imagery*, in Proceedings of IEEE International Geoscience and Remote Sensing Symposium, July 2007, pp. 3329–3332. http://dx.doi.org/10.1109/IGARSS.2007.4423557.

[15] M. D. King, S. Platnick, W. P. Menzel, S. A. Ackerman, and P. A. Hubanks, *Spatial and Temporal Distribution of Clouds Observed by MODIS Onboard the Terra and Aqua Satellites*, IEEE Transactions on Geoscience and Remote Sensing, 51 (2013), pp. 3826–3852. http://dx.doi.org/10.1109/TGRS.2012.2227333.

[16] K. S. Kuo, R. M. Welch, and S. K. Sengupta, *Structural and Textural Characteristics of Cirrus Clouds Observed Using High Spatial Resolution LANDSAT Imagery*, Journal of Applied Meteorology, 27 (1988), pp. 1242–1260. http://dx.doi.org/10.1175/1520-0450(1988)027<1242:SATCOC>2.0.CO;2.

[17] C. Latry, C. Panem, and P. Dejean, *Cloud detection with SVM technique*, in Proceedings of IEEE International Geoscience and Remote Sensing Symposium, July 2007, pp. 448–451. http://dx.doi.org/10.1109/IGARSS.2007.4422827.

[18] D. G. Lowe, *Distinctive image features from scale-invariant keypoints*, International Journal of Computer Vision, 60 (2004), pp. 91–110. http://dx.doi.org/10.1023/B:VISI.0000029664.99615.94.

[19] S. K. McFeeters, *The use of the Normalized Difference Water Index (NDWI) in the delineation of open water features*, International Journal of Remote Sensing, 17 (1996), pp. 1425–1432. http://dx.doi.org/10.1080/01431169608948714.

[20] U. Mueller-Wilm, *Sen2cor configuration and user manual*, Tech. Report 1, European Space Agency, 2016. S2-PDGS-MPC-L2A-SUM-V2.3.

[21] M.U. Mller, G. Ghergu, V. Scipione, J. Morrow, G. Lohss, and N. Flood, *Planet-tmask (version 1.0)*, 2018. https://github.com/planetlabs/planet-tmask.

[22] S. Özkan, M. Efendioglu, and C. Demirpolat, *Cloud detection from RGB color remote sensing images with deep pyramid networks*, in Proceedings of IEEE International Geoscience and Remote Sensing Symposium (IGARSS), 2018, pp. 6939–6942. http://dx.doi.org/10.1109/IGARSS.2018.8519570.

[23] O. Pele and M. Werman, *A linear time histogram metric for improved SIFT matching*, in Proceedings of the 10th European Conference on Computer Vision: Part III, ECCV '08, Berlin, Heidelberg, 2008, Springer-Verlag, pp. 495–508. http://dx.doi.org/10.1007/978-3-540-88690-7_37.

[24] ——, *The quadratic-chi histogram distance family*, in Proceedings of the 11th European Conference on Computer Vision: Part II, ECCV'10, Berlin, Heidelberg, 2010, Springer-Verlag, pp. 749–762. http://dl.acm.org/citation.cfm?id=1888028.1888085.

[25] M. Rais, *Fast and accurate image registration. Applications to on-board satellite imaging*, PhD Thesis, Université Paris-Saclay, Dec. 2016. https://tel.archives-ouvertes.fr/tel-01485321.

[26] R. Rossi, R. Basili, F. Del Frate, M. Luciani, and F. Mesiano, *Techniques based on Support Vector Machines for cloud detection on QuickBird satellite imagery*, in Proceedings of IEEE International Geoscience and Remote Sensing Symposium, July 2011, pp. 515–518. http://dx.doi.org/10.1109/IGARSS.2011.6049178.

[27] Y. Rubner, C. Tomasi, and L. J. Guibas, *The Earth Mover's Distance as a Metric for Image Retrieval*, International Journal of Computer Vision, 40 (2000), pp. 99–121. http://dx.doi.org/10.1023/A:1026543900054.

[28] P. L. Scaramuzza, M. A. Bouchard, and J. L. Dwyer, *Development of the Landsat Data Continuity Mission Cloud-Cover Assessment Algorithms*, IEEE Transactions on Geoscience and Remote Sensing, 50 (2012), pp. 1140–1154. http://dx.doi.org/10.1109/TGRS.2011.2164087.

[29] P. Singh and N. Komodakis, *Cloud-Gan: Cloud Removal for Sentinel-2 Imagery Using a Cyclic Consistent Generative Adversarial Networks*, in Proceedings of IEEE International Geoscience and Remote Sensing Symposium (IGARSS), 2018, pp. 1772–1775. http://dx.doi.org/10.1109/IGARSS.2018.8519033.

[30] A. Taravat, S. Proud, S. Peronaci, F. Del Frate, and N. Oppelt, *Multilayer Perceptron Neural Networks Model for Meteosat Second Generation SEVIRI Daytime Cloud Masking*, Remote Sensing, 7 (2015), pp. 1529–1539. http://dx.doi.org/10.3390/rs70201529.

[31] R. E. Tarjan, *Efficiency of a good but not linear set union algorithm*, Journal of ACM, 22 (1975), pp. 215–225. http://dx.doi.org/10.1145/321879.321884.

[32] P. Vinukonda, *A study of the scale-invariant feature transform on a parallel pipeline*, master's thesis, Louisiana State University and Agricultural and Mechanical College, 2011.

[33] G. Vivone, P. Addesso, R. Conte, M. Longo, and R. Restaino, *A class of cloud detection algorithms based on a MAP-MRF approach in space and time*, IEEE Transactions on Geoscience and Remote Sensing, 52 (2014), pp. 5100–5115. http://dx.doi.org/10.1109/TGRS.2013.2286834.

[34] R. M. Welch, S. K. Sengupta, and D. W. Chen, *Cloud field classification based upon high spatial resolution textural features: 1. gray level co-occurrence matrix approach*, Journal of Geophysical Research: Atmospheres, 93 (1988), pp. 12663–12681. http://dx.doi.org/10.1029/JD093iD10p12663.

[35] B. A. Wielicki and R. N. Green, *Cloud Identification for ERBE Radiative Flux Retrieval*, Journal of Applied Meteorology, 28 (1989), pp. 1133–1146. http://dx.doi.org/10.1175/1520-0450(1989)028<1133:CIFERF>2.0.CO;2.

[36] W. D. Zhang, M. X. He, and M. W. Mak, *Cloud detection using probabilistic neural networks*, in IEEE International Geoscience and Remote Sensing Symposium, vol. 5, 2001, pp. 2373–2375 vol.5. http://dx.doi.org/10.1109/IGARSS.2001.978006.

[37] Z. Zhu and C. E. Woodcock, *Automated cloud, cloud shadow, and snow detection in multitemporal Landsat data: An algorithm designed specifically for monitoring land cover change*, Remote Sensing of Environment, 152 (2014), pp. 217 – 234. http://dx.doi.org/10.1016/j.rse.2014.06.012.