



Published in Image Processing On Line on 2020-11-21.  
Submitted on 2019-07-10, accepted on 2020-06-03.  
ISSN 2105-1232 © 2020 IPOL & the authors CC-BY-NC-SA  
This article is available online with supplementary materials,  
software, datasets and online demo at  
<https://doi.org/10.5201/ipol.2020.271>

# Cloud Detection by Luminance and Inter-band Parallax Analysis for Pushbroom Satellite Imagers

Tristan Dagobert<sup>1</sup>, Rafael Grompone von Gioi<sup>1</sup>, Carlo de Franchis<sup>1,2</sup>,  
Jean-Michel Morel<sup>1</sup>, Charles Hessel<sup>1,2</sup>

<sup>1</sup> Centre Borelli, ENS Paris-Saclay, France ([tristan.dagobert@cmla.ens-cachan.fr](mailto:tristan.dagobert@cmla.ens-cachan.fr), {[grompone](mailto:grompone@ens-paris-saclay.fr),  
[jean-michel.morel](mailto:jean-michel.morel@ens-paris-saclay.fr),[carlo.de-franchis](mailto:carlo.de-franchis@ens-paris-saclay.fr),[charles.hessel](mailto:charles.hessel@ens-paris-saclay.fr)}@ens-paris-saclay.fr)

<sup>2</sup> Kayrros, France

*Communicated by* Charles Hessel

*Demo edited by* Tristan Dagobert

## Abstract

This paper proposes a cloud detection algorithm for Earth observation images obtained by pushbroom satellite imagers. The pushbroom technology induces an inter-band acquisition delay leading to a parallax effect for the clouds. We propose a method exploiting this characteristic thanks to the analysis of the inter-band disparity. Several other features discriminating clouds are also defined and all are merged to build a robust *a contrario* statistical decision. Experiments applied on scenes acquired by various pushbroom satellites such as Sentinel-2, RapidEye and WorldView-2 show the effectiveness of the proposed method. In particular, we demonstrate a balanced accuracy rate close to 98% for cloud and non cloud classification for Sentinel-2 images.

**Keywords:** clouds; pushbroom; optical flow; hypothesis testing; Sentinel-2; RapidEye; WorldView-2

## Source Code

The source code of a C implementation of the proposed algorithm is provided on the [web page<sup>1</sup>](#) of the article.

## Supplementary Material

The 9789 annotated Sentinel-2 366×366 tiles extracted from the Hollstein et al. dataset [9] and used for the experiments are provided as supplementary material.

<sup>1</sup><https://doi.org/10.5201/ipol.2020.271>

# 1 Introduction

Unless observed for meteorological purposes, clouds are an overwhelming nuisance for optical satellite imagery. Not only they hide the ground, but their detection is also a major concern to avoid detection and interpretation errors in automatic image analysis. Being so numerous and large, satellite images nevertheless require automatic analysis. Hence it is very important to detect automatically and accurately all clouds in any image.

Furthermore, a number of companies in the Earth observation industry are launching constellations of dozens to hundreds of satellites to get a short revisit time over any region. Most of these satellites only acquire a few spectral bands in the visible light range of the electromagnetic spectrum. Thus, many lack specifically designed spectral bands for cloud detection (such as, for example, the cirrus band B10 of Sentinel-2).

In this paper, our intention is to demonstrate that this limitation can be compensated, for cloud detection, in the particular case where images are acquired with a pushbroom satellite such as Sentinel-2, Landsat-8, Pléiades or RapidEye. The pushbroom imaging mode acquires spectral bands one after the other by a scanning process. This procedure causes an acquisition delay during which the satellite moves from a position  $S_0$  to  $S_1$  and induces a parallax effect (see Figure 1).

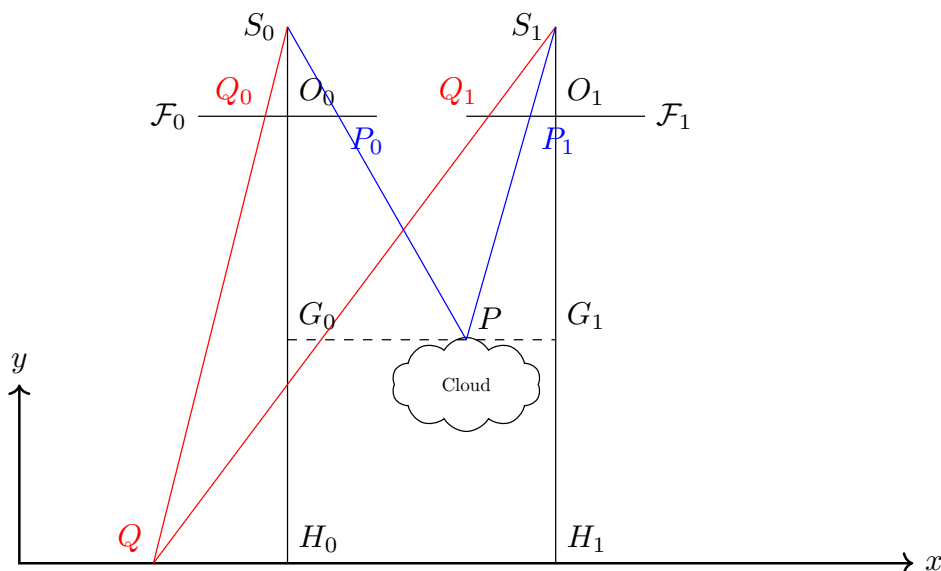


Figure 1: Principle of the parallax effect in the ideal case where the satellite has a translation motion and the ground has no relief. Let  $\mathcal{F}_0$  and  $\mathcal{F}_1$  be the focal planes of the satellite at positions  $S_0$  and  $S_1$  respectively at times  $t_0$  and  $t_1$ . Points  $O_0$  and  $O_1$  are their respective centers. Points  $H_0$  and  $H_1$  are the projection of  $O_0$  and  $O_1$  on the ground. For each point  $Q$  located on the ground (*i.e.* on the line  $(H_0H_1)$ ), its projections over the focal planes are respectively  $Q_0$  and  $Q_1$ . The observable disparity between  $Q_0$  and  $Q_1$  is  $\overrightarrow{O_0Q_0} - \overrightarrow{O_1Q_1} = \alpha \overrightarrow{H_0H_1}$ . The observable disparity for a point  $P$  located on a cloud is  $\overrightarrow{O_0P_0} - \overrightarrow{O_1P_1} = \beta \overrightarrow{H_0H_1}$ . After the registration process (corresponding, in the case of Sentinel-2 to the Level-1A processing [5]), operated as a global translation of the image  $u_1$  to  $u_0$ , the new observable disparity on the ground is zero while there is a nonzero disparity  $(\beta - \alpha) \overrightarrow{H_0H_1}$  for pixels located on clouds.

In this study, we present an algorithm based on the exploitation of the inter-band parallax information to detect clouds. As this method is intended to be general and common to the aforementioned satellites, we restrict our input data to spectral bands in the visible range of the spectrum. We evaluate qualitatively and quantitatively this method with the Hollstein et al. [9] ground truth Sentinel-2 dataset. In addition, we show some results on RapidEye, SkySat and WorldView-2 images.

This article is organized as follows: Section 2 describes related works. Section 3 describes the proposed algorithm. Section 4 deals with the setup of the experimental protocol and compares the

efficiency of the different strategies applied to Sentinel-2 data. Finally, we present the conclusions in Section 5.

## 2 Related Work

The idea to exploit the parallax information using a stereoscopic approach has been developed since the middle of the 90's. Given a series of satellite images in quick succession, a parallax effect is generally perceptible, caused by the clouds relative altitude with respect to the ground. In addition, knowing the geometrical properties and positions of the satellite, it is possible to estimate the clouds top height with stereo-matching algorithms. Shin and Pollard [16] proposed to detect clouds over seas from data acquired by an Along-Track Scanning Radiometer (ATSR). Manizade et al. [11] applied discrete correlation to binary series obtained by the slicing of the 8-12  $\mu\text{m}$  infrared band into 18 temperature ranges. This permitted to refine the cloud altitude estimation up to  $\pm 390$  m. Working with the Multi-angle Imaging SpectroRadiometer (MISR) of the Terra satellite, a radiometer composed of nine pushbroom cameras with a time delay between adjacent cameras of about 45 seconds, Mulleret al. [13] studied a set of fast algorithms to obtain cloud-top height and motion. They used a combination of area-based and feature-based stereo-matchers with only pixel-level acuity.

Wu et al. [21] proposed to merge stereo-matching results with external information coming from a digital elevation model acquired by the Shuttle Radar Topography Mission (SRTM). As shown by the authors, the wind speed cannot be neglected when working with high resolution images: the diffusion of the cloud layers during the acquisition can lead to bad estimations of the clouds altitudes. A significant enhancement has been presented by Panem et al. [14], using SPOT5 images which remedies this temporal sampling problem. The parallax of images is obtained by dense matching after aligning panchromatic and multispectral images with SRTM. As parallax changes are significant at the edges of cloudy areas, these areas can be automatically extracted. This means that given multi-spectral data recorded by a satellite at a single time  $t$ , it is possible to estimate the clouds apparent motion. The authors notice however, that given the short baseline between panchromatic and multispectral images, this method can extract only clouds higher than 600 m. From Research Scanning Polarimeter (RSP) data, Sinclair et al. [17] determine the clouds top height thanks to a set of selected consecutive nadir reflectances and the cross correlations between this set and the collocated sets at other viewing angles. Under the assumption that cloud reflectances are isotropic, local peaks in the obtained correlation profile indicate cloud layers.

Frantzet al. [4] proposed to separate clouds and land surfaces thanks to the spectrally correlated NIR bands 7, 8 and 8A of Sentinel-2 which are affected by a small view angle parallax. Defining first the two ratios  $R_{8A,8} = B_8/B_{8A}$  and  $R_{8A,7} = B_7/B_{8A}$  they convolve them with a variance filter to obtain the texture indices  $V_{8A,8}$  and  $V_{8A,7}$  respectively. Finally they define an index called Cloud Displacement Index (CDI) by  $\frac{V_{8A,7} - V_{8A,8}}{V_{8A,7} + V_{8A,8}}$ . CDI values lower than  $-0.5$  are considered to be clouds. The authors don't explain why they didn't involve the NIR bands 5 and 6 as well, despite the fact that these bands have the same resolution as the aforementioned, are spectrally close, and also induce pairwise parallaxes. Bands 8 and 8A seem to be used together because band 8A ( $864 \pm 10$  nm) is a spectral subset of band 8 ( $832 \pm 53$  nm); their extreme similarity makes equalization unnecessary. Some comparisons with our method are shown in Section 4.

### 3 Algorithm Overall Description

The algorithm input is an RGB image, seen as a triplet  $\mathbf{u} = (u_R, u_G, u_B)$  of gray level images defined on the same discrete domain  $\Omega = \llbracket 1, W \rrbracket \times \llbracket 1, H \rrbracket$  where  $W$  (resp.  $H$ ) is the image width (resp. height) and  $\llbracket a, b \rrbracket$  a discrete interval from integer  $a$  to integer  $b$ . The pixels positions are denoted by  $\mathbf{x} = (x, y)^\top$ . The algorithm output is a Boolean map  $\tilde{B}$  such that

$$\tilde{B}(\mathbf{x}) = \begin{cases} 1 & \text{if the ground is hidden by cloud,} \\ 0 & \text{if the ground is clear.} \end{cases} \quad (1)$$

Our algorithm is mainly based on the analysis of the parallax effects observed between the three color bands  $u_R$ ,  $u_G$  and  $u_B$ . Given a point of the scene that is observed at position  $\mathbf{x}_R$  in  $u_R$ , we study its respective positions  $\mathbf{x}_G$  and  $\mathbf{x}_B$  in  $u_G$  and  $u_B$ . From a theoretical point of view, denoting  $D_{RG}$  (resp.  $D_{GB}$ ,  $D_{RB}$ ) the displacement map defined on the continuous domain  $\Omega' = [1, W] \times [1, H]$  between  $u_R$  and  $u_G$  (resp.  $u_G$  and  $u_B$ ,  $u_R$  and  $u_B$ ), we should have for all  $\mathbf{x}_R \in \Omega'$  the relation

$$D_{RB}(\mathbf{x}_R) = D_{RG}(\mathbf{x}_R) + D_{GB}(\mathbf{x}_G). \quad (2)$$

Based on the three vectors of relation (2), we define three complementary and non redundant features, namely the angular error, the composition error ratio and the modulus, the meaning of which is clarified in Section 3.2. In order to leverage all the information available in a single image, we do not limit ourselves to the optical flow analysis. We also consider two features computed from the luminance and chrominance properties of the clouds.

Based on the empirical probability distribution function of each feature, we define an *a contrario* multi-feature hypothesis testing [3]. This leads to combine the five features into a single detector. For that we propose to sum a normalized measure of each feature. As a result, we obtain a probability threshold yielding a Boolean map. The application of an optional morphological post-processing to remove small artefacts gives the final Boolean map  $\tilde{B}$ . The algorithm is described in Algorithm 1. Its parameters are summarized in Table 1.

#### 3.1 Disparity Map Computation

Disparity map computation in stereovision is a very active field of research, so methods are numerous. In the present case, a subpixel precision is needed in order to detect small parallax effects. To this end we use an optical flow method.

**Histogram equalization** Optical flow algorithms assume that the aspect of observed objects remains nearly constant in consecutive frames  $u$  and  $v$ . Yet in satellite imaging the observable textures and luminances in the green and blue bands are generally quite close, but the red band produces large differences hence this hypothesis does not stand.

To stabilize locally the dynamics, several approaches exist: midway equalization [7] where each image is equalized by the midway histogram of both (in the sense of a transport equation) or the local approach by Sabater et al. [15] consisting in transferring the local mean and variance of  $v$  to  $u$ . Both methods have a tendency to transfer the texture of  $v$  to  $u$ , thus leading to the creation of artificial patterns. We therefore decided to apply the Zabih and Woodfill [22] rank transform (RT), which encodes for each pixel the rank of its grey value in its neighborhood  $\mathcal{V}_{RT}(\mathbf{x})$ :

$$RT(\mathbf{x}) = \#\{\mathbf{y} \in \mathcal{V}_{RT}(\mathbf{x}) \mid u(\mathbf{y}) < u(\mathbf{x})\}. \quad (3)$$

By construction, this method decreases dramatically the image dynamic range, which is reduced to the size of the transformed neighborhood. With this method the optical flow estimation becomes

---

**Algorithm 1:** COMPUTECLOUDMAP[ $\mathbf{u}, (f_c)_{c \in \{\lambda, \kappa, \xi, \rho, \phi\}}, \epsilon, s, a$ ] — Computation of the cloud map.

---

**Input:**  $\mathbf{u}$  the visible band triplet,  $(f_c)_{c \in \{\lambda, \kappa, \xi, \rho, \phi\}}$  the features probability laws,  $\epsilon$  the NFA threshold,  $s$  the post-processing threshold,  $a$  the step of the square 9-neighborhood.

**Output:**  $\tilde{B}$  the Boolean map.

```

1 // EQUALIZATION OF THE CHANNELS
2 for  $b \in \{R, G, B\}$  do
3    $\tilde{u}_b \leftarrow$  compute the rank transform (3) [ $u_b, \mathcal{V}_{RT} = 35 \times 35$ ]
4 // COMPUTATION OF THE DISPARITY MAPS
5  $\mathbf{d}_{RG} \leftarrow$  apply the optical flow (4) [ $u_R, u_G, \alpha = 20, \gamma = 1$ ]
6  $\mathbf{d}_{GB} \leftarrow$  apply the optical flow (4) [ $u_G, u_B, \alpha = 20, \gamma = 1$ ]
7  $\mathbf{d}_{RB} \leftarrow$  apply the optical flow (4) [ $u_R, u_B, \alpha = 20, \gamma = 1$ ]
8  $\mathbf{d}'_{GB} \leftarrow$  apply the optical flow (4) [ $u_G, u_B, \alpha = 200, \gamma = 1$ ]
9  $\mathbf{d}_{\tilde{G}B} \leftarrow$  interpolate by bilinear filter [ $\mathbf{d}_{RG}, \mathbf{d}_{GB}, \mathbf{d}_{RB}$ ]
10 // COMPUTATION OF THE FEATURE MAPS
11 for  $\mathbf{x} \in \Omega$  do
12    $x_{\mathbf{x}, \phi} \leftarrow$  compute feature  $\phi$  according to (5) [ $\mathbf{d}_{RG}, \mathbf{d}_{RB}, \mathbf{d}_{\tilde{G}B}$ ]
13    $x_{\mathbf{x}, \xi} \leftarrow$  compute feature  $\xi$  according to (6) [ $\mathbf{d}_{RG}, \mathbf{d}_{RB}, \mathbf{d}_{\tilde{G}B}$ ]
14    $x_{\mathbf{x}, \rho} \leftarrow$  compute feature  $\rho$  according to (7) [ $\mathbf{d}'_{GB}$ ]
15    $x_{\mathbf{x}, \lambda} \leftarrow$  compute feature  $\lambda$  according to (8) [ $\mathbf{u}$ ]
16    $x_{\mathbf{x}, \kappa} \leftarrow$  compute feature  $\kappa$  according to (9) [ $\mathbf{u}$ ]
17 // UNIFORMIZATION
18 for  $c \in \{\lambda, \kappa, \xi, \rho, \phi\}$  do
19   for  $\mathbf{x} \in \Omega$  do
20      $w_{\mathbf{x}, c} \leftarrow F_c(x_{\mathbf{x}, c})$ 
21 // COMBINATION OF THE FEATURES, THRESHOLDING OF THE NFA MAP
22 for  $\mathbf{x} \in \Omega$  do
23    $y_{\mathbf{x}} \leftarrow \sum_{\mathbf{y} \in \mathcal{V}(\mathbf{x})} \sum_c w_{\mathbf{y}, c}$  according to (20) [ $a$ ]
24    $B(\mathbf{x}) \leftarrow \text{NFA}(y_{\mathbf{x}}) < \epsilon$  according to (13) and (15) [ $9 \times \#c$ ]
25 // MORPHOLOGICAL FILTERING
26  $\tilde{B} \leftarrow$  remove small regions by Equation (23) [ $B, s$ ]
27 return  $\tilde{B}$ 

```

---

coarse which is *a priori* counterproductive. Yet it is not important that the ground be correctly matched provided that clouds are; this difference will actually be exploited in Section 3.2.

Experimentally (see Figure 2), we observe that, as clouds have similar reflectances in the red, green and blue bands (see Kokhanovsky [10]) their rank transform is very stable while the ground, which often induces contrast inversions, is chaotic.

**Optical flow estimation** Numerous methods are available to compute disparity maps. We apply the optical flow method of Monzónet al. [12]. It is preferred here because it has an automatic mechanism to weight the flow regularization term and preserve discontinuities. As shown in [2] its results are accurate. The functional to minimize is defined by

Table 1: Table summarizing the parameters of the proposed method. Among these, only the last three can be changed to refine the cloud maps. Indeed, the parameters  $\alpha$  and  $\gamma$  play an intermediate role and are fixed once for all and the function values ( $\phi, \xi, \rho, \lambda$  and  $\kappa$ ) have been computed empirically.

$\alpha$	The regularization term on the flow in the optical flow estimation
$\gamma$	The regularization term on the image in the optical flow estimation
$\phi$	The angular error function
$\xi$	The composition error ratio function
$\rho$	The modulus function
$\lambda$	The luminance function
$\kappa$	The color saturation function
$a$	The 9-neighborhood step of the pixel $\mathbf{x}$
$\epsilon$	The NFA threshold
$s$	The surface threshold of the morphological post-processing

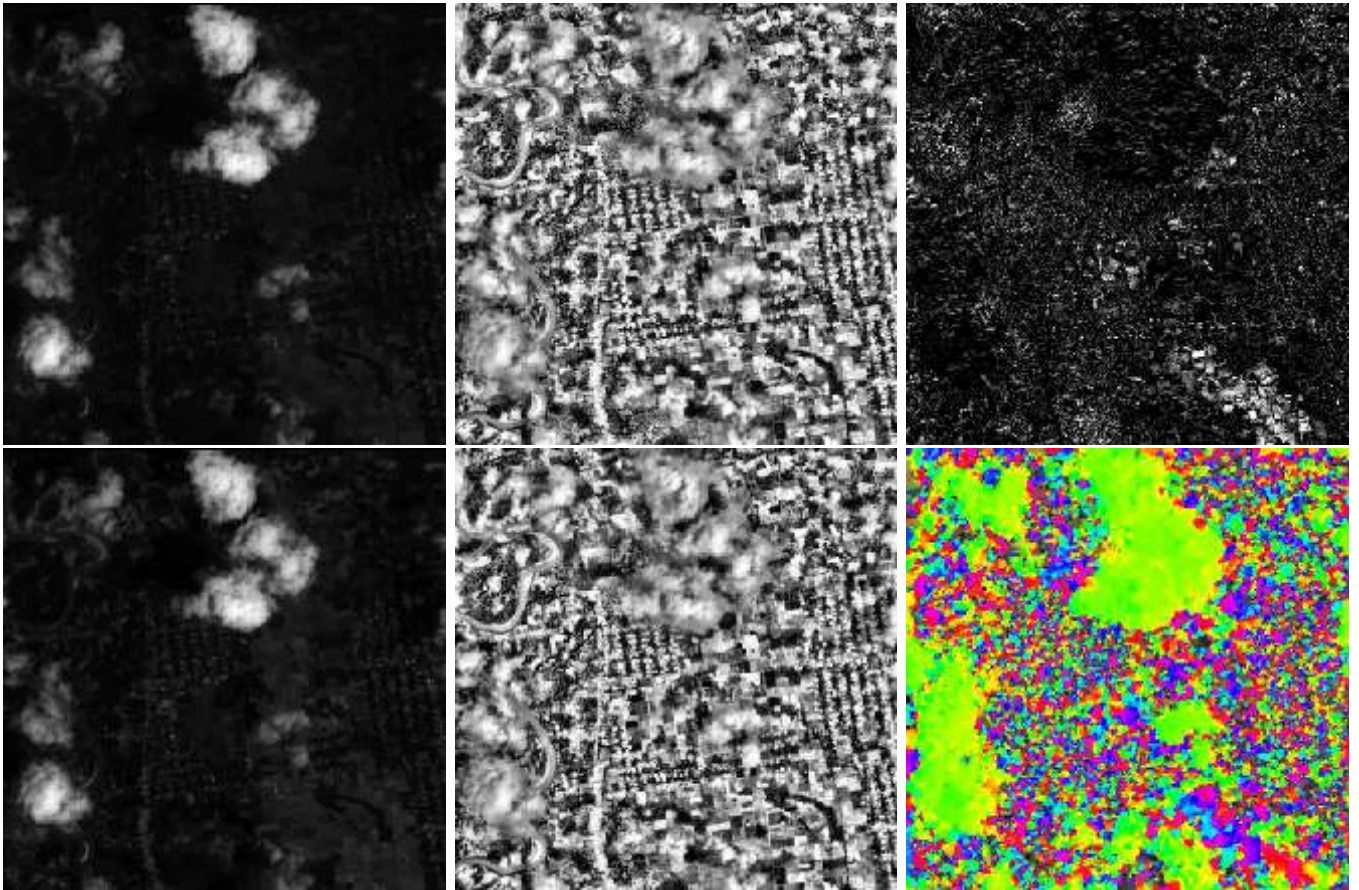


Figure 2: From top to bottom and left to right: the blue and red bands, their rank transforms with  $\mathcal{V}_{RT}$  a  $35 \times 35$  pixels square patch, the absolute difference (low values are in black) between the rank transforms, the disparity map computed from them. The rank transform exacerbates the differences of textures on the ground due to reflectance variations. This implies that the optical flow algorithm produces a heterogeneous vector field on cloud free regions. The flow is much more regular in cloudy regions.

$$\begin{aligned}
 E(\mathbf{d}) = & \int_{\Omega} \Psi \left( (v(\mathbf{x} + \mathbf{d}(\mathbf{x})) - u(\mathbf{x}))^2 \right) d\mathbf{x} + \gamma \int_{\Omega} \Psi \left( |\nabla v(\mathbf{x} + \mathbf{d}(\mathbf{x})) - \nabla u(\mathbf{x})|^2 \right) d\mathbf{x} \\
 & + \alpha \int_{\Omega} \Psi \left( \Phi(\nabla u(\mathbf{x})) (|\nabla d_1(\mathbf{x})|^2 + |\nabla d_2(\mathbf{x})|^2) \right) d\mathbf{x}
 \end{aligned} \tag{4}$$

where  $\mathbf{d}(\mathbf{x}) = (d_1(\mathbf{x}), d_2(\mathbf{x}))^\top$ ,  $\Psi(s^2) = \sqrt{s^2 + \epsilon^2}$ ,  $\epsilon = 0.001$  is a predefined small constant to ensure that  $\Psi$  is strictly convex,  $\Phi(\nabla u)$  is a smoothness function equal to  $\exp(-\lambda_{auto}|\nabla u|)$  where  $\lambda_{auto}$  is automatically computed to avoid instabilities and the coefficients  $\gamma$  and  $\alpha$  are the weights. In particular, the third term weighted by  $\alpha$  regularizes the field of displacements: the more  $\alpha$  is large, the more the field is smooth.

**Numerical considerations** Numerically, we work with the disparity maps  $\mathbf{d}$  computed on each pair extracted from  $(u_R, u_G, u_B)$ . We denote respectively by  $\mathbf{d}_{RG}$ ,  $\mathbf{d}_{GB}$  and  $\mathbf{d}_{RB}$  the disparity maps computed respectively between the frames  $u_R$  and  $u_G$ , the frames  $u_G$  and  $u_B$  and the frames  $u_R$  and  $u_B$ . As these three disparity maps are defined on  $\Omega$  and in order to study them according to the variable  $\mathbf{x}_R$ , we have to introduce a fourth map  $\widetilde{\mathbf{d}}_{GB}$  which is defined for each integer position  $\mathbf{x} \in \Omega$  by  $\widetilde{\mathbf{d}}_{GB}(\mathbf{x}) = \mathbf{d}_{GB}(\mathbf{d}_{RG}(\mathbf{x}) + \mathbf{x})$ . Indeed, the position  $\mathbf{y} = \mathbf{d}_{RG}(\mathbf{x}) + \mathbf{x}$  is not an integer in general. The value  $\widetilde{\mathbf{d}}_{GB}(\mathbf{x})$  is computed by bilinear interpolation from  $\mathbf{d}_{GB}$  at the four closest integer positions in the neighborhood of  $\mathbf{d}_{RG}(\mathbf{x}) + \mathbf{x}$ . In addition, for each disparity map we consider the associated modulus map  $\rho$  and direction map  $\theta$ .

### 3.2 Detection Features

In this section we will define the five features that we propose to consider. An illustration of the discriminative quality of each of the proposed features is shown on Figure 4. It represents the empirical probability laws of each feature, according to the nature of the classified pixel from Hollstein et al. [9] ground truth dataset (see Section 4).

**Angular error** We analyze the collinearity of the three optical flows using the angular error (see Figure 3) defined by

$$\phi(\mathbf{x}) = \min_{i,j,k} (|\theta_k(\mathbf{x}) - \theta_j(\mathbf{x})| \bmod 2\pi + |\theta_j(\mathbf{x}) - \theta_i(\mathbf{x})| \bmod 2\pi) \quad (5)$$

for distinct  $i, j, k \in \{RG, \widetilde{GB}, RB\}$ , where  $\theta_i(\mathbf{x})$  is the direction angle of disparity  $\mathbf{d}_i(\mathbf{x})$ . As the maximal angular error is reached when the three normalized optical flows are equidistant on the unit circle, it follows that  $\phi$  has values in  $[0, 4\pi/3)$ . The interest of formula (5) is that it has a large value when two vectors have opposite directions. On cloudy pixels, due to the parallax effect, the three optical flows should have a coherent direction and the angle error  $\phi$  is expected to be small. On the other hand, because the image bands are well registered, cloud free areas should have very small optical flow vectors and the noise should dominate; thus, we expect to observe larger angular errors. This property is illustrated in the middle-right graph of Figure 4.

For this feature, the disparity maps were computed after a rank transform with  $\mathcal{V}_{RT}$  a square patch of size  $35 \times 35$  pixels, and the optical flow ran with the weights  $\alpha = 20$  and  $\gamma = 1$ .

**Composition error ratio** We define the composition flow error ratio (see Figure 3) by

$$\xi(\mathbf{x}) = \frac{|\mathbf{d}_{RB}(\mathbf{x}) - \mathbf{d}_{RG}(\mathbf{x}) - \widetilde{\mathbf{d}}_{GB}(\mathbf{x})|}{|\mathbf{d}_{RB}(\mathbf{x})|}. \quad (6)$$

We justify the use of such a feature because the angular error may not be sufficient: experiments show that the regularization term of Equation (4) used in the optical flow induces sometimes local collinearity even in cloud-free regions. We generally observe the expected relation  $|\mathbf{d}_{RB}(\mathbf{x})| \simeq 2|\mathbf{d}_{RG}(\mathbf{x})| \simeq 2|\widetilde{\mathbf{d}}_{GB}(\mathbf{x})|$  in the cloudy regions while in the cloud-free regions these values are uncorrelated. So, considering in first approximation that these vectors are quite collinear, we

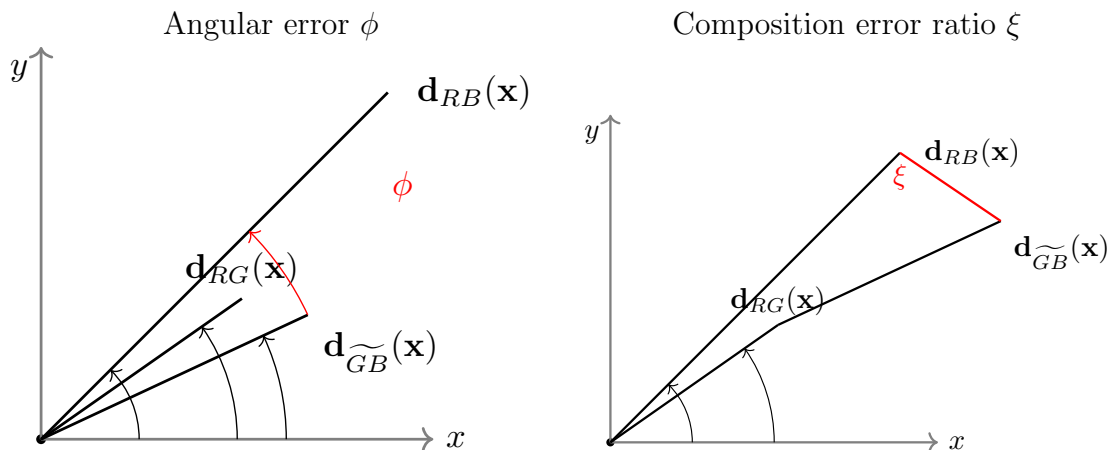


Figure 3: While the three optical flows  $\mathbf{d}_{RG}(\mathbf{x})$ ,  $\mathbf{d}_{GB}(\mathbf{x})$  and  $\mathbf{d}_{RB}(\mathbf{x})$  should theoretically verify Chasles' relation and a perfect collinearity, numerous factors such as the numerical approximations, the flow regularization, the noise, *etc.*, produce non collinear vectors. On the left, we define the angular error  $\phi$  as the angle formed by two of the vectors and corresponding to the smallest angle whose conic domain contains the third vector. On the right, we define the composition error ratio  $\xi$  based on the residual vector (in red).

should observe a ratio  $\xi(\mathbf{x}) \ll 1$  when position  $\mathbf{x}$  is in a cloud, and no particular value when  $\mathbf{x}$  is in a cloud-free region. These properties are illustrated in the bottom-left graph of Figure 4.

For this feature, the disparity maps were computed after a rank transform with  $\mathcal{V}_{RT}$  a square patch of size  $35 \times 35$  pixels, and the optical flow ran with the weights  $\alpha = 20$  and  $\gamma = 1$ .

**Modulus** We take into account the modulus of the displacement between the green and blue bands

$$\rho(\mathbf{x}) = \rho_{GB}(\mathbf{x}) = |\mathbf{d}'_{GB}(\mathbf{x})|. \quad (7)$$

This modulus is computed with a pair of parameters  $(\alpha, \gamma)$  that are different from those of relations (5) and (6). We denote by  $\mathbf{d}'_{GB}$  the associated disparity map. Using the rank transform is a good way to randomize the flow directions in the cloud-free regions. Yet the counterpart is that the flow modulus is often of the same order of magnitude in both cloudy and cloud-free regions. So to make the amplitude flow more accurate in cloud-free regions, we regularize more the flow and use the green and the blue bands which are similar. To save computation time, we limit ourselves to one optical flow for the modulus; we could have defined an average modulus as well, using the three flows. The middle-left graph of Figure 4 shows the empirical distributions of this feature, for both cloud and non-cloud classes.

For this feature, the disparity maps were computed after a rank transform with  $\mathcal{V}_{RT}$  a square patch of size  $35 \times 35$  pixels, and the optical flow ran with the weights  $\alpha = 200$  and  $\gamma = 1$ . The reason for applying a different factor  $\alpha$  relative to the previous two features, is that it acts on both the smoothness of the disparity field and the accuracy of its amplitude. While this second property has a few impact on the angle and the composition error ratio, it is not negligible for the modulus. In particular we observe that the amplitudes on cloud-free areas, which should have values very close to zero, are greater when  $\alpha = 20$  than when  $\alpha = 200$ .

**Luminance** It is straightforward that clouds have a high luminance and, if we except snow, are most of the time brighter than the ground (see the top-left graph of Figure 4). Hence we define the



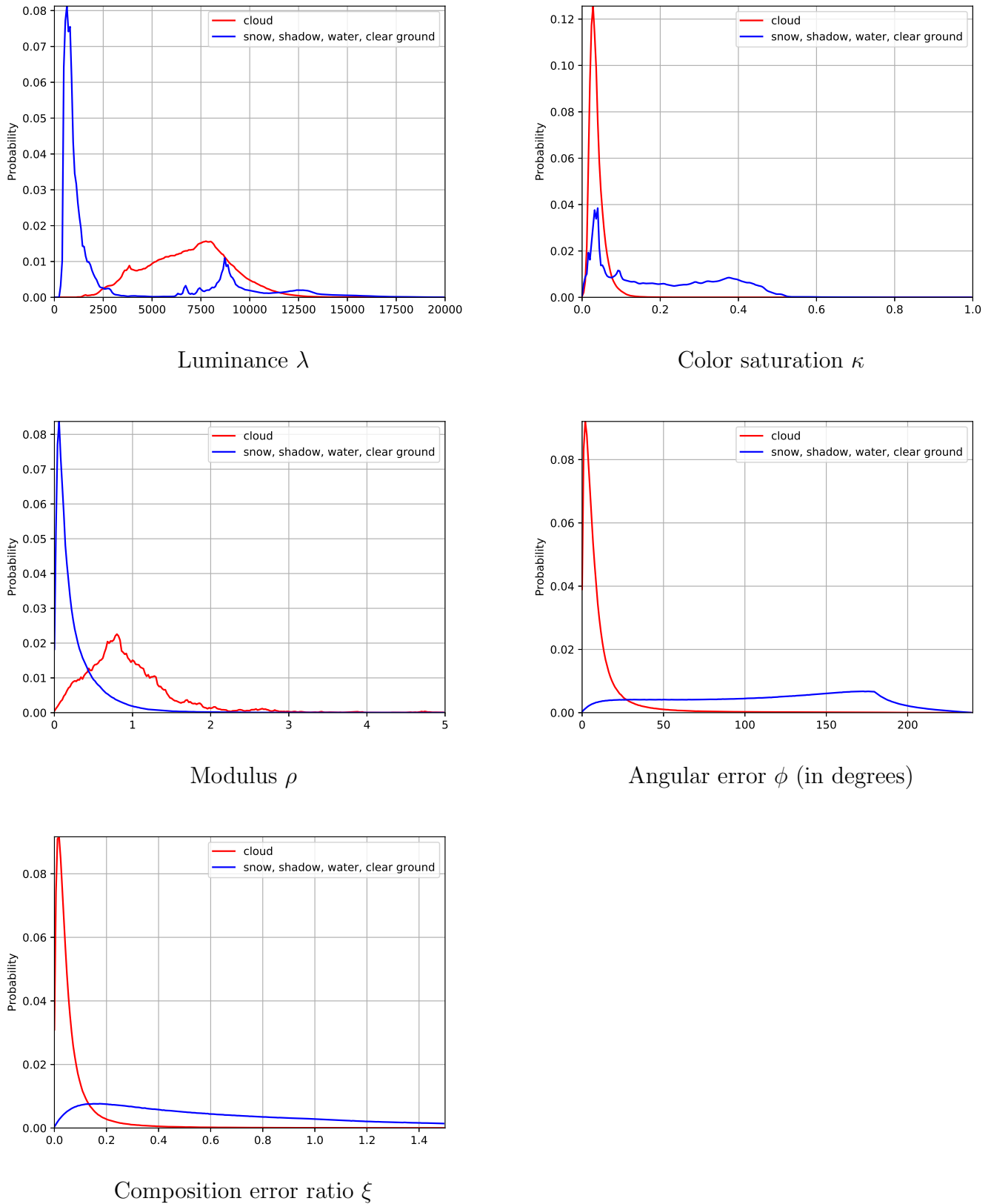


Figure 4: Empirical probability laws of each feature, obtained from the Hollstein et al. [9] ground truth dataset. The disparity maps were computed after a rank transform with  $\mathcal{V}_{RT}$  a square patch of size  $35 \times 35$  pixels. Optical flow ran with the weights  $\alpha = 20, \gamma = 1$  for the features  $\phi$  (Equation (5)) and  $\xi$  (Equation (6)) and  $\alpha = 200, \gamma = 1$  for feature  $\rho$  (Equation (7)). The angular error distribution of the cloud free classes is close to the theoretically uniform distribution in  $[0, 4\pi/3)$ . The peaks observed on the distributions of the cloud free classes for the luminance and the saturation features are mainly due to the snow category.

average luminance feature by

$$\lambda(\mathbf{x}) = \frac{1}{3}(u_R(\mathbf{x}) + u_G(\mathbf{x}) + u_B(\mathbf{x})). \quad (8)$$

**Color saturation** Another criterion is based on the color saturation of the clouds. As in the visible spectral bands their reflectance is quasi-uniform (see Kokhanovsky [10]), they should look gray with various intensities. To quantify this property, we then define  $\kappa$  as the standard deviation

$$\kappa(\mathbf{x}) = \sqrt{\text{var}(\log \mathbf{u}(\mathbf{x}))} = \sqrt{\frac{\sum_{i=1}^k \left( \log(u_i(\mathbf{x})) - \mu(\mathbf{x}) \right)^2}{k}}, \quad (9)$$

with

$$\mu(\mathbf{x}) = \frac{\sum_{i=1}^k \log(u_i(\mathbf{x}))}{k}, \quad (10)$$

where the index  $i$  runs on the  $k$  bands of the image  $\mathbf{u}$  (e.g.  $k = 3$  for RGB images). The usage of the log function penalizes the dark values more by increasing their standard deviation. It leads us to attribute a great value to  $\kappa(\mathbf{x})$  when the luminance is low, and conversely a low value when the intensity is high, for variations of  $\mathbf{u}(\mathbf{x})$  with the same order. The top-right graph of Figure 4 shows the empirical distributions of this feature, for both cloud and non-cloud classes. This feature complements the luminance. Indeed the luminance may be high as soon as one of the bands is saturated. Moreover, these two features assume that an appropriate radiometric calibration was applied to the images (which is the case for example with Sentinel-2 L2A images).

### 3.3 A *Contrario* Multi-features Hypothesis Testing Approach

Our objective is not to classify the clouds as one more category among others such as vegetation or urban, but merely to separate them from cloud-free areas. This amounts to considering only two categories. On the other hand, the characterization of clouds is far from obvious. Indeed, since they have extremely variable motions, shapes and textures, it seems difficult to establish cloud models that are both simple and robust. These reasons led us to favor an *a contrario* statistical method [3], where the data are compared to a null hypothesis  $\mathcal{H}_0$  which, in our case, is the absence of cloud.

For each pixel  $\mathbf{x}$  of  $\Omega$  we shall consider an *a contrario* ground stochastic model

$$\mathcal{H}_0 : Y_{\mathbf{x}} \sim f_{Y_{\mathbf{x}}} \quad (11)$$

where  $Y_{\mathbf{x}}$  is the random variable associated to  $\mathbf{x}$  and  $f_{Y_{\mathbf{x}}}$  is its probability density function, modeling the pixel behavior in absence of clouds. In the case of a simple binary decision, hypothesis testing would lead to threshold the p-value defined by

$$\mathbb{P}(Y_{\mathbf{x}} \geq y_{\mathbf{x}} | \mathcal{H}_0) \quad (12)$$

where  $y_{\mathbf{x}}$  is the observed value of  $Y_{\mathbf{x}}$ . However, as this test is applied separately on each pixel, i.e. more than 100,000 times per image, chances that false rejections of  $\mathcal{H}_0$  happen are huge. To overcome this situation by controlling the total number of false rejections several approaches exist (see [20, § 10.7]). The method computes a *number of false alarms* (NFA)

$$\text{NFA}(y_{\mathbf{x}}) = \#T \times \mathbb{P}(Y_{\mathbf{x}} \geq y_{\mathbf{x}} | \mathcal{H}_0) = \#T \times (1 - F_{Y_{\mathbf{x}}}(y_{\mathbf{x}})) \quad (13)$$

where  $\#T$  is the number of tests and  $F_{Y_{\mathbf{x}}}$  the cumulative distribution function of  $Y_{\mathbf{x}}$ . The NFA method verifies the consistency requirement, under the null hypothesis [6]:

**Definition 1.** Let  $(Y_{\mathbf{x}})_{\mathbf{x} \in \Omega}$  be a set of random variables. A function  $G$  is an NFA (number of false alarms) for the random variables  $\{Y_{\mathbf{x}}\}$  if

$$\forall \epsilon, \quad \mathbb{E} \left[ \sum_{\mathbf{x} \in \Omega} \mathbb{1}_{\{G(Y_{\mathbf{x}}) \leq \epsilon\}} \right] \leq \epsilon. \quad (14)$$

where  $\mathbb{1}$  is the indicator function.

The threshold  $\epsilon$  becomes the input decision parameter to compute the Boolean map  $B$  defined as

$$B(\mathbf{x}) = \mathbb{1}_{\{\text{NFA}(y_{\mathbf{x}}) \leq \epsilon\}}. \quad (15)$$

As we shall use several detection criteria,  $Y_{\mathbf{x}}$  will be a combination of the random variables associated to these features. For each one of these random variables, a probability density function must therefore be estimated.

### 3.4 Feature Combination

Hypothesis testing requires to define random variables for the different features and their associated probability and cumulative distribution functions. We consider the empirical distributions respectively denoted  $\hat{f}$  and  $\hat{F}$  and justify this choice in Section 3.6. To clarify the notation, we shall index by  $c$  the random variables  $X$  associated to the aforementioned criteria and suppose that

$$X_{\mathbf{x},c} \sim \hat{f}_c, \quad \forall \mathbf{x} \in \Omega, \quad \forall c \in \{\lambda, \kappa, \rho, \phi, \xi\}. \quad (16)$$

As the shapes of the distributions  $\hat{F}_c$  of the preceding criteria are all different we cannot apply directly the NFA test. We first apply the following transformations to the previous random variables:

$$W_{\mathbf{x},c} = \hat{F}_c(X_{\mathbf{x},c}), \quad \text{for } c \in \{\rho, \lambda\}, \quad (17)$$

$$W_{\mathbf{x},c} = 1 - \hat{F}_c(X_{\mathbf{x},c}), \quad \text{for } c \in \{\phi, \kappa, \xi\} \quad (18)$$

in order to work with distributions  $W_{\mathbf{x},c}$  which are uniform on  $[0, 1]$  and such that values  $W_{\mathbf{x},c} \approx 1$  should correspond to clouds in all cases. In this way, the hypothesis testing will be performed always on the right tail of the distribution  $Y_{\mathbf{x}}$  (11). (Transformation (18) flips the empirical distributions  $\hat{F}_\phi$ ,  $\hat{F}_\kappa$  and  $\hat{F}_\xi$  so as to evaluate the rare events in the right tail of  $Y_{\mathbf{x}}$ .)

Among the numerous ways to combine the random variables  $W_{\mathbf{x},c}$  into the random variable  $Y_{\mathbf{x}}$  of Equation (11), we propose to use their sum. In addition, to increase the discrimination and ensure some regularity, we consider the 9-neighborhood  $\mathcal{V}(\mathbf{x})$  of the tested pixel defined by

$$\mathcal{V}(\mathbf{x}) = \{(x \pm a, y \pm a), (x, y \pm a), (x \pm a, y), (x, y)\} \quad (19)$$

where  $a \in \mathbb{N}^*$ . In practice, we use  $a \geq 2$  which results in a fairly good decorrelation of the spatial variables. At the image borders we use the constant boundary extension.

**Sum sample based combination** We propose to combine the different features by their sum

$$Y_{\mathbf{x}} = \sum_{\mathbf{y} \in \mathcal{V}(\mathbf{x})} \sum_c W_{\mathbf{y},c}. \quad (20)$$

In the background model  $\mathcal{H}_0$  all the random variables  $W_{\mathbf{y},c}$  follow a uniform distribution. We will also assume that they are also independent random variables. While not strictly true, they are fairly

decorrelated. Then, the sum of independent random variables following a uniform distribution has a closed form distribution called the Irwin-Hall [8] law. Denoting  $n \stackrel{\text{def}}{=} \#\{\lambda, \kappa, \rho, \phi, \xi\} \times |\mathcal{V}(\mathbf{x})|$  the size sample, the cumulative distribution associated to Equation (20) is

$$F_Y(x) = \begin{cases} \frac{1}{n!} \sum_{i=0}^{\lfloor x \rfloor} (-1)^i \binom{n}{i} (x-i)^n, & \forall x \in [0, n], \\ 1, & n \leq x, \\ 0, & x \leq 0. \end{cases} \quad (21)$$

When the neighborhood size increases, the probability distribution  $Y_{\mathbf{x}}$  asymptotically tends to the normal distribution [1]

$$\mathcal{N}\left(\mu = \frac{n}{2}, \sigma = \sqrt{\frac{n}{12}}\right). \quad (22)$$

We use this approximation as soon as  $n$  is higher than 10.

This approach leads to take into account  $\#T = 3|\Omega|$  tests in Equation (13). Indeed, given  $\mathbf{y}$  in the neighborhood  $\mathcal{V}(\mathbf{x})$ , the five features are combined into a single test per pixel from the definition (20) of  $\sum_c W_{\mathbf{y},c}$ . However, when we evaluate *a posteriori* in Equation (5) the realization of the angular error  $\phi$  with the minimal value we need to perform three tests. These three tests must be taken *a priori* into account when we introduce the feature  $\phi$  into the sum  $\sum_c$ . Moreover, the summation over  $\mathcal{V}(\mathbf{x})$  does not change this number per pixel. Finally, summing over the whole image of size  $|\Omega|$  leads to the aforementioned result.

### 3.5 Morphological Post-processing

A number of false or no detection areas may appear spatially isolated in the Boolean map  $B$  defined by Equation (15). These are generally detection errors that can be easily eliminated by considering a surface threshold  $s \leq |\Omega|$  consistent with the image resolution and the minimal acceptable surface of clouds. Denoting  $\mathcal{R}$  the set of all connected regions of image  $B$ , we fill the undesirable holes according to

$$\forall \omega \in \mathcal{R}, \forall \mathbf{x} \in \omega, \tilde{B}(\mathbf{x}) = \begin{cases} 1 - B(\mathbf{x}) & \text{if } |\omega| < s, \\ B(\mathbf{x}) & \text{otherwise.} \end{cases} \quad (23)$$

This task is performed thanks to the Boolean image partitioning approach of Tarjan and Endre [19]. Numerically it is computed by the module `ccproc` of the `imscript` library<sup>2</sup>.

### 3.6 Discussion

With hypothesis testing, using empirical distributions entails an over-fitting risk, because of their explicit relation to the input data. This objection is legitimate for the features  $\lambda$  and  $\kappa$  which are implicitly based on the Sentinel-2 data calibration process. Yet the objection is not serious for optical flow based features. Indeed, the optical flow does not work directly on the input data, and its behaviour depends more on the observable motions coupled with its input parameters. Consequently, images from other satellites than Sentinel-2 can be processed using the same five empirical distributions, provided that their dynamic is adapted to match the typical dynamic of Sentinel-2 images (see Section 4.2).

<sup>2</sup><https://github.com/mnhrdt/imscript>

Another remark concerns the Independence of the variables. The Irwin-Hall law we use to combine the features assumes such a property. On the one hand, it is clear that the regularized optical flow maps induces dependent values in close neighborhoods and so are spatially the three features ( $\phi$ ,  $\xi$  and  $\rho$ ) deduced from them; using  $a \geq 2$  in Equation (19) decreases this dependence. On the other hand there is no evidence of dependency between these features and the pixel value-based ones (i.e. the luminance and the color saturation). In order to take into consideration the dependency of variables, it could be possible to define a more complex statistical model including other mathematical entities (e.g. covariance matrices). Nevertheless, given the fairly decorrelation of the variables, this would probably not lead to a dramatic enhancement of the performances. Thus, we preferred to keep a simple formulation (with some theoretical approximations) which produces good results in practice.

Another remark concerns the use of supplementary input data if such are available. In particular, it is legitimate to question the interest of integrating the near infrared bands 5, 6, 7, 8 and 8A in the algorithm, since some authors like Frantzet al. [4] use some of them. The fact is that these bands are physically located on the same detectors (see Gasconet al. [5]) so that a parallax is observable between these 5 bands. However some of these bands are less resolved, which implies *a priori* less accuracy for the optical flow. Their use would involve calculating a much larger number of optical flows and reviewing the definition of the current criteria, which are mainly based on the existence of a triplet of flows. Furthermore, it is not obvious that increasing the complexity of the algorithm would significantly improve the results.

A final remark concerns our approach itself. Since the parallax observed is due to the displacement of the satellite, one could compute a 1D optical flow in its direction. In addition, this method could increase the flow accuracy and speed up the calculations. Yet we found experimentally that directions differing from the direction of the one induced by the satellite motion are observable. This situation occurs mainly at the border of the 12 staggered detectors mounted in Sentinel-2 focal plane (see Figure 5), or when the clouds are pushed by a sustained wind. Consequently, our 2D approach is more robust to these cases.

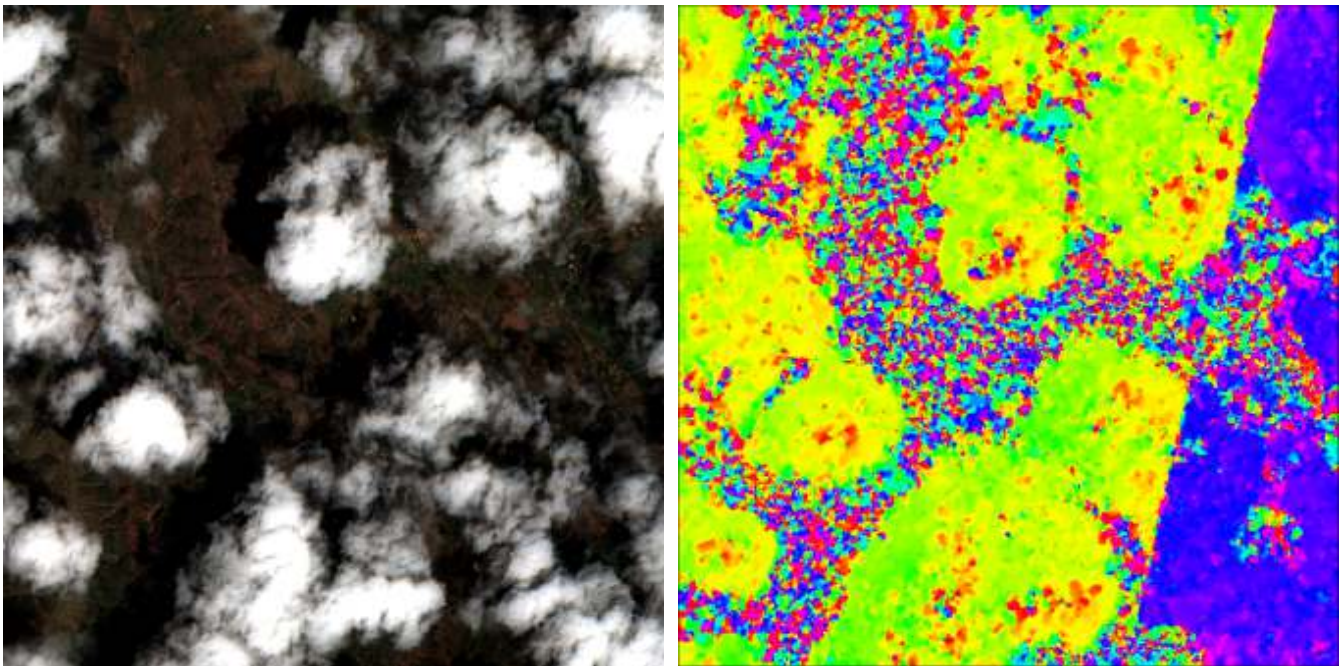


Figure 5: Sentinel-2 images produced by ESA can contain shifted parts due to the staggered positioning of the 12 detectors on the focal planes. This leads to observe opposite directions when computing the optical flow (yellow and purple patterns). This does not decrease the performance of our detector.

## 4 Experiments

We present in this section quantitative results obtained by our method on a Sentinel-2 dataset and show some cloud masks computed on Sentinel-2, RapidEye, SkySat and WorldView-2 images.

### 4.1 Experimental Setup

In order to evaluate quantitatively the performance of the proposed algorithm, we used the Sentinel-2 dataset created by Hollstein et al. [9]. This dataset is composed of 108 scenes located all over the world (each is a  $10980 \times 10980$  pixels tile) with their ground truth represented by manually labelled polygons. We retrieved only 100 tiles of them, the others being unavailable (see Table 3 for the complete list of the tested tiles). As the ground truths are sparse, we split each tile into images of size  $366 \times 366$ , giving a series of 900 images per tile. We processed only the annotated ones, leading to a full data set of 9789 images.

The disparity maps were computed by the optical flow algorithm parameterized with  $\alpha = 20, \gamma = 1$ , for the features (5) and (6) and  $\alpha = 200, \gamma = 1$  for the modulus (7) after the application of the rank transform on the visible R,G,B bands with  $\mathcal{V}_{RT}$  a square patch of size  $35 \times 35$  pixels. The 9-neighborhood step  $a$  is set to 3.

The original classification of the dataset has six categories: opaque clouds, cirrus, clear ground, water, shadows and snow. The cirrus category has not been taken into account because of the numerous ambiguities in its annotation. Indeed, the ground truth classification of cirrus is based on their meteorological definition. However, in terms of visibility, a cirrus can be translucent when thin, or opaque when several layers superimpose. According to the capabilities of our algorithm, we reduced this classification to two groups: opaque clouds and cloud-free, the latter containing clear ground, water, shadow and snow. We consider the first group as the positive examples and the second as the negative examples whose populations are respectively  $P = 3,567,188$  pixels and  $N = 16,188,312$  pixels.

### 4.2 Experimental Results

**Sentinel-2 dataset evaluation** To identify the various feature combinations that we evaluate, we use the following abbreviations:

- M: average modulus  $\rho$ , Equation (7);
- A: angular error  $\phi$ , Equation (5);
- R: composition error ratio  $\xi$ , Equation (6);
- K: color saturation  $\kappa$ , Equation (9);
- L: luminance  $\lambda$ , Equation (8).

The metrics with which we evaluate the performances of classification are

- the balanced accuracy :  $BACC = (TNR + TPR)/2$ ;
- the accuracy :  $ACC = (TP + TN)/(P + N)$ ;
- the  $F_1$  score :  $F_1 = \frac{2TP}{2TP + FP + FN}$ ;

where  $TNR = TN / (TN + FP)$  is the true negative rate,  $TPR = TP / (TP + FN)$  the true positive rate and TP, TN and FN are respectively the number of pixels correctly identified as opaque cloud, non cloud and incorrectly identified as non cloud. In the present case, we favor the balanced accuracy in the interpretation. It is more appropriate than the accuracy metric and the  $F_1$  score because the populations are imbalanced: positive examples represent only 18,05% of the ground truth (see Sokolova et al. [18]).

We present in Figure 6 the results of the true positive rate, false positive rate and balanced accuracy, obtained for different combinations of features and a 9 pixel neighborhood with an  $a = 3$  pixel step and after a post-processing step where  $s = 100$  when varying the  $\log \epsilon$  threshold. The value of the parameter  $s$  is equivalent to a surface of 1 ha, according to the Sentinel-2 resolution. Table 2 shows the maximal scores in term of balanced accuracy, accuracy and  $F_1$  score achieved by these combinations.

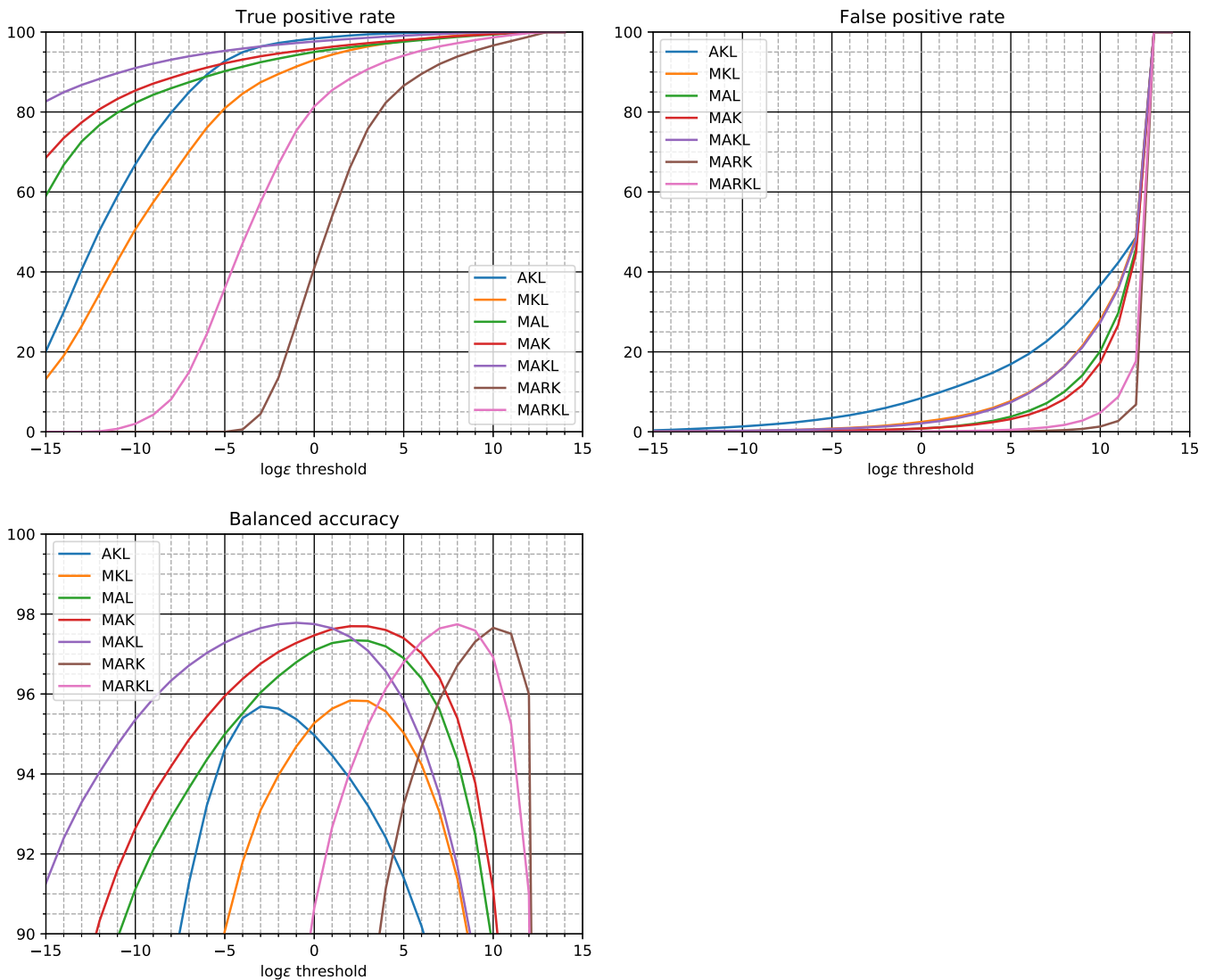


Figure 6: True positive rate, false positive rate and balanced accuracy expressed in percents for various combinations.

Overall the results are quite satisfactory, because the combinations considered have in majority an average balanced accuracy higher than 97.5%. These results quantitatively show the advantage of using the parallax phenomenon. On the one hand, by taking the displacement modulus into

Table 2: Maximal values obtained by various combinations. Taking into account the acceptable number of false alarms, the MAK and MAKL combinations have the best performances.

Combination	$\log \epsilon$	Balanced accuracy	Accuracy	F1 score
AKL	-3	95.69	95.83	88.92
MKL	2	95.84	96.83	91.20
MAL	2	97.35	98.45	95.63
MAK	<b>2</b>	<b>97.70</b>	98.60	96.07
MAKL	<b>-1</b>	<b>97.78</b>	98.56	95.98
MARK	8	97.66	98.57	96.00
MARKL	10	97.75	98.54	95.94

account, the balanced accuracy is increased by at least 2 points by going from the AKL combination to MAKL. On the other hand, the measurement of the coherence of the optical flow, represented by the angular error, is also a preponderant factor to increase the precision: while the combinations MKL and MAKL have the same rate of false alarms, the rate of true positives is significantly higher for MAKL. The balanced precision is thus increased by nearly two points when switching from MKL to MAKL. Adding the R criterion does not improve the scores and shifts the MARK and MARKL curves to higher  $\log \epsilon$  values, increasing the risk of false alarms.

When considering the MAL and MAK combinations, the K criterion seems more discriminating than the luminance. When the two criteria K and L are combined, there is little improvement in the scores, and we observe a translation of the MAKL curve of the balanced accuracy, compared to those of MAK and MAL, according to a lower value  $\log \epsilon$ . These two phenomena seem to indicate that the parameters K and L are quite dependent.

Thus, the criterion K based on the color saturation seems necessary to achieve good performance, confirming that the optical flow based features are incomplete. This justifies carrying out a good correction and a white balance during the preprocessing of the raw images in order to stay within the Kokhanovsky observations [10].

Finally, according to the curves and the results of the table, the best combinations are MAK and MAKL. Figures 7 and 8 show some qualitative results on various grounds, using the MAK combination.

**CDI index comparison** Figure 9 shows several results obtained with the Frantz et al. [4] approach (see Section 2). The CDI index over-detects cloudy regions most of the time. This behavior seems to confirm (see [4, §3.3]) that the CDI index is more of a means to corroborate the cloud detections in the Fmask processing workflow, rather than a standalone detector.

**Algorithm generalization evaluation** In order to test the generalization of the method we applied it on scenes from RapidEye, WorldView-2 and SkySat. As we could not compute the empirical probabilities of the aforementioned images by lack of data, we adapted their dynamics to fit those of the Sentinel-2 images. While the RapidEye, SkySat and Sentinel-2 images have the same dynamics, we linearly stretched the SkySat dynamics by a factor 16.

Results of the RapidEye and SkySat scenes are shown in Figure 10 and WorldView-2 in Figure 11. The results are globally satisfactory, all the more so in the case of the WorldView-2 image, because we no longer use the criteria M, K and L. Indeed, the pushbroom mechanism of WorldView-2 has only one shift so that there is no parallax between the blue and green bands.



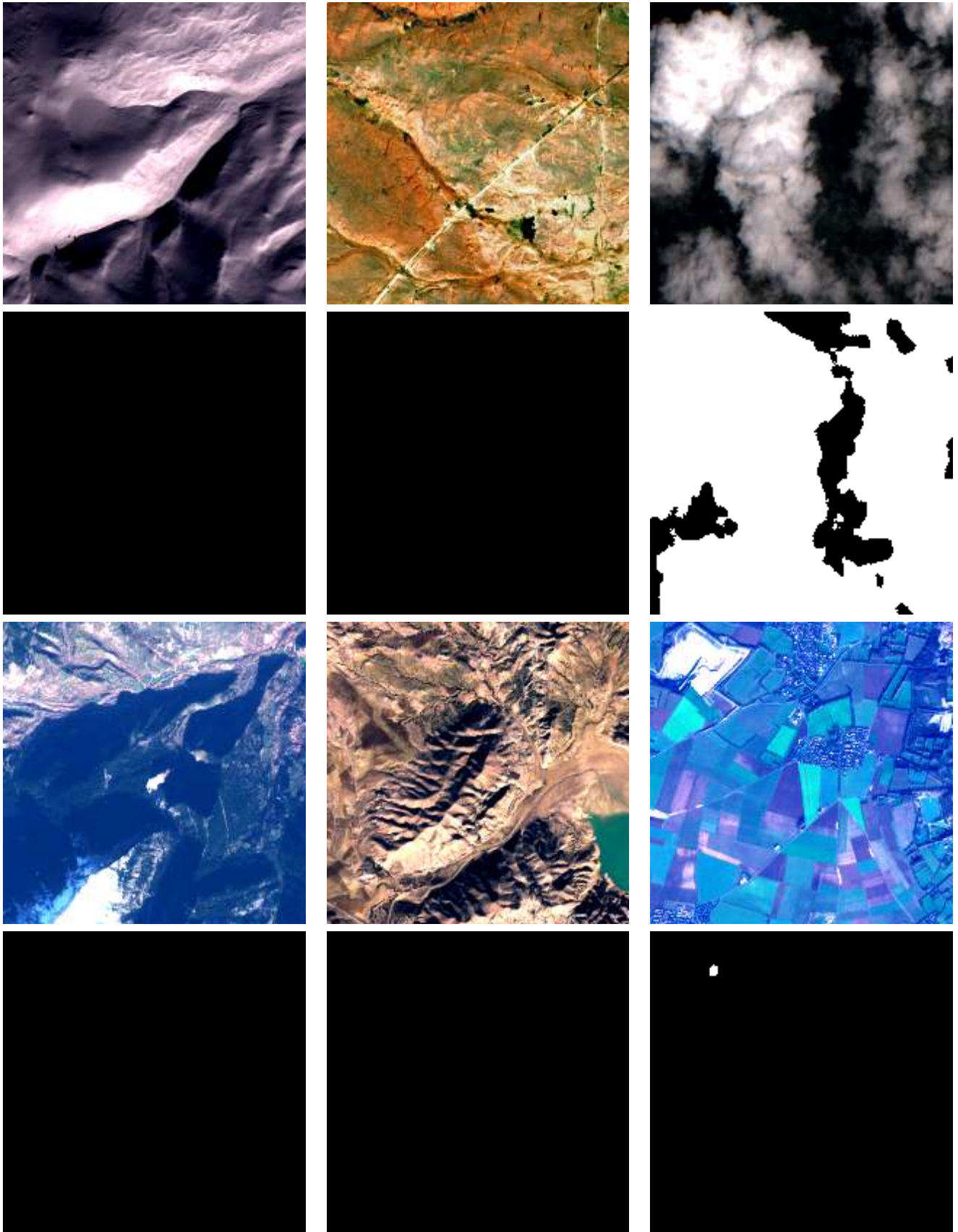


Figure 7: Examples of scenes extracted from the Sentinel-2 Hollstein dataset with their associated  $\tilde{B}$  cloud maps estimated by our method (MAK combination with  $a = 3$ ,  $\epsilon = 1$ ,  $s = 100$ ).

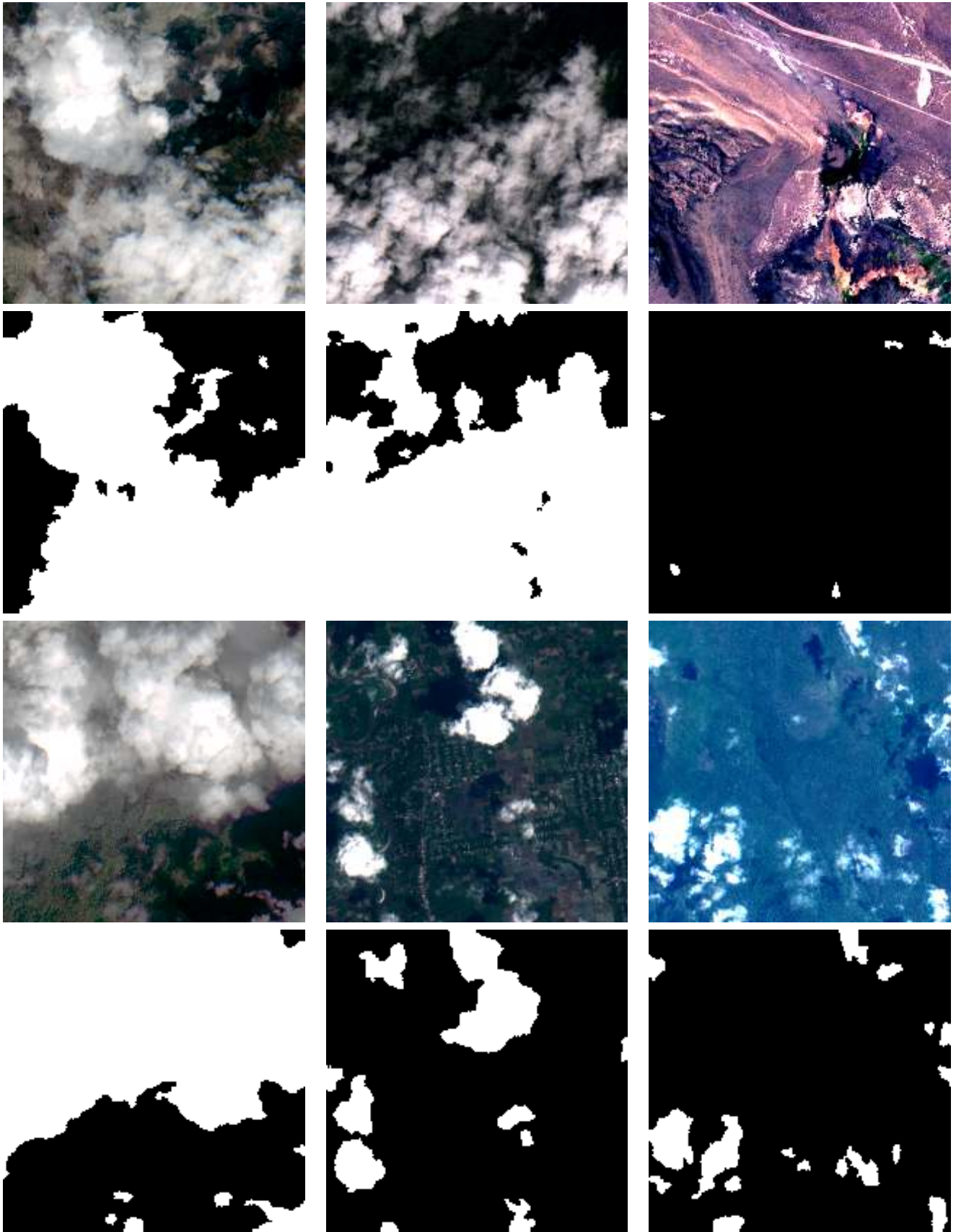


Figure 8: Examples of scenes extracted from the Sentinel Hollstein dataset with their associated  $\tilde{B}$  cloud maps estimated by our method (MAK combination with  $a = 3$ ,  $\epsilon = 1$ ,  $s = 100$ ).

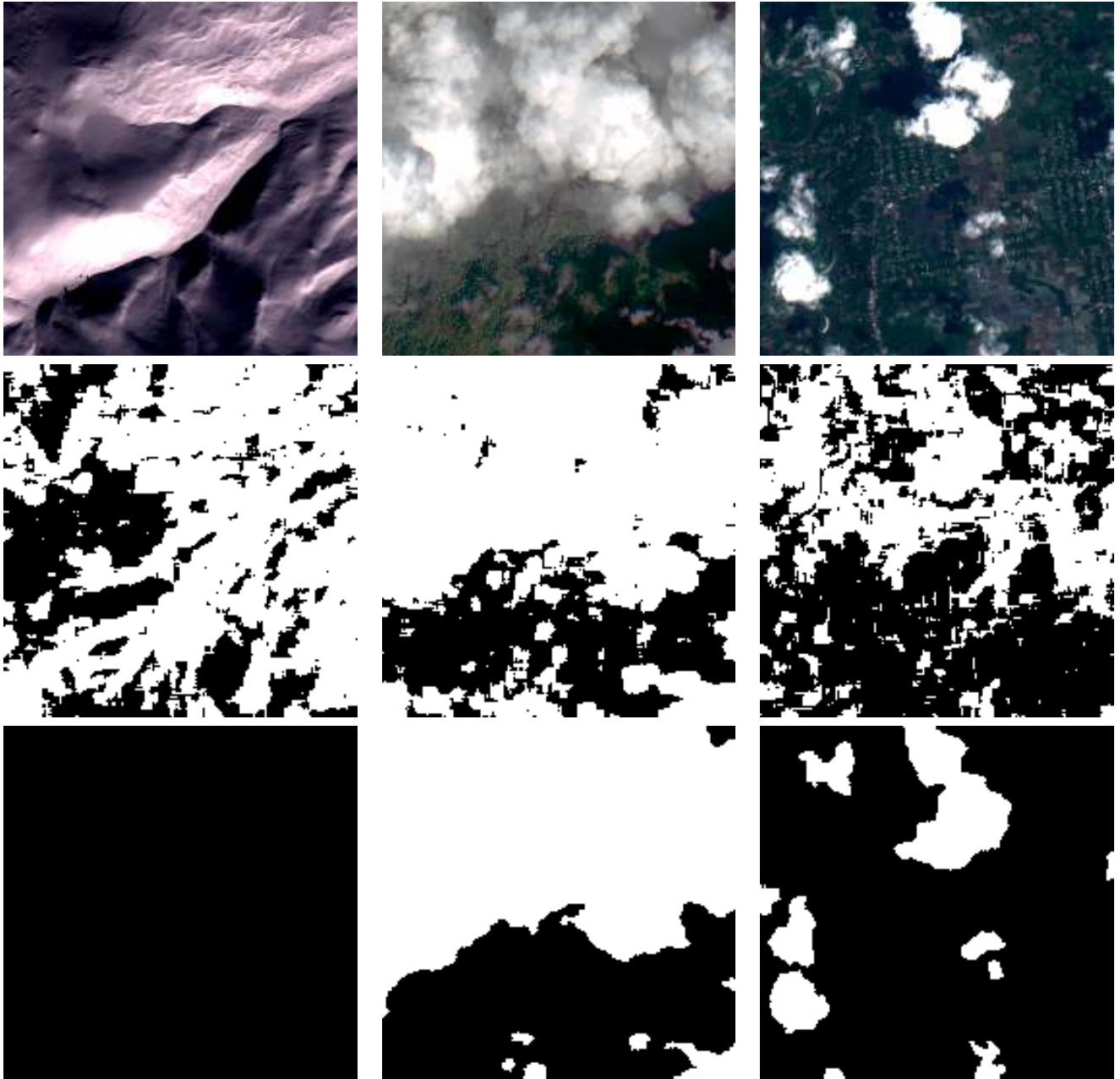


Figure 9: Examples of scenes (top row) extracted from the Sentinel Hollstein dataset with their associated CDI maps estimated by the Frantz et al. [4] method (middle row) and by our method (bottom row).

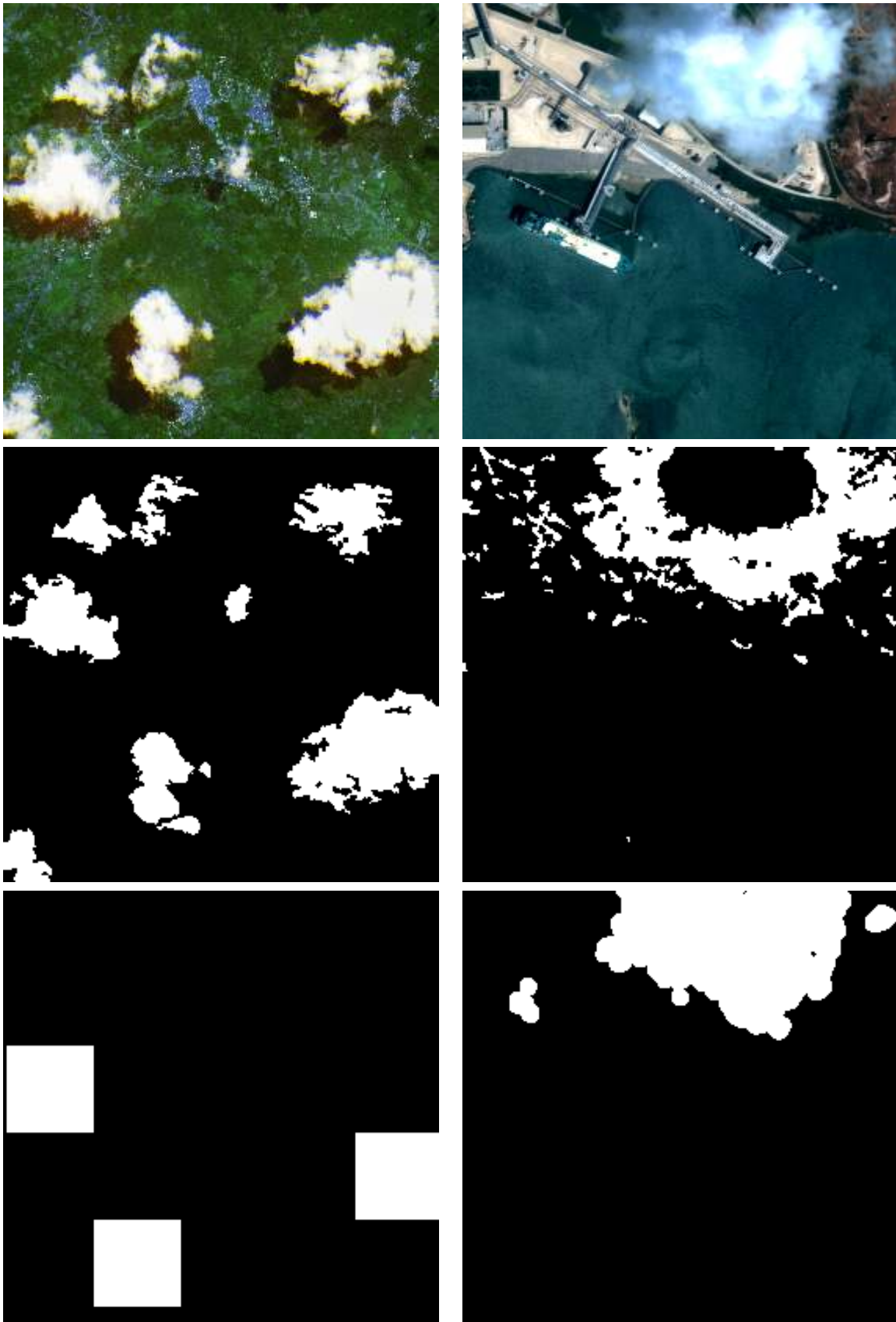


Figure 10: Application of our method to a scene acquired by RapidEye (up left) and SkySat (up right) imagers. Middle line, our cloud maps computed according to the MAK summation criteria strategy with  $a = 3, \log \epsilon = 0$  and  $s = 100$ . Bottom, the UDM cloud maps provided by PlanetScope. The relative failure in the Skysat detection (up right) comes from the absence of an observable motion at the center of the cloud impacting the features M and A.



Figure 11: On the left, a scene acquired by WorldView-2. On the right, our cloud map computed according the MKL combination with  $a = 3$ ,  $\log \epsilon = 7$  and  $s = 100$ . Notice that the features A and R are ineffective because the pushbroom has only one shift, so that green and blue bands are synchronized. The cloud map is correctly estimated, in particular in the translucent regions located on the top and with few false alarms.

## 5 Conclusion

In this paper we have presented an unsupervised algorithm for detecting clouds in images made of three asynchronous spectral bands acquired by a pushbroom satellite. Our method has three steps: histogram equalization of the image bands by the rank transform, disparity map estimation, cloud detection by multi-features hypothesis testing. Quantitative experiments were conducted on a representative dataset of Sentinel-2 scenes. Our method reaches success rates of over 97.5% for both the cloudy and cloud free areas. Our experiments on RapidEye, SkySat and WorldView-2 scenes also show satisfactory results when compared with their native cloud maps.

In addition, this study allowed us to show that the parallax effect is highly discriminant for the cloud detection problem. We think therefore that optical flow maps could be used as an input in deep learning methods to reduce the architecture complexity or the training dataset size.

## Acknowledgments

Work partly financed by Office of Naval research grant N00014-20-S-B001, DGA Astrid project “filmer la Terre” n° ANR-17-ASTR-0013-01 and Kayrros, Inc.

## Image Credits

Images of Figures 2 (up left, bottom left), 7, 8 and 9 are excerpts from the Hollstein et al. dataset [9]. All the other images in this manuscript were produced by the authors (license CC-BY-SA).

# Input Data

The complete list of the tiles used in this study from the Hollstein et al. [9] dataset is shown on Table 3.

Table 3: Complete list of the tiles used in this study from the Hollstein et al. [9] dataset.

S2A_OPER_MSI_L1C_TL_MPS__20160210T132231_A003321_T38TKL_N02_01	S2A_OPER_MSI_L1C_TL_MPS__20160210T132231_A003321_T38TLL_N02_01
S2A_OPER_MSI_L1C_TL_MPS__20160506T134142_A004552_T35VLJ_N02_02	S2A_OPER_MSI_L1C_TL_MPS__20160831T133451_A006225_T35VLJ_N02_04
S2A_OPER_MSI_L1C_TL_MPS__20160921T112552_A006524_T36JTS_N02_04	S2A_OPER_MSI_L1C_TL_MTI__20151204T202604_A002354_T15TVJ_N02_00
S2A_OPER_MSI_L1C_TL_MTI__20151204T202604_A002354_T15TVJ_N02_00	S2A_OPER_MSI_L1C_TL_MTI__20151206T095033_A002377_T37PFP_N02_00
S2A_OPER_MSI_L1C_TL_MTI__20151206T095033_A002377_T37PGQ_N02_00	S2A_OPER_MSI_L1C_TL_MTI__20151209T100106_A002420_T38SKB_N02_00
S2A_OPER_MSI_L1C_TL_MTI__20151209T100106_A002420_T38SLB_N02_00	S2A_OPER_MSI_L1C_TL_MTI__20151209T100106_A002420_T38SLF_N02_00
S2A_OPER_MSI_L1C_TL_MTI__20151209T100106_A002420_T38SMD_N02_00	S2A_OPER_MSI_L1C_TL_MTI__20151209T100106_A002420_T38SMG_N02_00
S2A_OPER_MSI_L1C_TL_MTI__20151218T200553_A002555_T11SPD_N02_01	S2A_OPER_MSI_L1C_TL_MTI__20151218T200553_A002555_T12SUJ_N02_01
S2A_OPER_MSI_L1C_TL_MTI__20151225T115351_A002649_T36NTG_N02_01	S2A_OPER_MSI_L1C_TL_MTI__20151227T203753_A002681_T19FCD_N02_01
S2A_OPER_MSI_L1C_TL_MTI__20151227T203753_A002681_T19FCE_N02_01	S2A_OPER_MSI_L1C_TL_MTI__20151227T203753_A002681_T19FEE_N02_01
S2A_OPER_MSI_L1C_TL_MTI__20151231T201721_A002741_T11SLT_N02_01	S2A_OPER_MSI_L1C_TL_MTI__20160122T190934_A003053_T19HEB_N02_01
S2A_OPER_MSI_L1C_TL_MTI__20160204T095011_A003235_T36MYV_N02_01	S2A_OPER_MSI_L1C_TL_MTI__20160204T095011_A003235_T37MCS_N02_01
S2A_OPER_MSI_L1C_TL_MTI__20160207T114227_A003278_T35JPJ_N02_01	S2A_OPER_MSI_L1C_TL_MTI__20160207T114227_A003278_T35JQK_N02_01
S2A_OPER_MSI_L1C_TL_MTI__20160210T195026_A003324_T27VWL_N02_01	S2A_OPER_MSI_L1C_TL_MTI__20161115T215646_A007317_T11SNC_N02_04
S2A_OPER_MSI_L1C_TL_SGS__20151203T163340_A002336_T30SXH_N02_00	S2A_OPER_MSI_L1C_TL_SGS__20151203T163340_A002336_T30SYH_N02_00
S2A_OPER_MSI_L1C_TL_SGS__20151206T075514_A002375_T46RFU_N02_00	S2A_OPER_MSI_L1C_TL_SGS__20151206T075514_A002375_T46RGV_N02_00
S2A_OPER_MSI_L1C_TL_SGS__20151206T113048_A002378_T34SDH_N02_00	S2A_OPER_MSI_L1C_TL_SGS__20151206T113048_A002378_T34SEJ_N02_00
S2A_OPER_MSI_L1C_TL_SGS__20151206T113048_A002378_T34SFH_N02_00	S2A_OPER_MSI_L1C_TL_SGS__20151207T091210_A002390_T44SKE_N02_00
S2A_OPER_MSI_L1C_TL_SGS__20151207T091210_A002390_T44SLF_N02_00	S2A_OPER_MSI_L1C_TL_SGS__20151207T091210_A002390_T44TMQ_N02_00
S2A_OPER_MSI_L1C_TL_SGS__20151207T091210_A002390_T44TNP_N02_00	S2A_OPER_MSI_L1C_TL_SGS__20151208T152443_A002407_T32SNE_N02_00
S2A_OPER_MSI_L1C_TL_SGS__20151208T152443_A002407_T32SPE_N02_00	S2A_OPER_MSI_L1C_TL_SGS__20151209T152100_A002422_T30UWB_N02_00
S2A_OPER_MSI_L1C_TL_SGS__20151209T152100_A002422_T30UXC_N02_00	S2A_OPER_MSI_L1C_TL_SGS__20151209T152100_A002422_T30UXD_N02_00
S2A_OPER_MSI_L1C_TL_SGS__20151209T152100_A002422_T30UXV_N02_00	S2A_OPER_MSI_L1C_TL_SGS__20151209T152100_A002422_T30UYB_N02_00
S2A_OPER_MSI_L1C_TL_SGS__20151211T123450_A002449_T36RUU_N02_00	S2A_OPER_MSI_L1C_TL_SGS__20151211T123450_A002449_T36RVV_N02_00
S2A_OPER_MSI_L1C_TL_SGS__20151211T173733_A002450_T32SLE_N02_00	S2A_OPER_MSI_L1C_TL_SGS__20151211T173733_A002450_T32SMF_N02_00
S2A_OPER_MSI_L1C_TL_SGS__20151216T151003_A002522_T30SVJ_N02_01	S2A_OPER_MSI_L1C_TL_SGS__20151216T151003_A002522_T30TVK_N02_01
S2A_OPER_MSI_L1C_TL_SGS__20151223T181253_A002624_T19HDE_N02_01	S2A_OPER_MSI_L1C_TL_SGS__20151223T181253_A002624_T19HEE_N02_01
S2A_OPER_MSI_L1C_TL_SGS__20151223T181253_A002624_T19JFG_N02_01	S2A_OPER_MSI_L1C_TL_SGS__20151224T161331_A002636_T33UUU_N02_01
S2A_OPER_MSI_L1C_TL_SGS__20151224T161331_A002636_T33UVU_N02_01	S2A_OPER_MSI_L1C_TL_SGS__20151226T112710_A002663_T36LYN_N02_01
S2A_OPER_MSI_L1C_TL_SGS__20151226T112710_A002663_T36LYN_N02_01	S2A_OPER_MSI_L1C_TL_SGS__20151228T085259_A002687_T54HYD_N02_01
S2A_OPER_MSI_L1C_TL_SGS__20151228T085259_A002687_T55HCA_N02_01	S2A_OPER_MSI_L1C_TL_SGS__20151228T085259_A002687_T55HCU_N02_01
S2A_OPER_MSI_L1C_TL_SGS__20151229T114601_A002706_T37PCN_N02_01	S2A_OPER_MSI_L1C_TL_SGS__20151229T114601_A002706_T37PDQ_N02_01
S2A_OPER_MSI_L1C_TL_SGS__20160102T195442_A002768_T16PHS_N02_01	S2A_OPER_MSI_L1C_TL_SGS__20160102T195442_A002768_T17PKK_N02_01
S2A_OPER_MSI_L1C_TL_SGS__20160102T195442_A002768_T17PKL_N02_01	S2A_OPER_MSI_L1C_TL_SGS__20160109T211805_A002867_T18FXM_N02_01
S2A_OPER_MSI_L1C_TL_SGS__20160109T211805_A002867_T18GYP_N02_01	S2A_OPER_MSI_L1C_TL_SGS__20160109T230542_A002870_T10UEV_N02_01
S2A_OPER_MSI_L1C_TL_SGS__20160109T230542_A002870_T10UGA_N02_01	S2A_OPER_MSI_L1C_TL_SGS__20160112T094610_A002903_T49MGM_N02_01
S2A_OPER_MSI_L1C_TL_SGS__20160112T094610_A002903_T49MGN_N02_01	S2A_OPER_MSI_L1C_TL_SGS__20160112T094610_A002903_T49MHM_N02_01
S2A_OPER_MSI_L1C_TL_SGS__20160116T091329_A002960_T50MRD_N02_01	S2A_OPER_MSI_L1C_TL_SGS__20160116T091329_A002960_T51MTT_N02_01
S2A_OPER_MSI_L1C_TL_SGS__20160121T065055_A003030_T55GDN_N02_01	S2A_OPER_MSI_L1C_TL_SGS__20160121T065055_A003030_T55GEN_N02_01
S2A_OPER_MSI_L1C_TL_SGS__20160126T055615_A003103_T49JGL_N02_01	S2A_OPER_MSI_L1C_TL_SGS__20160126T055615_A003103_T49JHM_N02_01
S2A_OPER_MSI_L1C_TL_SGS__20160205T174515_A003251_T31TGL_N02_01	S2A_OPER_MSI_L1C_TL_SGS__20160205T174515_A003251_T32TLR_N02_01
S2A_OPER_MSI_L1C_TL_SGS__20160215T212109_A003398_T16WFT_N02_01	S2A_OPER_MSI_L1C_TL_SGS__20160215T212109_A003398_T16WFU_N02_01
S2A_OPER_MSI_L1C_TL_SGS__20160217T115519_A003421_T37PDT_N02_01	S2A_OPER_MSI_L1C_TL_SGS__20160217T115519_A003421_T37PET_N02_01
S2A_OPER_MSI_L1C_TL_SGS__20160217T182738_A003423_T29RMP_N02_01	S2A_OPER_MSI_L1C_TL_SGS__20160217T182738_A003423_T29RNQ_N02_01
S2A_OPER_MSI_L1C_TL_SGS__20160217T182738_A003423_T29RPQ_N02_01	S2A_OPER_MSI_L1C_TL_SGS__20160217T182738_A003423_T29SQR_N02_01
S2A_OPER_MSI_L1C_TL_SGS__20160310T173833_A003737_T32TNR_N02_01	S2A_OPER_MSI_L1C_TL_SGS__20160323T160631_A003923_T32TMR_N02_01
S2A_OPER_MSI_L1C_TL_SGS__20160327T140224_A003980_T35VLJ_N02_01	S2A_OPER_MSI_L1C_TL_SGS__20160417T165340_A004281_T29RPQ_N02_01
S2A_OPER_MSI_L1C_TL_SGS__20160419T160334_A004309_T32TNR_N02_01	S2A_OPER_MSI_L1C_TL_SGS__20160422T142445_A004352_T32TMR_N02_01
S2A_OPER_MSI_L1C_TL_SGS__20160624T173213_A005253_T32TMR_N02_04	S2A_OPER_MSI_L1C_TL_SGS__20161029T172901_A007069_T32TMR_N02_04
S2A_OPER_MSI_L1C_TL_SGS__20161108T154744_A007212_T32TNR_N02_04	S2A_OPER_MSI_L1C_TL_SGS__20161108T154744_A007212_T32TNS_N02_04

## References

- [1] H. Cramér. *Mathematical Methods of Statistics (PMS-9)*. Princeton University Press, 1999. ISBN 9780691005478, <http://www.jstor.org/stable/j.ctt1bpm9r4>.
- [2] T. Dagobert, N. Monzón, and J. Sánchez. Comparison of Optical Flow Methods under Stereomatching with Short Baselines. *Image Processing On Line*, 9:329–359, 2019. <https://doi.org/10.5201/ipol.2019.217>.
- [3] A. Desolneux, L. Moisan, and J. M. Morel. *From Gestalt Theory to Image Analysis*. Springer, 2008.
- [4] D. Frantz, E. Haß, A. Uhl, J. Stoffels, and J. Hill. Improvement of the Fmask algorithm for Sentinel-2 images: Separating clouds from bright surfaces based on parallax effects. *Remote Sensing of Environment*, 215:471 – 481, 2018. <https://doi.org/10.1016/j.rse.2018.04.046>.
- [5] F. Gascon, C. Bouzinac, O. Thépaut, M. Jung, B. Francesconi, J. Louis, V. Lonjou, B. Lafrance, S. Massera, A. Gaudel-Vacaresse, F. Languille, B. Alhammoud, F. Viallefont, B. Pflug, J. Bieniarz, S. Clerc, L. Pessiot, T. Trémas, E. Cadau, R. De Bonis, C. Isola, P. Martimort, and V. Fernandez. Copernicus Sentinel-2A Calibration and Products Validation Status. *Remote Sensing*, 9(6), 2017. <https://doi.org/10.3390/rs9060584>.
- [6] B. Grosjean and L. Moisan. A-contrario detectability of spots in textured backgrounds. *Journal of Mathematical Imaging and Vision*, 33(3):313, Sep 2008. <https://doi.org/10.1007/s10851-008-0111-4>.
- [7] T. Guillemot and J. Delon. Implementation of the Midway Image Equalization. *Image Processing On Line*, 6:114–129, 2016. <https://doi.org/10.5201/ipol.2016.140>.
- [8] P. Hall. The Distribution of Means for Samples of Size N Drawn from a Population in which the Variate Takes Values Between 0 and 1, All Such Values Being Equally Probable. *Biometrika*, 19(3/4):240–245, 1927. <http://dx.doi.org/10.2307/2331961>.
- [9] 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(8), 2016. <https://doi.org/10.3390/rs8080666>.
- [10] A. Kokhanovsky. Optical properties of terrestrial clouds. *Earth-Science Reviews*, 64(3):189 – 241, 2004. [https://doi.org/10.1016/S0012-8252\(03\)00042-4](https://doi.org/10.1016/S0012-8252(03)00042-4).
- [11] K. F. Manizade, J. D. Spinhirne, and R. S. Lancaster. Stereo Cloud Heights From Multi-spectral infrared imagery via Region-of-Interest Segmentation. *IEEE Transactions on Geoscience and Remote Sensing*, 44(9):2481–2491, Sept 2006. <https://doi.org/10.1109/TGRS.2006.873339>.
- [12] N. Monzón, A. Salgado, and J. Sánchez. Robust Discontinuity Preserving Optical Flow Methods. *Image Processing On Line*, 6:165–182, 2016. <https://doi.org/10.5201/ipol.2016.172>.
- [13] J. Muller, A. Mandanayake, C. Moroney, R. Davies, D. J. Diner, and S. Paradise. MISR stereoscopic image matchers: techniques and results. *IEEE Transactions on Geoscience and Remote Sensing*, 40(7):1547–1559, July 2002. <https://doi.org/10.1109/TGRS.2002.801160>.

- [14] C. Panem, S. Baillarin, C. Latry, H. Vadon, and P. Dejean. Automatic cloud detection on high resolution images. In *IEEE International Geoscience and Remote Sensing Symposium*, volume 1, 2005. <https://doi.org/10.1109/IGARSS.2005.1526222>.
- [15] N. Sabater, S. Leprince, and J. Avouac. Contrast invariant and affine sub-pixel optical flow. In *IEEE International Conference on Image Processing*, pages 53–56, 2012. <https://doi.org/10.1109/ICIP.2012.6466793>.
- [16] D. Shin and J. K. Pollard. Cloud height determination from satellite stereo images. In *IEEE Colloquium on Image Processing for Remote Sensing*, pages 4/1–4/7, Feb 1996. <https://doi.org/10.1049/ic:19960158>.
- [17] K. Sinclair, B. van Dienenhoven, B. Cairns, J. Yorks, A. Wasilewski, and M. McGill. Remote sensing of multiple cloud layer heights using multi-angular measurements. *Atmospheric Measurement Techniques Discussions*, pages 1–23, 02 2017. <https://doi.org/10.5194/amt-2017-2>.
- [18] M. Sokolova, N. Japkowicz, and S. Szpakowicz. Beyond Accuracy, F-Score and ROC: A Family of Discriminant Measures for Performance Evaluation. volume Vol. 4304, pages 1015–1021, 01 2006. [https://doi.org/10.1007/11941439\\_114](https://doi.org/10.1007/11941439_114).
- [19] R. E. Tarjan. Efficiency of a good but not linear set union algorithm. *ACM*, 22(2):215–225, Apr. 1975. <https://doi.org/10.1145/321879.321884>.
- [20] L. Wasserman. *All of Statistics: A Concise Course in Statistical Inference*. Springer Publishing Company, Incorporated, 2010. ISBN 1441923225, 9781441923226.
- [21] T. Wu, X. Hu, Y. Zhang, L. Zhang, P. Tao, and L. Lu. Automatic cloud detection for high resolution satellite stereo images and its application in terrain extraction. *ISPRS Journal of Photogrammetry and Remote Sensing*, 121:143 – 156, 2016. <http://dx.doi.org/10.1016/j.isprsjprs.2016.09.006>.
- [22] R. Zabih and J. Woodfill. Non-parametric local transforms for computing visual correspondence. In J.-O. Eklundh, editor, *Proceedings of Third European Conference on Computer Vision, Volume II*, pages 151–158. Springer Berlin Heidelberg, 1994. <http://dx.doi.org/10.1007/BFb0028345>.