

Automatic Detection of Clouds from Aerial Photographs of Snowy Volcanoes

Carolina Chang^{1,2}(✉) and Fernando Vaca³

¹ Grupo de Inteligencia Artificial, Universidad Simón Bolívar, Caracas, Venezuela

² PROMETEO Researcher, Instituto Geográfico Militar, Quito, Ecuador
cchang@usb.ve

³ Instituto Geográfico Militar, Quito, Ecuador
fernando.vaca@mail.igm.gob.ec

Abstract. We propose a method for cloud detection from RGB aerial photographs of snow-capped volcanoes of Ecuador. For cartography purposes, clouds are undesired objects that occlude the terrain, while snow-covered areas are valid regions of a map. The traditional approach of image thresholding does not suffice when snowy areas cannot be dismissed from the image in advanced. We combine image thresholding with region growing and neural networks classification to detect clouds at the object level. We show that there is overlap at the pixel level of clouds and snow. At the classification task a fuzzy ARTMAP neural net achieves 91.4% of success in fast learning mode and 95.5% of success in slow learning mode at the same vigilance level, for 32×32 pixel images. Incremental learning is achieved at a loss of 0.4% of the network performance.

Keywords: Cloud detection · Object-based image analysis · Fuzzy artmap neural network

1 Introduction

The Military Geographic Institute of Ecuador is in charge of generating the base cartography of the country. The land is mapped from aerial photographs taken from an airplane by means of a high resolution digital camera. Through the technique of orthophotography the images are corrected to produce a photographic presentation of the land without perspectives effects and other distortions. It is well known that clouds hinder the production of maps, as the clouds occlude the ground. The problem of automatic cloud detection has been extensively researched, but the existing solutions do not suite the Institute's task for two reasons:

a. The Institute's Maps are Produced from Aerial Digital RGB Photographs: Several of the existing techniques use multi-temporal information from satellite images of the same spot. The difference in brightness between two images of a region provides a starting point for identifying clouds. Additionally,

the movement of the clouds in the image sequence helps identify them from other areas of high brightness. In the case of aerial photography, the flight plan covers some overlap between pictures, but not high enough or with sufficient time span to apply the techniques of image comparison.

b. The Ecuadorian Andes have Perpetual Snow: The Continental Ecuador has a mountain range known as “*Volcanoes Avenue*”, about 300 Km long and 50 Km wide. In this “*Avenue*” there are many volcanoes and most of the highest mountains of the country. Due to the height, there are many snowy-capped peaks. Among them there is the Chimborazo (6,268 m Earth’s closest point to the sun) and the Cotopaxi (5,897 m, one of the highest active volcanoes in the world). High brightness of perpetual snow mountains in the photos difficult to distinguish clouds through the technique of brightness of clouds. Unlikely clouds, snow-covered areas are valid areas of land to be retained in cartography.

It is necessary to differentiate the clouds from the snow-capped peaks in the Institute’s photos. Currently, the process of cloud detection is performed entirely manually, which means a large investment of time.

The goal of this research is to detect the clouds by using only the RGB bands of a given photo and taking into account the possibility of the existence of snow-capped peaks in the photography. In order to guarantee this, linear transformations techniques, region growing and pattern classification through neural networks will be combined.

2 Related Work

Cloud detection has been an important issue since the beginning of remote sensing because half of the planet is covered with clouds at any given instant of time [12]. Several methods are based on the assumptions that clouds are bright and cold regions, therefore the RGB bands and infrared bands such as near-infrared (NIR), short wave infrared (SWIR) or mid-infrared (MIR) from satellites are widely used [4,6,9,13,14]. Since many images of a scene can be captured by satellite, cloud detection approaches often compare several images of the scene, at different times of the day, or over several days. For example, Champions [3] uses a pile of 6 overlapping images. Pixel-to-pixel comparison is done to detect clouds, based on the assumptions that clouds produce a significant increase of reflectance, and that the clouds should move over time (have different locations across the image time series).

Other methods use information such as the solar zenith angle or the solar irradiance. The D transform uses the the normalized difference vegetation index (NDVI) [5]. Pixels from vegetation have a positive NDVI, from water negative, and from clouds approximately zero. However, soil and snow also have NDVI close to zero as well.

Tasseled-Cap linear transformation (*TC*) is effective for detection of thin clouds and haze, and uses only information from the red (*R*) and blue (*B*) channels of the Landsat satellite:

$$TC = 0.846B - 0.464R \quad (1)$$

Variations of the tasseled-cap transformation such as the Haze Optimised Transform (*HOT*) use the image data to calculate the weights of the blue and red channels [15]. *HOT* calculates a linear regression between B (independent variable) and R (dependent variable). ϕ is the angle of the adjusted line:

$$HOT = \sin(\phi)B - \cos(\phi)R \quad (2)$$

A similar transformation is proposed by Le Hégarat-Masclé and André, [9], but using the green and the MIR channels.

These conversions from multi-channel images to a single channel image (greyscale) that can be thresholded for cloud detection is a commonly used method [11] because it is simple, fast, and cheap. Marais et al. [10] studied how to transform a four-dimensional image into a greyscale image. They computed optimal thresholds for the D transform, the *HOT* transform and the Heteroscedastic Discriminant Analysis transformation (HDA). They showed that HDA discriminates the clouds better, yet they are aware that HDA is slower than *HOT*, and that it will fail in the detection of clouds over snow or ice. It is quite clear to us that any method based on a single threshold will fail to discriminate clouds from snowy mountain peaks. Instead of using thresholds Jang et al. trained multi-layer perceptrons to classify pixels of SPOT Vegetation images [8].

Based on the work of Le Hégarat-Masclé and André, [9], Champion [3] states that clouds are connected objects and are brighter than the underlying landscape. This is a simple way of understanding the underlying process of approaches based on region growing algorithms. The brightest pixels are chosen as seeds for the algorithm. Neighboring pixels that satisfy certain criteria are aggregated with the region. All the neighbors of a pixel that belong to the region are evaluated. While Champion grows regions based on the brightness of pixels of panchromatic images, Sedano et al. implemented the procedure on the SWIR band after cloud patches were segmented by comparing cloud-free and cloudy images [14]. The cloud free images were obtained within a 17-day window from the date the cloudy image was obtained. Notice that in the multi-temporal approach most of the bright pixels that belong to the terrain will not be chosen since their brightness will not vary significantly across the multi-temporal images, and if so, they will not move across images as clouds do.

3 Input Data

The data set consists of 105 photos from 6 different regions of Ecuador: Antisana, Balzar, Chaguarpamba, Chillanes, Cotopaxi and Ilinizas. The photos are high-resolution RGB images in TIFF format. The photos were shoot from a *Cessna Citation II IGM-62* airplane equipped with an *UltraCam XP* aerial digital camera installed on a gyrostabilized mount. The mount compensates for abrupt movements of the airplane, and combined with a GPS/IMU system provides information for image correction through the process of orthophotography.

Antisana, Cotopaxi and Ilinizas are volcanoes of Ecuador. 44 of the photos of the data set include parts of the snowy peaks. As compared to related works, this

is a large dataset of high resolution images. For example, Sedano and colleagues worked on 7 satellite images [14], while Marais et al. worked on 13 images from which 32 sub-scenes were extracted [10]. Each sub-scene measured 1000×1000 pixels. Le Héhart & Mascle worked with 39 images of the African Monsoon Multidisciplinary Analysis (AMMA) dataset [9].

17 additional photos of the Guayas region were obtained at the end of this study, so they were not used in experiments I and II (Sect. 5). This is an urban region that is very different from the other six. Therefore, these photos were ideal for testing the incremental learning capability of the system.

Table 1 shows a description of the input images. The ground sample distance (GSD) is the distance between pixel centers measured on the ground. For example adjacent pixels in the images of the Antisana region are 30 cm apart on the ground. We take into account that images have different GSD, as explained in Sect. 4.

Table 1. Description of the input data

Region	Number of images	Size (pixel x pixel)	GSD (cm)
Antisana	42	11310×17310	34
Balzar	5	11310×17310	14
Chaguarpamba	14	11310×17310	30
Chillanes	8	11310×17310	34
Cotopaxi	17	7680×13824	40
Ilinizas	19	11310×17310	30
Guayas	17	11310×17310	14

4 Methods

Our method for cloud detection consists of three steps: Preprocessing, Classification and Postprocessing. In this section we describe each of these steps.

4.1 Preprocessing

The high-resolution RGB images are preprocessed to segment bright objects. Preprocessing consists of Cloud Masking, Region Growing, and GSD image normalization.

Cloud Masking: In Sect. 2 we described several filters for cloud detection. From the RGB images we only have access to the color channels, so our best options for cloud masking are the Tasseled-Cap transformation and the HOT transformation. Our goal is neither to evaluate which transform is better nor to

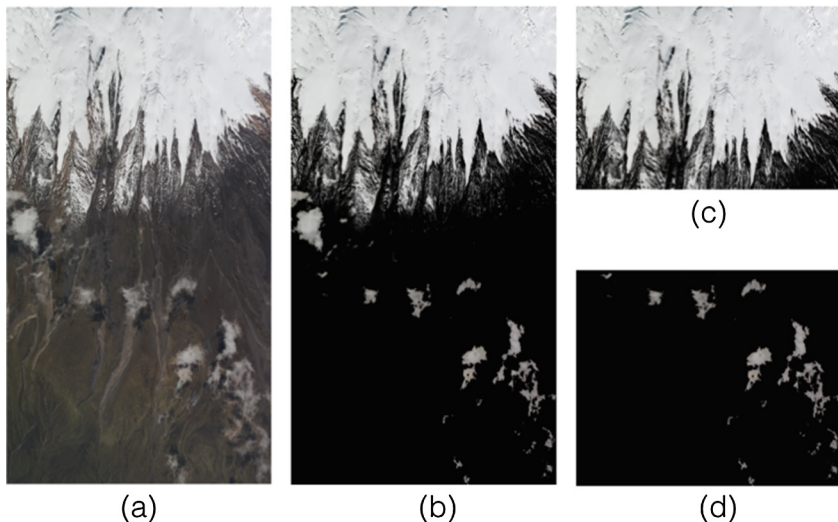


Fig. 1. Tasseled-Cap transformation of an photograph of the Cotopaxi region. (a) Original hophoto. (b) Pixels masked with the Tasseled-Cap transformation. (c) Region containing only pixels from snow. (d) Region containing only pixels of the Cloud class.

find the optimal thresholds, but to test whether a neural network can generalize what clouds are from object features.

The simplest and lowest-cost transformation is the Tasseled-Cap (TC) transformation (Eq. 1). This transformation was proposed for agricultural applications. Some vegetation can pass the filter as the green channel is not weighted in the equation. This transformation does not take into account the overall brightness of the image as the HOT transformation does, and this is the reason why we chose the TC transform. We decided to challenge the neural network approach with objects segmented by sub-optimal cloud masks. If the network can learn to recognize clouds from very variable input examples, then a better image preprocessing stage should improve the overall performance of the system. We used the following cloud mask for all images:

$$(TC > 35) \text{ AND } (G > 200) \quad (3)$$

Setting a threshold for the green channel (G) allows us to discard dark green pixels that have high TC values. The threshold of the Tasseled-Cap transform (TC) and the green channel (G) are smart guesses that we must improve in the future. Approaches for choosing these thresholds involve a process of manually labelling each pixel of the image in order to evaluate the filter's threshold accuracy. This is a time-consuming process that we cannot afford for the time being. The size of our images and data set are too large for this endeavour. For example Marais and colleagues invested two weeks labelling 32 1000×1000 pixel images [10]. This whole set is about 1/5 of the size of a single high-resolution image from our data set.

Region Growing: Experiment I (Sect. 5.1) shows that there is overlap between the values of pixels belonging to clouds and pixels belonging to snow. We could try to find an upper threshold for pixels belonging to clouds at the cost of some misclassification. However, we are interested in studying the clouds at the object level instead of at the pixel level. We believe that there are features that differentiate clouds from other objects of a photograph. To segment objects we follow the approach of region growing ([3, 9, 14]) previously discussed in Sect. 2. Land-occluding clouds are connected bright pixels. Contrary to those authors, we will encounter regions of snow in the images. Snow-covered land will grow bright regions as well. Images were compressed to 10% of their size to increase the speed of the process. We will need to compress the regions anyway later on to train the neural network, so we would rather add some speed at this point. The brightest pixel of a masked image is the seed for the region-growing algorithm. The bounding-box of the resulting region is saved for labelling the clouds during post processing. The process is repeated until all the pixels of the image are aggregated to an object.

Ground Sampling Distance (GSD) Normalization: Images in our data set have different sizes and GSD (see Sect. 3). For this reason, the sizes of the segmented objects are not normalized. We would like to feed the neural network with normalized objects, even though we are aware that there are clouds of many sizes (Table 2).

The least common multiple (lcm) of all the GSDs is 14280 cm. Since images are compressed to 10% of their size, the lcm of the compressed GSDs is 142800 cm. Hence, we created squared windows from the regions of 1428×1428 m.

Large regions are divided into several squared images of the size of the window. Tiny objects are discarded. They may belong to scattered snow, or other shiny objects. If they were indeed clouds, they will not invalidate the photo as there is some tolerance on the amount of occlusion allowed in the maps. The minimum object size was set to 2% of the window. The other regions are completed to a square image.

4.2 Classification

There are many options of neural networks that can be used in this classification task. However, we chose to use fuzzy ARTMAP neural networks because

Table 2. Size of the regions window according to the image GSD

Compressed image GSD (cm)	Window size (pixel x pixel)
140	1040×1040
300	476×476
340	420×420
400	357×357

they are capable of incremental learning and fast learning ([1,2]). Incremental learning allows the network to learn new inputs after an initial training has been completed. Many other architectures would require a complete retraining of the network in order to learn the new inputs (*i.e.* catastrophic learning). With fast learning, inputs can be encoded in a single presentation while traditional networks may require thousands of training epochs. Usually fast learning is not as accurate as slow learning, but it has a greater capability for real-time applications and exploratory experimentation.

Fuzzy ARTMAP is a biologically inspired neural network architecture that combines fuzzy logic and adaptive resonance theory (ART). The parameter ρ (*vigilance*) determines how much generalization is permitted ($0 \leq \rho \leq 1$). Small vigilance values lead to greater generalization (*i.e.* fewer recognition categories or nodes are generated) while larger values lead to more differentiation among inputs. The parameter β controls the speed of learning ($0 \leq \beta \leq 1$). At $\beta = 1$ the network is set to fast learning.

4.3 Postprocessing

Recognized clouds are marked back in the original image using the coordinates of the regions, which were saved during preprocessing. The percentage of pixels of the image that belong to clouds is computed and reported.

5 Experiments and Results

We show the results of 3 experiments conducted on the image data set. In the first experiment we compare the pixels of regions that belong to clouds with those that belong to snow. In the second experiment we train ARTMAP neural networks with the square images of the regions. In the third experiment we test the incremental learning capability of the network.

5.1 Experiment I: Image Thresholding Counterexample

Some related works focus on finding an optimal threshold for the image transformations applied to detect clouds. These works do not deal with the problem of having brighter-than-cloud pixels. We hypothesize that there is no error-free way to classify clouds at the pixel level because some snow pixels and cloud pixels may have the same brightness value.

To confirm this hypothesis we conducted the following experiment. We applied the Tasseled-Cap transformation to a photo that has well defined snow regions (Fig. 1c) and cloud regions (Fig. 1d). The threshold was set to $\tau > 35$ manually. We compared the histograms of both regions. Results show that for every single brightness level of cloud pixels there were also snow pixels with the same value. In this single image, 7,846,843 pixels of the snow class have the same brightness as clouds. Therefore, it is not possible to classify pixels by setting a brightness threshold without producing classification errors.

5.2 Experiment II: ARTMAP Neural Network for Cloud Recognition

We obtained 1302 images after preprocessing the 105 initial photos (Sect. 4.1). The images were classified manually as belonging to the *Cloud Class* or to the *Not Cloud Class*. Examples of the images are shown in Fig. 2 (upper and middle rows). There are 794 images in the *Cloud Class* and 508 images in the *Not Cloud Class*. We used a 7-fold cross-validation approach to train and test the neural networks. Thus, the data set is partitioned into 7 disjoint subsets. Each subset contains 186 images. Training and validation is performed 7 times, each time using a different subset as the validation set, and using the remaining 6 subsets as the training set. We report on the average performance of the networks over the 7 runs of the cross-validation procedure.

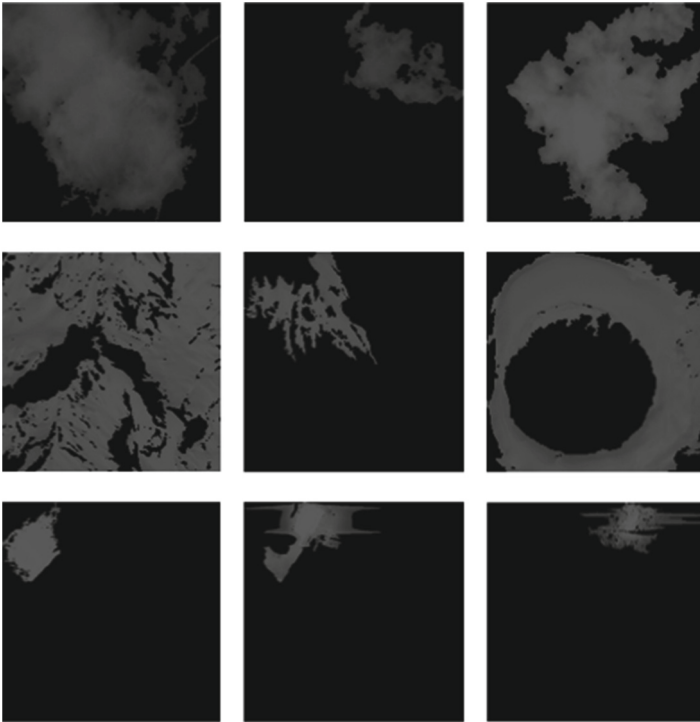


Fig. 2. Examples of high-resolution images obtained. **Upper Row:** Clouds. **Middle Row:** Snow-covered volcanoes. **Bottom Row:** Objects from the Guayas region.

We run the cross-validation procedure for 3 different input image sizes trained with fast learning at different base vigilance values. ρ was varied from 0 to 1 at increments of 0.01. Table 3 shows the minimum vigilance value and number of category nodes required to achieve a target performance.

The best results are obtained for all image sizes at success rates between 80% and 90%. Better performances are achieved at the cost of little network generalization, *i.e.* too many category nodes are created.

For the three image sizes we compare the performance of the network during fast learning and slow learning. We choose the vigilance value obtained at the 90% success value. We vary the learning rate β starting at 0.5 at increments of 0.05 and let the neural network learn during 100 epochs. Results of the best performances are presented in Table 4. In all cases we obtained a better performance at slow learning, keeping the vigilance value obtained during fast learning. The number of category nodes increased in all cases, but the number of nodes are about half of the nodes obtained during fast learning at the 95% of performance (see Table 3). For example for 32×32 images, we obtained 95.9% of success with 429 category nodes, while 992 category nodes were created during fast learning. Thus, slow learning lets the network achieve a better generalization.

Table 3. ARTMAP neural network performance in fast learning for 3 different image sizes

Target %	16×16				24×24				32×32			
	Test	Train	ρ	Nodes	Test	Train	ρ	Nodes	Test	Train	ρ	Nodes
min	75.8	73.9	0.00	20	78.0	75.8	0.00	20	67.7	75.1	0.00	18
80	80.1	78.0	0.32	83	82.2	78.4	0.29	73	83.9	77.6	0.24	66
85	85.5	85.2	0.59	188	85.5	82.9	0.52	163	85.5	84.8	0.42	124
90	91.9	90.1	0.82	418	90.1	90.1	0.79	362	91.4	89.9	0.74	317
95	-	-	-	-	95.2	96.2	0.99	980	95.6	96.7	0.99	992
Max	94.6	95.8	0.99	975	96.2	96.9	1.00	1114	96.2	97.1	1.00	1114

Table 4. ARTMAP neural network performance after slow learning

Image size	ρ	Fast Learning		Slow Learning		
		Test %	Nodes	β	Test %	Nodes
16×16	0.82	91.9	418	0.70	95.7	488
24×24	0.79	90.1	362	0.65	95.7	454
32×32	0.74	91.4	317	0.65	95.9	429

Table 5 shows the confusion matrix of the network for 32×32 images after slow learning. We note a slightly better prediction of the Cloud Class. Recall however that there are less examples in the Not Cloud Class, and that this class include a variety of objects.

Table 5. Average confusion matrix of the ARTMAP network for 32×32 input images and slow learning. Recall that there are 794 images in the Cloud Class and 508 images in the Not Cloud Class

Slow Learning 32×32		Predicted	
		Cloud	Not Cloud
Actual	Cloud	770 (97.0 %)	24 (3.0 %)
	Not Cloud	22 (4.3 %)	486 (95.7 %)

5.3 Experiment III: Incremental Learning from Urban Scenes

We used an ARTMAP neural network trained previously at slow learning ($\beta=0.65$) with 32×32 image inputs. At a vigilance level of $\rho=0.74$ the network has a success rate of 95.5% on all the 1302 images. Then, we obtained 41 objects (20 Clouds, and 21 Not Clouds) from the Guayas region photos. The not cloud objects are building roofs and image distortions not seen by the network before (see the bottom row of Fig. 2). We tested the 41 images. 80% of the new misclassified images were presented at the network for fast learning.

Table 6. Incremental Learning. The network is tested on a total of 1343 images

Guayas objects	20 Clouds	21 Not Clouds	Network Total Performance
Before fast learning	17 (85 %)	2 (1.5 %)	95.2 %
After fast learning	19 (95 %)	17 (80 %)	94.8 %

Table 6 shows the network performance for incremental learning. As expected, the network performed poorly on the not cloud objects. After 17 randomly chosen images are presented to the network for fast learning, classification of the new images improved significantly. However, we observe that new learning occurs at the cost of a small loss of the previously acquired knowledge.

6 Discussion

Results show that an object-based image analysis approach is suitable for a cloud recognition task. We have focused on exploring whether bright regions of the images encode enough information for a neural network to learn the differences between clouds and other objects of the scene. We achieved over 95% of accuracy in the classification task, which is a promising result. We must address now the issue of image preprocessing optimization and multi-resolution learning. If we can depure the inputs, learning should improve. We could evaluate other neural networks architectures, voting mechanisms, and a variety of classification techniques as well. However we would like to preserve the incremental learning capability that the ARTMAP neural network has to some extent.

Acknowledgement. This work has been financed by the PROMETEO Project from Secretaría de Educación Superior, Ciencia, Tecnología e Innovación (SENESCYT) of the Republic of Ecuador.

We would like to thank Mireya Chuga and Lenin Jaramillo for their help collecting the input data set.

References

1. Carpenter, G.A., Grossberg, S., Markuzon, N., Reynolds, J.H., Rosen, D.B.: Fuzzy ARTMAP: a neural network architecture for incremental supervised learning of analog multidimensional maps. *IEEE Trans. Neural Netw.* **3**, 698–713 (1992)
2. Carpenter, G.A.: Default ARTMAP. In: *Proceedings of the International Joint Conference on Neural Networks (IJCNN 2003)*, Portland, Oregon (2003)
3. Champion, N.: Automatic cloud detection from multi-temporal satellite images: towards the use of pléiades time series. *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*. vol. XXXIX-B3, pp. 559–564 (2012)
4. Chen, P.Y., Srinivasan, R., Fedosejevs, G., Narasimhan, B.: An automated cloud detection method for daily NOAA-14 AVHRR data for Texas, USA. *Int. J. Remote Sens.* **23**(15), 2939–2950 (2002)
5. Di Girolamo, L., Davies, R.: The image navigation cloud mask for the multi-angle imaging spectroradiometer (MISR). *J. Atmos. Ocean. Technol.* **12**, 1215–1228 (1995)
6. Gutman, G.G.: Satellite daytime image classification for global studies of earth's surface parameters from polar orbiters. *Int. J. Remote Sens.* **13**(2), 209–234 (1992)
7. Hagolle, O., Huc, M., Pascual, D.V., Dedieu, G.: A multi-temporal method for cloud detection, applied to FORMOSAT-2, VENS, LANDSAT and SENTINEL-2 images. *Remote Sens. Environ.* **114**(8), 1747–1755 (2010)
8. Jang, J.-D., Viau, A., Anctil, F., Bartholomé, E.: Neural network application for cloud detection in SPOT VEGETATION images. *Int. J. Remote Sens.* **27**, 719–736 (2006)
9. Le Hégarat-Masclé, S., André, C.: Use of markov random fields for automatic cloud/shadow detection on high resolution optical images. *ISPRS J. Photogramm. Remote Sens.* **64**(4), 351–366 (2009)
10. Marais, I.V.Z., Du Preez, J.A., Steyn, W.H.: An optimal image transform for threshold-based cloud detection using heteroscedastic discriminant analysis. *Int. J. Remote Sens.* **32**(6), 1713–1729 (2011)
11. Rossow, W.B.: Measuring cloud properties from space: a review. *J. Clim.* **2**, 201–213 (1989)
12. Saunders, R.W.: An automated scheme for the removal of cloud contamination from AVHRR radiances over western europe. *Int. J. Remote Sens.* **7**(7), 867–886 (1986)
13. Saunders, R.W., Kriebel, K.T.: Improved method for detecting clear sky and cloudy radiances from AVHRR data. *Int. J. Remote Sens.* **9**, 123–150 (1988)
14. Sedano, F., Kempeneers, P., Strobl, P., Kucera, J., Vogt, P., Seebach, L., San-Miguel-Ayanz, J.: A cloud mask methodology for high resolution remote sensing data combining information from high and medium resolution optical sensors. *ISPRS J. Photogramm. Remote Sens.* **66**(5), 588–596 (2011)
15. Zhang, Y., Guindon, B., Cihlar, J.: An image transform to characterize and compensate for spatial variations in thin cloud contamination of landsat images. *Remote Sens. Environ.* **82**(2), 173–187 (2002)