https://doi.org/10.2352/ISSN.2470-1173.2017.13.IPAS-206

Cloud and Shadow Detection using Sequential Characteristics on Multi-Spectral Satellite Images

Herman G.J. Groot[a], Arjen Oostdijk[b], Mark van Persie[b], and Peter H.N. de With[a]
[a]Eindhoven University of Technology, Dep. Electrical Eng., Video Coding and Architectures Res. Group, Eindhoven, the Netherlands
[b]NLR – Netherlands Aerospace Centre, division Aerospace Systems Space, Amsterdam, the Netherlands

Abstract

For land and sea surface monitoring applications that rely on optical Earth Observation satellite images, it is required that cloud and cloud shadow areas in the images are detected and removed. As a consequence, the frequency of obtaining new images can be increased and short-term changes can be studied more effectively. In this paper, an algorithm is designed that is able to automatically detect clouds and shadows in medium-resolution optical multi-spectral images.

The developed system is a frame-based image processing technique utilizing multiple spectral bands for feeding a cloud and shadow detector, where possible cloud contamination is recursively removed from the input images. The cloud detector utilizes Brightness Temperature Differences in the spectral regions of Far IR (FIR) and Thermal IR (TIR). After careful considerations, the reflective band FIR was adopted for usage in this Difference Image. The shadow detector uses Background Subtraction, which iteratively constructs its Reference Image automatically. This iterative nature is exploited to utilize time-sequential characteristics among the input images.

After experiments, 94.6% of the clouds are detected, with a precision of 86.5%, as determined using per-pixel ground-truth data. For shadows, these statistics are 77.1% and 75.8%, respectively and may be further improved in future work. Selected mid-resolution Landsat images have been used for the validation.

Index Terms— Cloud Detection, Shadow Detection, Multi-Spectral Satellite Images, Time-Sequential Characteristics, Ground-Truth Based Validation

I. Introduction

For most studies that rely on optical Earth Observation satellite images, it is required that the images do not contain clouds or cloud shadows. For example, in applications like weather analysis, weather forecast, and land- and sea-surface temperature analysis, the extraction of cloud locations plays a crucial role [1].

Since most satellite images contain cloud contamination, cloud and shadow removal is useful. This paper describes a study on cloud and shadow removal using multi-spectral images. A well-designed algorithm will take all possible cloud types into account, since specific cloud configurations like e.g. cirrus clouds, are known to be difficult to identify [1][2][3].

The proposed approach utilizes multiple spectral bands, but limits the exploited bands to those that are typically present in most medium-resolution multi-spectral Earth Observation Satellites, like Landsat and ASTER. For coarser resolution multi-spectral satellites, like GOES, AVHRR and MODIS, the designed algorithm is expected to be compatible. The algorithm uses medium-resolution Landsat imagery, obtained from either the Thematic Mapper Satellite (L5) or newer.

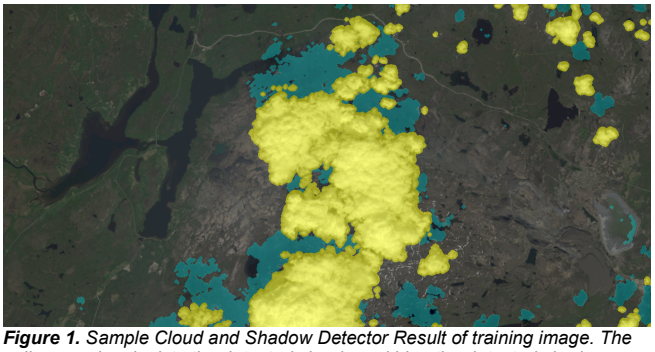

Figure 1. *Sample Cloud and Shadow Detector Result of training image. The yellow overlay depicts the detected clouds and blue the detected shadows.*

In previous work [1] and [2], it was found that Brightness Temperature (BT) difference images are used, which is attractive due to the high contrast between cloud and ground. However, the algorithm from [1] adapts some of its thresholds using a 20-day history of images, which makes it more flexible to changes, while [2] is based on fixed thresholds. Approaches that do not employ thresholds on more or less raw image data have been encountered in [4] and [5], which use feature extractors and clustering techniques, respectively. As such techniques seem plausible, they were adopted into the proposed approach.

The sequel of this paper first describes the detector in Section II. Then, Section III discusses the processing structure that utilizes sequential characteristics. The experimental results are given in Section IV. The recommendations and conclusion are presented in Sections V and VI, respectively.

II. Cloud and Shadow Mask Generation

The proposed approach is shown in Figure 2, which consists of blocks *Cloud and Shadow Mask Generation*, *Previous Result Insertion*, *Storage* and *Visualization Post-Processing*. The first block is the most important function, generating a cloud- and shadow mask. Both detectors are addressed in detail in this section.

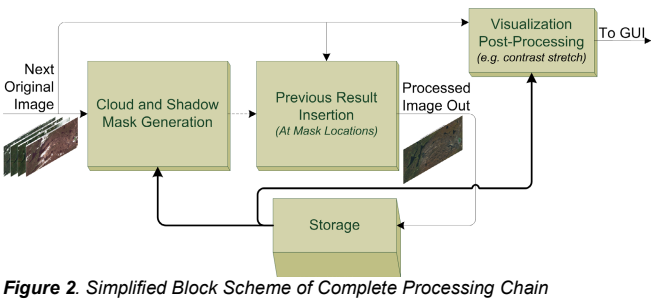

Figure 2*. Simplified Block Scheme of Complete Processing Chain*

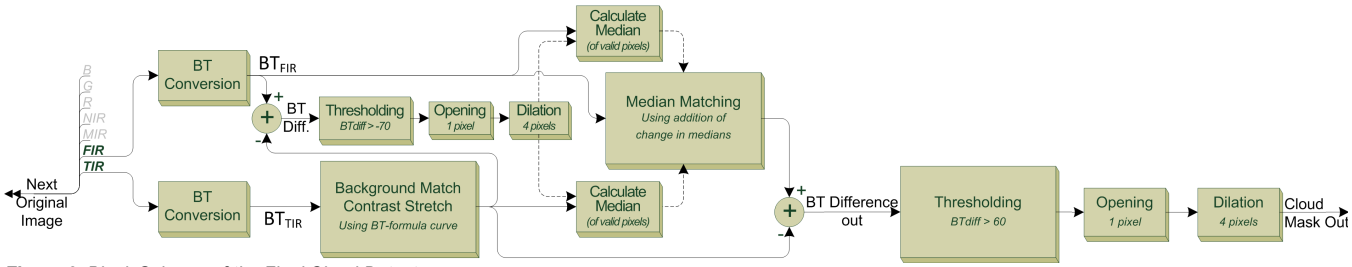

Figure 3. *Block Scheme of the Final Cloud Detector.*

A) Cloud Detector

The block scheme of the cloud detector is shown in Figure 3, where the input for the detector is a series of Landsat radiance images. A Brightness Temperature (BT) Differencing approach is adopted which is motivated below, where the detector first converts two spectral bands of the input to BTs. Next, one of the channels is stretched in contrast, to increase the contrast between cloud and background for the thresholding steps that follow. To make the threshold more adaptive, a median matching step is included, which matches the median of the background parts of the BT images that should be subtracted.

To structure the upcoming discussion of the system, a list of topics is defined. In order to quickly refer to these topics, a label has been assigned to each of them. This list involves motivating the use of BTs (T1), the complications using Landsat Images (T2), and is followed by an extra Contrast Stretch (T3). Then further optimizations follow for this extra Contrast Stretch (T4), succeeded by Median Matching optimizations (T5). Finally, the morphological operations are discussed (T6) and some conclusions are given (T7).

(T1) Motivation of using Brightness Temperature

In literature, multiple cases are found that applied Brightness Temperature (BT) differences between spectra around 3.7 μm and 11 μm [1][2][6][7] (where [2] referenced to [6] and [7] for its use of this BT difference method, but for [6] the use of this method could not be checked). Both of these spectral channels are part of the emissive bands [8]. However, the 3.7-μm band is so close to the borderline between emissive and reflective channels that reflection of sunlight still plays the most important role in this channel [8]. On the other hand, at the 11-μm band, the influence of sunlight reflection is negligible because it involves the larger wavelength part of the IR spectrum and is thus almost a fully emissive channel [8]. These properties are used in the BT difference. Since clouds are colder than the background, their appearance is dark in the fully emissive 11-μm band, while clouds appear bright in the 3.7-μm band, due to the reflected sunlight [1][2][8]. As a result, the BT difference has a much higher contrast between clouds and background.

Another advantage of using these spectra is that snow will have little or no influence on the detection. This is because, in [3], snow has a strong absorption ability for wavelengths around 1.64 μm. For the considered Landsat imagery, similar behavior was encountered for the reflective bands that have longer wavelengths. Also in the emissive band, snow appeared as a visual signal that mixes into the background.

(T2) Complications using Landsat Images

Adopting this BT approach on the Landsat Imagery yields complications in the available spectra and their BT conversions. Landsat does not use Lookup Tables, like MODIS does, to come to Brightness Temperatures for their emissive bands, but calculates them using a formula that stretches image contrast, according to:

$$BT = \frac{K_2}{ln\left(\frac{K_1}{L_\lambda} + 1\right)} \quad . \tag{1}$$

Here, constants K_1 and K_2 are provided in the image meta data and L_λ is the spectral radiance of the corresponding emissive band [9].

Unfortunately, none of the Landsat satellites measure spectral data around 3.7 μm, which is the source of the complications when implementing the BT difference method as applied in [1][2][6][7]. The only emissive channel that the Landsat missions offer, is the Thermal IR (TIR) channel. This channel is sensitive for light with wavelengths between 10.4 and 12.5 μm (for L5). Because the constants of Eq. (1) are only available for emissive channels, only the TIR channel has a method available to come to BTs. The channel of the Landsat satellites closest to the other required spectra of around 3.7 μm is the Far-IR channel (FIR, also known as SWIR), which measures in the spectral range of $2.08 - 2.35\ \mu m$ (for L5). Newer than L5 satellites measure in similar spectral ranges.

Applying Eq. (1) with the metadata-provided constants K_1 and K_2 ensures that units of degrees Kelvin are achieved. Since it is not typical to acquire Brightness Temperatures for reflective bands, it will be difficult for the FIR channel to come to the exact constants for BT conversion. However, the application for this study is cloud and shadow detection, so that coming to BTs with exact units of degrees Kelvin is of little importance, as long as a similar contrast stretch is applied – both *BT conversion* blocks of Figure 3 are therefore identical – and the reason for using BT differences in this application and its channel selection is sound, this is treated next.

If a typical spectral characteristics plot is considered that shows the ratio between Earth-emissive and Sun-reflective energy of the detected radiance in orbit, as is present in [8], it can be seen that using spectra around 2.2 μm rather than 3.7 μm, has no significant influence: in both cases the reflective energy is the major part of the in-orbit received radiance. Both FIR and TIR channels will thus respond similarly to the previously proposed spectra (by [1][2][6][7]).

(T3) Extra Contrast Stretching

In the ideal case, it would be desirable that the cloud-free areas (the backgrounds) in both individual images of the BT difference are as similar as possible to make the clouds stand out more. This can be achieved by an extra contrast stretch. The result is that the backgrounds have a BT difference close to zero, while the clouds are then the only areas that significantly differ and consequently are better to differentiate from the background. Because of practical reasons, the extra stretch is only applied to the TIR channel, as is done by the processing block *Background*

Match Contrast Stretch in Figure 3, with fixed constants K_1 and K_2.

(T4) Further Optimizations for the extra Contrast Stretch

To optimize the extra contrast stretch, BT_{TIR} (see Figure 3) is first subtracted by a scalar, followed by removing the then negative values. This ensures that clouds are differently stretched than the backgrounds, because otherwise values are inserted in the roughly linear range of the curve of Eq. (1). The scalar that is subtracted is selected as the 0.5th percentile of the pixel values present. This ensures that always a constant area of the clouds is saturated which results in more constant behavior and makes it inherently adaptive to radiance variations in the actually considered image. The scalar to subtract is not allowed to be higher than $273.15\,°K$, as not all images contain clouds. Lastly, the performance curve of the extra contrast stretch is selected such that it is similar to the curves of the processing blocks *BT Conversion* of Figure 3, which proved to be effective.

(T5) Median Matching Optimizations

The next step of the cloud detector is *Median Matching*, which makes the system adaptive to a broad range of situations, as demonstrated in Figure 4. This step adds the difference between the medians of the two input spectra (measured at blocks *Calculate Median*) to one of the spectra, such that these inputs equalize and lead to the same median outputs. It results in a more robust behavior of the cloud detection, thereby effectively making the threshold dependent on the currently processed image, and hence making it adaptive. Having the same median outputs ensures that the background pixels will correspond to a small BT difference value. However, if there are more cloud-contaminated pixels than there are background pixels, the median will correspond to the clouds. To solve this, the detector is first executed partially with a different threshold, to provide an estimate that excludes most of the cloud pixels for the *Median Matching*. The detector therefor has two instantiations of the block series *BT Difference*, *Thresholding*, *Opening* and *Dilation*.

(T6) Morphological Operations

As a final step of the cloud detector, morphological *Opening* is performed as most of the smallest areas returned by both blocks *Thresholding* are typically errors. Likewise, *dilation* is also applied, since the detector appeared to have difficulties on cloud edges. Lastly, the structuring element used has a circular shape, since clouds typically have an approximately round shape.

(T7) Conclusions on the Cloud Detector

Ultimately, a detector is realized that is able to differentiate cirrus clouds which are hard to detect. Although cirrus clouds have a more grayish color, they are still detectable in most situations, due to the enhanced contrast. For the same reason, the scheme also

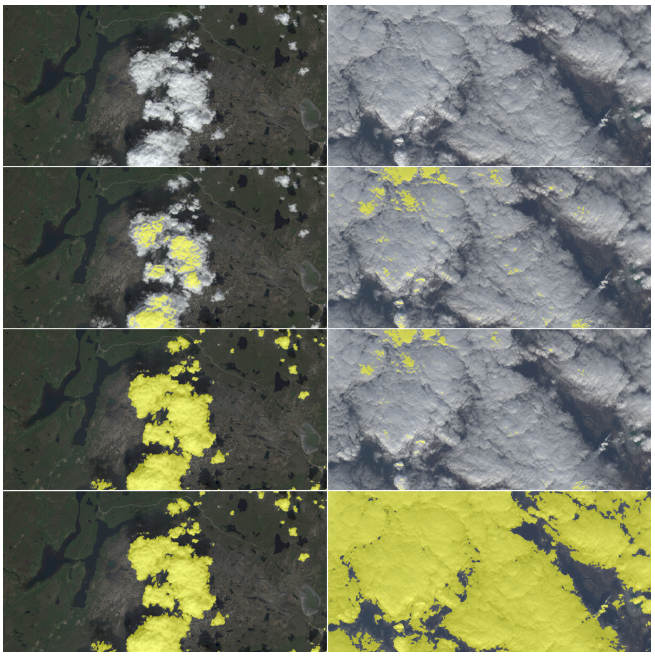

Figure 4. *Median Matching (MM) effectiveness for two images with different cloud coverage at the same location (left/right). From top to bottom: Original RGB image; cloud detection mask overlay without MM; ditto with MM where median is calculated over all pixels; and ditto where MM is calculated over valid pixels only. The detection threshold is constant for all these situations.*

works well for the thicker white clouds. However, it was found that once the cirrus type of clouds became more transparent and fog like, the method has more difficulties in detecting this type of clouds. It should be noticed that it depends on the application whether this is acceptable or not. If the application is for instance visible inspection, it may still be possible to distinguish objects from the background, like cars, planes, etc. In such a case, the detection difficulties will then be of less importance.

B) Shadow Detector

The shadow detector, partly inspired by [1], is depicted in Figure 5. For now, the BS image is considered to be constructed by subtracting an image which is known to be cloud-free. As such, clouds yield positive differences and shadows yield negative differences. This principle is exploited in the shadow detector.

The final shadow detector utilizes the two channels of all spectra that are found to be most effective. The selected channels are the Near IR (NIR) and the Mid IR (MIR) bands, which are combined as shown in Figure 5 to reduce errors. The combination can reduce errors, since both spectra mostly produce errors at different locations, as shown in Figure 6. Typically, the NIR channel generates errors at vegetation (NIR is used in the Normalized Difference Vegetation Index). For the MIR channel,

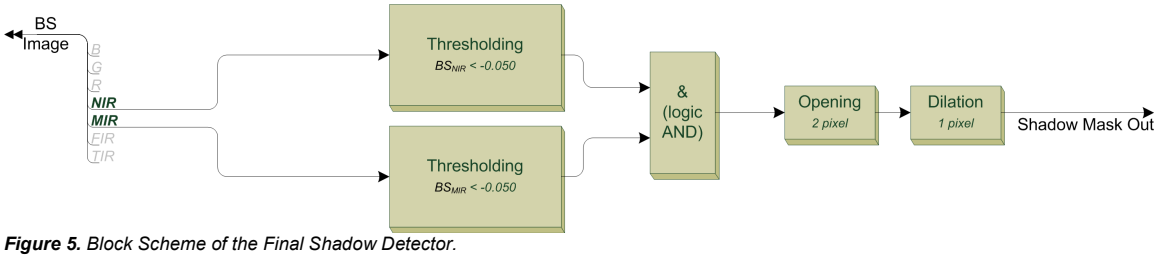

Figure 5. *Block Scheme of the Final Shadow Detector.*

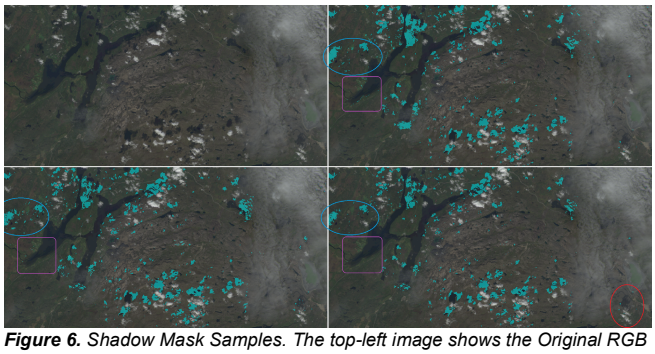

Figure 6. *Shadow Mask Samples. The top-left image shows the Original RGB image, the others show it overlaid with a shadow mask. Top-right: thresholding the NIR spectrum of the Background Subtracted image, bottom-left: ditto for MIR spectrum, and bottom-right: logical AND of the previous two. The circles indicate areas that show interesting trends.*

the errors are typically apparent at different locations than those present in the NIR result.

Unfortunately, it is unavoidable that errors remain in the combination of the two spectra, as indicated by the circles in Figure 6. This occurs mostly because the errors have similar intensities as true shadows. The blue circles demonstrate the case of effective error removal from MIR and the purple colors likewise for NIR. In the middle of these images, land can be seen that is ultimately labeled erroneously as shadows. Finally, the red circles show an undetected shadow, due to the very transparent cirrus clouds on top of it.

Additionally, water area is also difficult, likely because these areas have already low reflection values. The detector is also dependent on the sun angle, which influences the natural position of shadows projected by the ground and other objects. Its effect on the detection depends on the time of day and season.

However, difficulties for the shadow detector could be expected, since shadows are harder to detect. For instance, they are visually very similar to water lakes. Due to the here described issues and research time limitations, the resulting performance is limited, as will become clear in Section IV.

III. Sequential Characteristics Pipeline

The objective in this study is to develop a cloud- and shadow detection algorithm for time series of images, performed in the framework of monitoring applications, such that clouds and shadows are time-sequentially filtered out. The advantage of such a sequential approach is that the Background Image can be constructed iteratively, since it is not always feasible that there is an image present that contains no clouds whatsoever. Even if there are clouds, it is preferable that this image is taken recently due to landscape differences. The block scheme of the proposed approach is shown in Figure 7.

A) Preparation and Pre-Processing

Each iteration of the pipeline starts with loading the actual original image and calculating the Background Subtracted (BS) image, where both images are already atmospherically corrected. That is, the channels R, G, B, NIR, and MIR are already converted to reflectances, and the FIR and TIR channels are already converted to Brightness Temperatures.

B) Detection and Insertion

Once these images are available, the *Cloud and Shadow Detector* (Section II) is executed. The detector returns three masks: the cloud mask, the shadow mask, and the union of both. These are saved in *Saved Masks* and the combined mask is directly used by *Insert Previous Result*. For ideal masks, the latter produces a cloud-free processed image. It accomplishes this by inserting the previous iteration's output at the locations that the current combined mask labeled as cloud-contaminated pixels. The inserted image is corrected in brightness, to match the current *Original Image* as closely as possible and to prevent insertion borders, which will be further described (in Subsection C). The resulting image is then stored at *Saved Processed Images*.

C) Brightness Correction

To decrease the apparent insertion borders, the brightness correction is applied, as proposed in [5], using:

$$I_{out} = \left(I_{OI} - \mu_{I_{OI}}\right) \cdot \frac{\sigma_{I_T}}{\sigma_{I_{OI}}} + \mu_{I_T} \ ,$$

where I_{out} is the image of interest (I_{OI}), which is corrected in brightness towards the target image (I_T). Parameters $\mu_{I_{OI}}$ and μ_{I_T} are the mean of I_{OI} and I_T, respectively, while similarly, $\sigma_{I_{OI}}$ and σ_{I_T} describe the standard deviation of these images. In this case, I_{OI} denotes the *previous Processed Image*, and I_T is the current *Original Image*. The cloud-contaminated pixels are excluded from these statistics to improve accuracy.

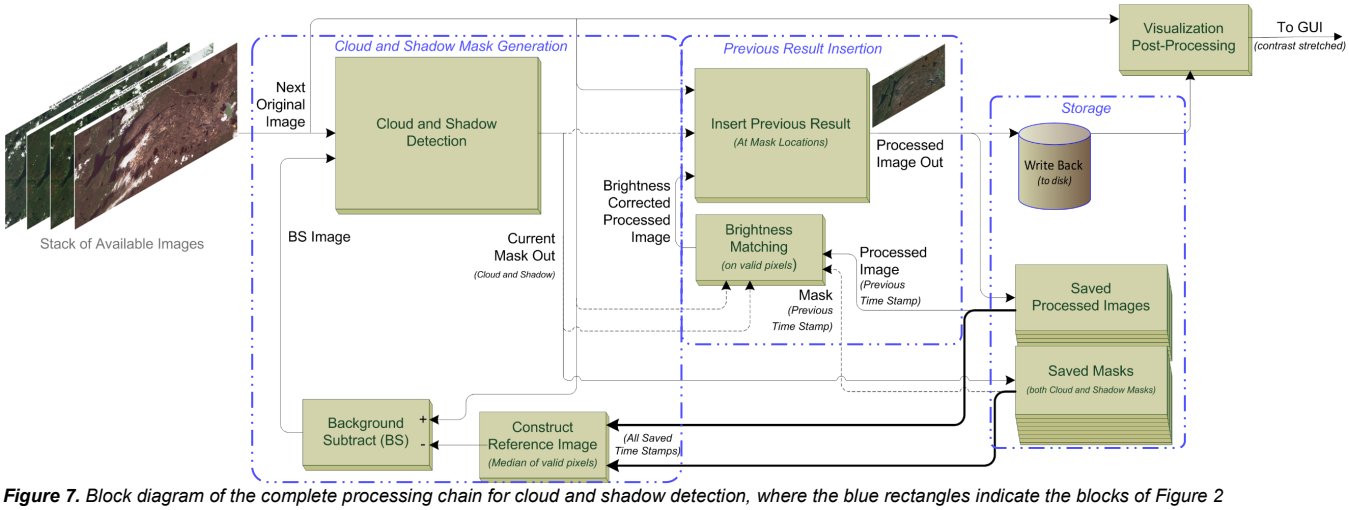

Figure 7. *Block diagram of the complete processing chain for cloud and shadow detection, where the blue rectangles indicate the blocks of Figure 2*

D) Reference Image

Blocks *Saved Processed Images* and *Saved Masks* currently save up to a total of 5 images for easy access and according to the FIFO principle. *Construct Reference Image* then accesses these images and masks to compute a median of the pixels having the same image-coordinates (loosely based on [1]). If one of these pixels (with the same coordinates) do not describe a cloud or a shadow, they are labeled valid and then the median of up to five values are computed, based on the label value. The non-valid pixels are excluded, to assure that none of the *Original Image* pixels are counted multiple times due to the insertion of the *previous Processed Image*.

The number of saved images is selected as 5, such that the complete system becomes more robust against temporal effects and because of the limited number of images in the validation and training sets. After completion, the full processing chain is repeated on image basis and the next iteration begins.

It is important to note that the background subtraction makes the shadow detector dependent on the cloud detector. Also, it is troublesome to encounter snow on the input images. When such a case first occurs, the detector cannot directly switch to the new situation and makes shadow errors, until it can change its reference image. This adaptation time could not be further addressed and as such it was decided to focus on non-snow imagery, for which a solution is presented in Section V.

IV. Results and Validation

This section focuses on the separation of datasets, the validation of the proposed scheme and the experimental results.

A) Optimizing the Method and Image Selection

A number of images have been made available by the NLR institute. These images have been divided into a set strictly used for training purposes, and a set reserved for validations. Since generating ground-truth data is a process that is rather time consuming, only for the validation set ground truths were made. Therefore, the training and the setting of involved parameters was performed empirically by manual inspection. For the validation set, the images were selected as consistently as possible. To do so, all Landsat imagery of 2015 that was available for the selected region was downloaded. The only imposed selection was the exclusion of images containing snow on approximately more than 10% of the images, due to the previously described limitations of the shadow detector.

The images from the training set were selected more specifically by the experts of the NLR institute. To be able to reuse these images, they were selected as the training set. The selection was done such that images are covering the same area approximately once every month in 2013 and 2014, to investigate short-term changes. Experts searched for the absence of clouds and snow as much as possible, which typically occurs more often in the summer. To compensate, one of the 2015 images that mostly consisted of clouds was interchanged with the training set for one of the 2014 images. At last, both sets now contain 8 images.

B) Ground-Truth Generation

The ground truths were generated by a remote-sensing expert of NLR, having extensive experience in image calibration, atmospheric correction and pre-processing. Typically, ground truths were annotated manually by experts and at pixel accuracy. However, in the case of clouds, the annotation is a painstaking drudgery work, due to the many small clouds occurring in certain situations. As a result, the ground truths were made by manually setting the optimal threshold per image for some spectral classification. In some cases, this required a threshold that includes also some false positives in order to involve as many clouds as there are present. Next, these false positives have been manually removed.

The manual annotations made by the expert proved to be partly based on the cloud detector, because it provided an effective classification compared to other spectral classification. Also, the shadow ground-truth was limited due to lack of images, using only known cloud-free images. Summarizing, the above selections yield some limitations to the generated ground truths. The missed clouds of the proposed algorithm are likely also missed in the ground-truth annotations. However, it is also possible that the ground truth lacks clouds that the proposed algorithm could rightfully detect, due to the induced enhanced contrast. Both statements also hold for the shadows. Besides this, also manual annotation uncertainties occur. It is generally known that if an image is annotated by multiple experts, their resulting ground truths are not fully identical. In order to address this issue, a multi-level ground truth would be required that takes these undefined borders into account [10]. As there is only one ground truth per image, these regions are finally generated by morphological dilation such that the undefined borders are included in the dilated ground truths.

C) Validation Methodology

Figure 8 visualizes a selection of the algorithm's mask overlay results. The ground-truth data are used to classify the algorithm masks into True- and False Positive regions (TP and FP), and True- and False Negative regions (TN and FN), excluding the 'don't care' regions. These regions are then used to compute:

$$R = \frac{|M_{GT} \cap M_D|}{|M_{GT}|} = \frac{|TP|}{|TP| + |FN|} \quad ,$$

$$P = \frac{|M_{GT} \cap M_D|}{|M_D|} = \frac{|TP|}{|TP| + |FP|} \quad ,$$

$$J = \frac{|M_{GT} \cap M_D|}{|M_{GT} \cup M_D|} = \frac{|TP|}{|TP| + |FP| + |FN|} \quad ,$$

with Recall score R, Precision score P, and Jaccard-Index score J. Parameter M_D denotes the Detector Mask, and M_{GT} is the Ground-Truth Mask.

The R, P, and J scores are three useful metrics, describing the detector performance, which is computed for every image in the validation set and then averaged. Recall shows what percentage of the ground truth is detected, while Precision describes how much of the total mask is correctly identified. The Jaccard-Index combines both types of possible mistakes and determines a score based on both. The Jaccard-Index equally weights both types of mistakes in its computation.

Figure 9 presents a full overview of the performances of the cloud, shadow, and overall detector using the validation set.

D) Validation Results

From the samples of Figure 8, it becomes clear that the cloud detector performs well, nearly all clouds are detected, only the more transparent clouds are missed. Typically, these cirrus clouds are most difficult to detect, and the results indicate that most of the cirrus can be detected. The shadow detector finds the majority of the shadows, but unfortunately also generates false positives in the process.

It stands out that the three images containing nearly no cirrus clouds, particularly image 4 and 5, score poorly on precision for cloud detection. The loosely-tuned dilation operation thus appears to form a compromise in the precision for these cloud types.

With respect to image 8, it should be noticed that although the cloud detector's precision is high, the recall is rather low. This image contains a difficult kind of cirrus cloud over the whole image, which results in a ground truth that is virtually impossible to tune with the described method of generation. For the shadow detector, the high recall is not surprising in this case, since the shadow mask includes a large portion of the image. It does include the darkest areas of the image, but it is hard to visually identify that these areas are indeed shadows.

Generally, the shadow detector performs reasonably good, but not as good as the cloud detector. This result was expected since its development was time constrained for project reasons. Most of the found errors are the result of faults in the reference image, indirectly caused by false negatives from the cloud detector.

Figure 9 presents a full overview of the precision, recall, Jaccard-Index performances of the cloud, shadow, and overall detector using the validation set. The red bars denote the performance of both cloud and shadow detectors in combined form as a union operation. This effectively looks at cloud-contamination in general. Its overall score is higher than the average of both detectors, which can be explained by the pixels that are labeled erroneously at both detectors: if both detector masks are combined, these pixels count as a single error instead of double errors.

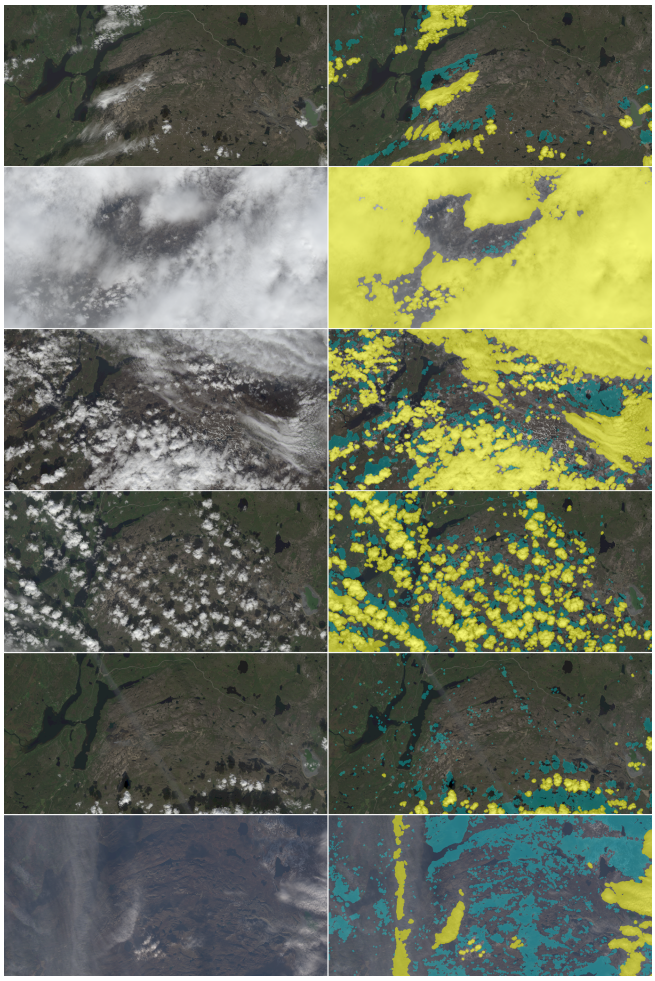

Figure 8. *Selection of validation results, shown as mask overlays where yellow depicts clouds and blue depicts shadows. From top to bottom, these images correspond respectively to images 1, 2, 3, 5, 7, and 8 in Figure 9.*

E) Proposed approach compared to previous work

From all encountered previous work as cited in this paper, none of the authors validated their work with pixel-based ground-truth data, where refs [1] and [2] performed a better validation than others. Furthermore, none of the validations of the related work involved recall and precision, their statistics have thus been re-interpreted subjectively. At last, most of the related work do not include a shadow detector. Therefore, in an attempt to compare the proposed approach with previous work as reliably as possible, their statistics have to be reinterpreted to estimate the precision and recall, as far as this is possible. The result of this comparison is shown in Table 1. Due to these issues concerning incompleteness of the experiments in literature, the table's legend describes how all scores are obtained.

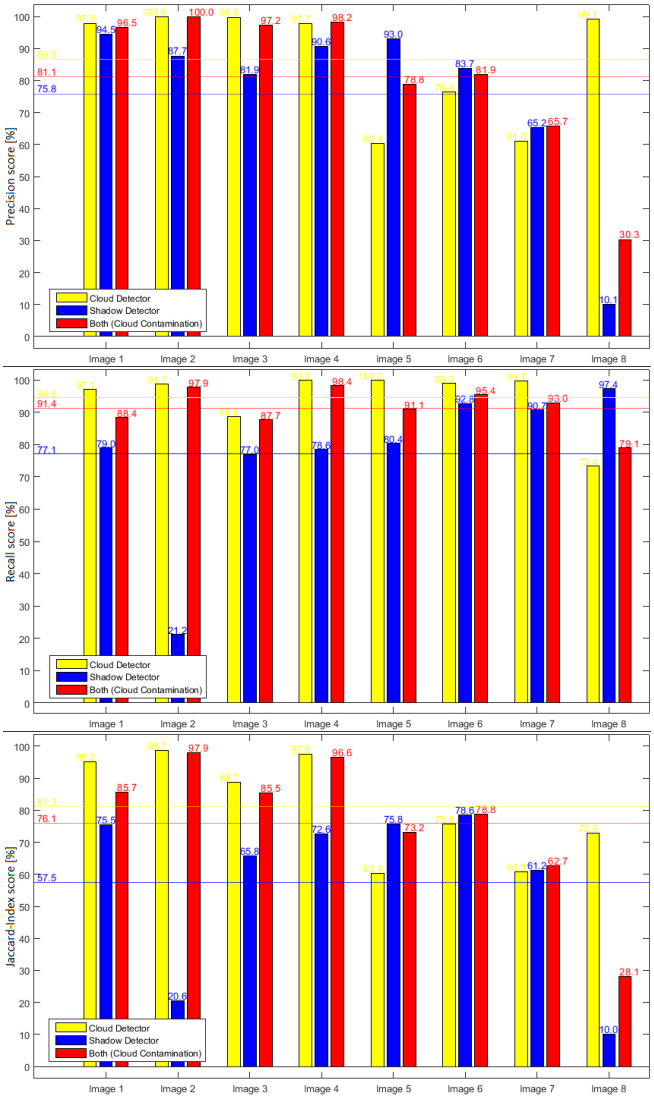

Figure 9. *Precision, Recall, and Jaccard-Index score, respectively, for all validation images. The horizontal lines correspond to the mean value of all measurements for the considered color and score.*

Table 1: Cloud Detector (CD) or Shadow Detector (SD) Validation scores of proposed and related approaches. These scores are averaged over all performed measurements and the legend describes further how the scores are obtained

	Approach (values in %)				
	Proposed	Jedlovec et al. [1]	Chen et al. [2]	Tong et al. [4]	Surya et al. [5]
CD Precision	86.5	93.5[1]	95.8[2]	-[3]	-[4]
CD Recall	94.6	84.5[1]	86.9[2]	68.0[3]	86.7[4]
CD Jacc.-Ind.	81.3	79.8[1]	83.6[2]	-[3]	-[4]
SD Precision	75.8	-[1]	-[2]	-[3]	-[4]
SD Recall	77.1	-[1]	47.4[2]	68.0[3]	-[4]
SD Jacc.-Ind.	57.5	-[1]	-[2]	-[3]	-[4]

Legend:

[1] approximated using the given numbers of the used validation technique. This technique involves defining 16675 blocks of 30x30 km, for 30 unique locations. For each of these blocks only 3 labels were given by experts: clear, partly cloudy, and cloudy. Shadow detector not part of research.

[2] derived from the extractable TP, FP and FN numbers, which was based on pixel sets from 24 images strictly used for validation. Thus, not all pixels from these images were processed. No expert mentioned for annotations. Note that cloud detector results are forced to include shadow detector results with the given statistics.

[3] probably recall is given, distinction between recall and precision is not clearly stated. Score based on counting the number of clouds detected for only five images. No expert mentioned for annotations. Only recall could be obtained.

[4] based on counting the clouds in only 3 images. No expert mentioned for annotations. Only recall could be obtained, shadows not part of research.

In conclusion, Table 1 in combination with the coarse validation methods described in its legend, shows that the proposed approach outperforms all related work found in the literature research, as referenced in this document. This can be concluded because none of the related work validate their work on a true per-pixel basis, and if done per-pixel, the regions used for the validation is based on some (subjective) selection.

V. Recommendations

As already mentioned, the performance evaluation would benefit from experiments with a larger dataset. It is essential that the issues regarding annotation and free selection of cloud types are resolved prior to performing such larger scale experiments.

Furthermore, keeping a history for snow and for non-snow images instead of a single global history of images, will also benefit the detection performance. The NIR, MIR and R channels are expected to effectively identify if a pixel describes snow-covered-landscape and can be used to put an image in one of these histories. In addition, further detailed examination of frequency characteristics can be used to identify the fog areas that the cloud detector currently misses.

Also, the Cloud Mask could be used to find shadows in the Shadow Mask with a matching shape of the detected clouds, similarly as done in [11]. Since clouds are easier to detect, this may improve the shadow detector results. Due to known sun angles, only the cloud height determines the locations of the corresponding shadows with respect to the cloud mask, which limits the search space.

VI. Conclusions

Inspired by previous work, an algorithm is designed that is able to detect clouds and shadows in optical Earth Observation satellite images. After careful considerations, the cloud detector is based on an adaptive scheme that applies Brightness Temperature Differences on an emissive and a reflective spectral band and further enhances contrast in this difference image. As a result, this cloud detector outperforms existing proposals and achieves an average recall of 94.6% as measured on an independent validation set of 8 images, with an average precision of 86.5% and an average Jaccard-Index of 81.3%.

Based on a set of previous results that replaced the cloud-contaminated areas with cloud-free data, the shadow detector iteratively constructs a reference image and subtracts this from the original image. The shadow detector then thresholds two of the spectral bands of this subtraction result. Finally, the shadow detector achieves an average recall, precision and Jaccard-Index of 77.1%, 75.8%, and 57.5%, respectively, measured on the same validation set.

For both cloud and shadow detection experiments, the dataset was quite limited due to specific selection of images and manual annotation issues. However, the proposed approach presents a better way of validation in the form of using annotated ground-truth data of an expert. From all encountered previous work, as cited, none of the authors validated their work with pixel-based ground-truth data and used subjective interpretations, rather than objective precision and recall measurements with expert annotations.

In comparison with related work, it has been shown that the introduced algorithm outperforms all related work found in the literature research, due to the incomplete validation techniques employed. This conclusion applies to both the cloud and the shadow detector. The proposed approach has been designed such that it is adaptive as much as possible and is also expected to be robust when snow is encountered, due to the selected IR spectral channels. Finally, there are various options described to improve the proposed approach, e.g. by making the reference image construction and/or the contrast stretch more advanced.

VII. Acknowledgement

The authors would like to thank A. Oostdijk and M. v. Persie, of Netherlands Aerospace Centre (NLR) for hosting the research work. The author would also like to thank H. Noorbergen of NLR for his semi-manual assessment of cloud cover and his help in performing atmospheric corrections.

VIII. References

[1] G.J. Jedlovec, S.L. Haines, and F.J. LaFontaine. "Spatial and Temporal Varying Thresholds for Cloud Detection in GOES Imagery", in IEEE Transactions on Geoscience and Remote Sensing, pp 1705-1717, June 2008.

[2] P. Y. Chen, R. Srinivasan, G. Fedosejevs, and B. Narasimhan, "An automated cloud detection method for daily NOAA-14 AVHRR data for Texas, USA", in International Journal of Remote Sensing, pp 2939-2950, 2002.

[3] T. Cheng, X. Gu, L. Chen, T. Yu, and G. Tian, "Cloud detection based on the spectral, multi-angular, and polarized characteristics of cloud", in IEEE International Geoscience and Remote Sensing Symposium (IGARSS), pp 3321-3324, July 2007.

[4] Y. Tong, and W. Ming-shu, "Clouds and Shadows Detection in Multi-Spectral Satellite Image Based on Maximally Stable Extremal Regions", in International Conference on Multimedia and Signal Processing (CMSP), pp 184-187, May 2011.

[5] S.R. Surya, and P. Simon, "Automatic Cloud Detection Using Spectral Rationing and Fuzzy Clustering", in Second International Conference on Advanced Computing, Networking and Security (ADCONS), pp 90-95, December 2013.

[6] T. Yamanouchi, and S. Kawaguchi, "Cloud distribution in the Antarctic from AVHRR data and radiation measurements at the surface", in International Journal of Remote Sensing, pp 111–127, January 1992.

[7] K.D. Hutchison, B.J. Etherton, and P.C. Topping, "Cloud top phase determination from the fusion of signatures in daytime AVHRR imagery and HIRS data", in International Journal of Remote Sensing, pp 3245–3262, January 1997.

[8] P. Antonelli, Lecture: "Quick Review of Remote Sensing Basic Theory", University of Wisconsin-Madison, June 2007, available from: https://cimss.ssec.wisc.edu/rss/benevento/source/lecture3a_7jun_emissive_and_reflective_antonelli.ppt

[9] Official description page of applying atmospheric corrections on Landsat 8 data. Available from: https://landsat.usgs.gov/using-usgs-landsat-8-product.

[10] F. van der Sommen, S. Zinger, E.J. Schoon, and P.H. de With, "Sweet-spot training for early esophageal cancer detection", in Medical Imaging 2016: Computer-Aided Diagnosis, pp. 97851B-1 to 7, March 2016.

[11] R. Ren, L. Gu, and H. Wang, "Clouds and Clouds Shadows Detection and Matching in MODIS Multispectral Satellite Images", in International Conference on Industrial Control and Electronics Engineering (ICICEE), pp 71-74, August 2012.

Author Biography

Herman Groot is starting a PhD study at the Eindhoven University of Technology, Eindhoven, The Netherlands, at the faculty of Electrical Engineering. He is specializing himself in various forms and signal processing techniques for Video and Image Processing with focus on the Space section of Aerospace Engineering. At present, he is working on Image Reconstruction and Restoration processing and 3-D Reconstruction of video scenes.

Mark van Persie is expert in Remote Sensing and Arjen Oostdijk is Systems Engineer, both at Netherlands Aerospace Centre (NLR), division Aerospace Systems, Amsterdam, the Netherlands. They are involved in the interpretation of Remote Sensing images for various application areas.

Peter H. N. deWith (MSc. EE) received his PhD degree from University of Technology Delft, The Netherlands. After positions at Philips Research, University Mannheim, LogicaCMG and CycloMedia, he became full professor at Eindhoven University of Technology. He is an (inter-) national expert in surveillance for safety/security and was involved in multiple EU projects on video surveillance analysis with the Harbor of Rotterdam, Dutch Defense, Bosch Security, TKH-Security, ViNotion, etc. He is board member of DITSS and R&D advisor to multiple companies. He is IEEE Fellow, has (co-)authored over 300 papers on video analysis, systems and architectures, with multiple awards of the IEEE, VCIP, and EURASIP.