Development of a cloud, snow and cloud shadow mask for VEGETATION imagery

Gil Lissens, Pieter Kempeneers, and Freddy Fierens

Flemish Institute for Technological Research (VITO)

Centre of Expertise Remote Sensing and Atmospheric Processes (TAP)

Boeretang 200, B-2400 Mol, Belgium.

Context and Objectives

Clouds obstruct the viewing of the earth's surface by satellites operating in the visual and infrared spectrum. On the basis of their particular spectral behaviour, they can be detected, and eliminated from the image before further processing. Thus, a reliable cloud mask is an essential and early step in the processing chain of VEGETATION imagery.

The objective of the project is to develop such a cloud mask, as well as a snow mask and a cloud shadow mask. A 'good' cloud mask depends on a variety of factors, which have all been included, when the multicriteria decision was made to proceed with the current masks, of which the methodology is described in the next paragraph. Parameters that describe the quality of a cloud mask include : flexibility, compatibility (with the current cloud masking algorithm), speed, interpretability (or acceptability to the user community), results. Five broad types of algorithmic approaches have been developed, and many subtypes of these have been tested in an initial phase.
The broad types are described as :

- **Thresholding:** a number of thresholds on spectral bands have to be surpassed before a pixel is labelled 'cloudy'
- **Fragmentation :** the world is first divided into regions, and for each region a different algorithm is tuned to the particular situation.
- **Derived variables:** the same methodology as thresholding, except that the input variables can be all sorts of combinations of spectral bands
- **Neural networks:** various neural networks have been tuned to detect clouds
- **Fuzzy cloud mask:** instead of saying it is a cloud or not, we can assign a degree of belief, that a certain pixel is either cloudy or clear

Elements from all of the broad categories appear in the final mask, but the types that do the major part of the classification are thresholding for the cloud mask, and derived variables for the snow mask.
Little is found in literature with respect to cloud shadow detection. Nevertheless, cloud shadows alter significantly the spectral signature of ground cover. A novel approach was developed. It uses information from the cloud mask, radiometry and geometry (observation and illumination), to forecast cloud height, and consequently mask shadowed pixels.

Methodology

Early on in the conceptual phase of the project, it was decided that the cloud mask to be developed was to be based on a supervised classification. The utmost important step of such an approach is the development of a suitable training and test set. The training and test sets have to include a sufficiently large number of examples from all types of clouds (haze to thick clouds, ice clouds to water clouds,…), and over all types of surfaces (dark, bright, water, snow,…). The training and test sets used for the different algorithms that have been developed in the course of the project, have been continuously updated and enhanced.

The next step is to select an algorithmic approach to follow. As indicated, a number of approaches have been investigated, but the thresholding approach has finally been withheld (for snow, we added some derived variables).

After this was established, a way to find the ideal thresholds had to be looked for. These thresholds should be able to maximally discriminate clouds from clear sky. To this end, a dedicated optimisation engine was developed, using a genetic algorithm.

Three level cloud mask

In the course of the development of the binary cloud mask, the desire to have a more flexible mask became ever more clear. Different applications need different degrees of cloud detection. Therefore, immediately after the binary cloud mask scheme was developed we started to look for ways to add a third level of cloudiness (in the rest of this document called 'uncertain'). To achieve this, we proceeded in the following way:

We applied the binary cloud mask (based on thresholding with derived variables) to a large set of test images, covering all regions in the world (in total about 38 million km^2), and partitioned these pixels into two sets, one corresponding to cloudy pixels, and one corresponding to clear pixels. For each of those two sets, we made a cumulative distribution of the scores of the pixels within the set on each of the variables on which thresholds were set in the binary cloud mask (i.e. the BLUE and the MIR band, and two derived variables). What we did next was to apply a region of uncertainty to the binary thresholds on each of the thresholds, working in both directions, i.e. we will lower the binary threshold a fraction to arrive at a lower bound for the three level threshold, and we increase the binary threshold a fraction to find an upper bound for the three level threshold. The amount of 'uncertainty' allowed for was measured in percent of the total number of pixels declared cloudy/clear by the binary mask. Several 'amounts of uncertainty' have been evaluated, and a 10% amount of uncertainty was judged about optimal. It gets rid of most of the thin haze in images, without affecting too much erroneous classification of bright surfaces as cloudy. For snow covered regions, a post processing step was needed, because with the three level cloud mask, (partially) snow covered pixels were frequently classified as uncertain by the cloud mask, due to the very close resemblance in spectral signature of ice clouds and snow cover. This problem we solved by applying a neural network to the ambivalent pixels defined as both cloudy and clear.

The cloud mask.

Two variables determine the classification of a pixel into clear, uncertain or cloudy. These two variables correspond to the digital numbers of the reflectances of a pixel in the BLUE and the SWIR band of the VEGETATION instrument.

The following values correspond with an amount of uncertainty, described earlier, of 10%. The digital numbers of the reflectances in a certain VEGETATION spectral band are noted as ρ_{BAND}.

A pixel is declared CLEAR if:

$\rho_{BLUE} < 493$ OR $\rho_{SWIR} < 180$

A pixel is declared CLOUDY if:

$\rho_{BLUE} \geq 720$ AND $\rho_{SWIR} \geq 320$

A pixel is declared uncertain if :

It is not declared clear or cloudy by the two previous rules.

The snow mask.

A pixel is declared snow covered if all of the following conditions are met:

$\rho_{RED} \geq 615$

$\rho_{MIR} < 481$

$$\frac{\rho_{BLUE} - \rho_{NIR}}{\rho_{BLUE} + \rho_{NIR}} \cdot 1000 \geq -773$$

$$\frac{\rho_{BLUE} - \rho_{SWIR}}{\rho_{BLUE} + \rho_{SWIR}} \cdot 1000 \geq 87$$

$$\frac{\rho_{BLUE} + \rho_{RED}}{2} - \rho_{SWIR} \geq 77$$

The shadow mask

For the shadow mask, a somewhat different approach has been developed. First, the line along which a cloud pixel (cloud mask information) may cast a shadow on the earth is calculated from the observation and illumination geometry (geometric information). Then, differences of NIR channel reflectances (radiometric information) are used to determine the edge of a cloud shadow along that line. From this information, the cloud height is estimated and a shadow mask is calculated using geometry again. The algorithm is able to detect shadows at the border of clouds as well as shadows which are several pixels away from the projected cloud. This feature is quite unique for a geometric based algorithm.

Outputs and results

The outputs of the algorithm include a three level cloud mask (cloud-uncertain-clear), a two level snow mask and a two level shadow mask. A first version of the cloud mask, which was specifically tuned to the format of the current cloud mask algorithm in the VEGETATION processing chain, but which altered the thresholds to those calculated with the genetic algorithm, dramatically improved the quality of the original cloud mask. The new developed algorithm, will improve the quality of the cloud, snow and shadow mask even more when implemented.

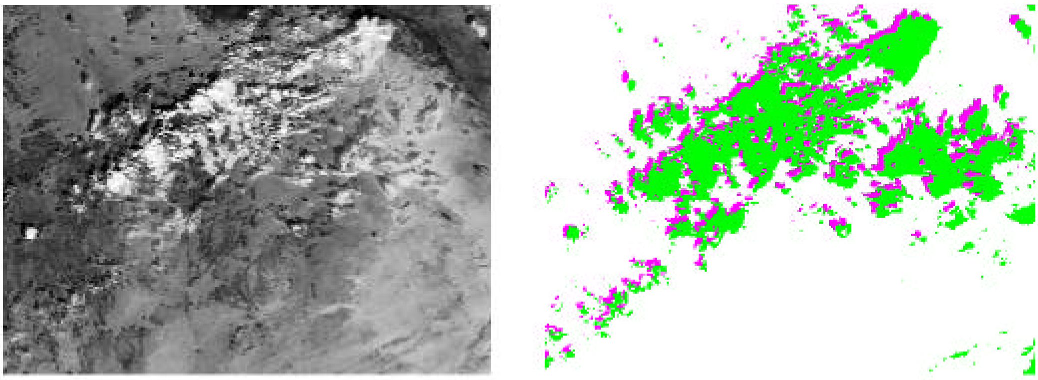

Figure 1: (Above) B3 channel of SPOT4-VEGETATION image of the Sahara desert, Algeria. (Below) cloud mask (light) and shadow mask (dark) calculated with the described algorithms

References

R. Levett, D. Campbell, Vegetation Cloud Detection Algorithm Study, EOS document EOS-95/138-RP-001, April 1995.

R.W. Saunders, K.T. Kriebel, An improved method for detecting clear sky and cloudy radiances from AVHRR data, International Journal of Remote Sensing, Vol.9, Nr.1, 1988.

J.J. Simpson, J.R. Stitt, A procedure for the Detection and Removal of Cloud Shadow from AVHRR Data over Land, IEEE Transactions on Geoscience and Remote Sensing, Vol.36, Nr.3, May 1998.

Contract and financing

Contract no. SP67008 is financed by the Belgian Federal Office for Scientific, Technical and Cultural Affairs (OSTC) in the framework of the SPOT4 - VEGETATION preparatory programme and under the Fourth Framework of the European Commission’s Directorate General XII convention with CNES No. 95/CNES/0449.

Contact

Email : Gil.Lissens@vito.be