CLOUD DETECTION MACHINE LEARNING ALGORITHMS FOR PROBA-V

Luis Gómez-Chova, Gonzalo Mateo-García, Jordi Muñoz-Marí, Gustau Camps-Valls

Image Processing Laboratory (IPL), University of Valencia, Spain

ABSTRACT

This paper presents the development and implementation of a cloud detection algorithm for Proba-V. Accurate and automatic detection of clouds in satellite scenes is a key issue for a wide range of remote sensing applications. With no accurate cloud masking, undetected clouds are one of the most significant sources of error in both sea and land cover biophysical parameter retrieval. The objective of the algorithms presented in this paper is to detect clouds accurately providing a cloud flag per pixel. For this purpose, the method exploits the information of Proba-V using statistical machine learning techniques to identify the clouds present in Proba-V products. The effectiveness of the proposed method is successfully illustrated using a large number of real Proba-V images.

Index Terms— Proba-V, cloud detection, ML

1. INTRODUCTION

The main objective of this work is to propose a cloud detection algorithm for Proba-V [1]. Images acquired by Proba-V instrument, which works in the visible and infrared (VIS-IR) ranges of the electromagnetic spectrum, may be affected by the presence of clouds.

Cloud detection approaches, also referred to as cloud masking, are generally based on the assumption that clouds present some useful characteristics for its identification. The simplest approach to cloud detection in a scene is the use of a set of static thresholds (e.g. over reflectance or temperature) applied to every pixel in the image, which provides a cloud flag (binary classification) [2]. However, the Proba-V instrument presents a limited number of spectral bands (Blue, Red, NIR and SWIR) which makes cloud detection particularly challenging since it does not present thermal channels or a dedicated cirrus band (Fig. 1). On the one hand, thicker clouds should be easily detected and masked out from visible and near-infrared Proba-V bands, but this is not true for thiner clouds, which are semitransparent to solar radiation. Moreover, bright pixels, such as ice and snow in the surface, can be misclassified as clouds. Bright land covers and clouds have a similar reflectance behavior, thus thresholds on reflectance values do not solve the problem. On the other hand, signal
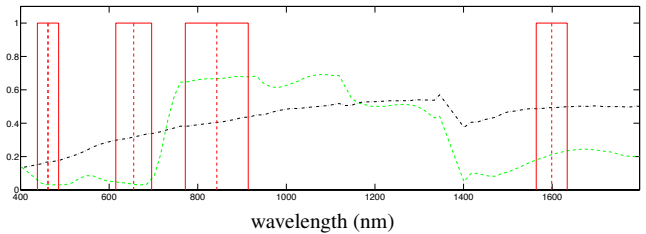
Fig. 1. Proba-V channels (boxes) superimposed to a reflectance spectra of healthy vegetation and bare soil (dash-dotted lines).

coming from optically-thin semitransparent clouds is mostly affected by surface contribution, and it ranges from very low to extremely high values depending on whether the cloud is over water or ice, respectively. Therefore, they are extremely difficult to detect from reflectance properties in VNIR data. These problems preclude the use of simple approaches based on static thresholds and suggest the use of more advanced cloud masking methods [3, 4].

Current Proba-V cloud detection uses multiple thresholds applied to the blue and the SWIR spectral bands [5], but the definition of global thresholds is practically impossible. Hence, for next Proba-V reprocessing [6], monthly composites of cloud-free reflectance in the blue band are used to define dynamic thresholds depending on the land cover type. Few works using more sophisticated machine learning tools have been presented so far in the cloud detection literature, such as Bayesian methods [7], fuzzy logic [8], artificial neural networks [9], or recently kernel methods [4, 10].

In this context, the European Space Agency (ESA) has started the study *'Clouds Detection Algorithms for Proba-V'* in order to propose and compare different cloud detection approaches for Proba-V. This paper presents our contribution in the framework of this Proba-V Clouds Detection Round Robin experiment. The proposed cloud detection algorithms rely on advanced non-linear methods capable of exploiting the information of Proba-V features in order to improve the cloud masking products.

2. METHODOLOGY

The cloud masking process relies on the extraction of meaningful physical features (e.g. brightness and whiteness) that are combined with spatial features to increase the cloud detection accuracy. Then, a supervised pixel-based classifica-

tion, based on the TOA reflectance and on a manually labeled training set, is applied to these features providing the pixel label (*cloud* or *cloud free*). The supervised classifiers to be developed and tested on this data should allow the use of a high number of input features and allow an easy integration of heterogeneous sources of information.

2.1. Proba-V Data and Ground Truth

In this work, we consider as input data Proba-V Level 2A products with TOA reflectance, i.e. the four Proba-V bands are radiometrically and geometrically corrected and resampled at 333m. The available data set for the Proba-V Round Robin[1] exercise consists of 331 products acquired in four days covering the four seasons: 21/03/2014, 21/06/2014, 21/09/2014, and 21/12/2014. A reduced set of Proba-V data for a number of representative sites worldwide is used to train the algorithm and validate its performance.

In order to train statistical machine learning models from real data, a representative number of samples have to be labelled as *cloud-contaminated* or *cloud-free* samples. To label in a semi-automatic way a sufficient number of pixels from the Proba-V images, we have adapted the user-driven methodology proposed for MERIS in [3] to the Proba-V images, where the labeling of cloud clusters found in the image is done by the user.

2.2. Feature Extraction

Several physically-inspired features can be extracted from the spectrum before applying the classification methods in order to improve their performance. In this work, we take advantage of previous research and, rather than working with the spectral reflectance only, physically-inspired features are extracted in order to increase the separability of clouds and surface covers [11]. The final set of features that are analyzed in the frame of this work are listed in Table 1.

Together with the spectral features, basic spatial features are extracted at different scales: the mean (μ) and standard deviation (σ) are computed for each pixel-based feature at two different scales in 3×3 and 5×5 windows. Summarizing, we consider the four Proba-V spectral channels (4), the spectral features described in Table 1 (10), and the mean (μ) and standard deviation at two different scales, which are computed for each pixel-based feature ($(4 + 10) \times 4$). That results in a total number of 70 possible input features (4 reflectance bands, 10 spectral features, 56 spatio-spectral features). Moreover, in order to reduce the complexity of the trained classifiers, we define two sets of features with the first 20 and 40 most relevant features that will be compared in the classification experiments.

Table 1. Cloud features extracted from Proba-V images.

Cloud Feature	Feature
Brightness	x_{Br}
Brightness VIS	$x_{Br,VIS}$
Brightness NIR	$x_{Br,NIR}$
Whiteness	x_{Wh}
Whiteness VIS	$x_{Wh,VIS}$
Whiteness NIR	$x_{Wh,NIR}$
Snow NDSI	$x_{(Blue-NIR)/(Blue+NIR)}$
Snow NDSI	$x_{(Blue-SWIR)/(Blue+SWIR)}$
Red-SWIR ratio	$x_{Red/SWIR}$
NDVI	$x_{(NIR-Red)/(NIR+Red)}$

2.3. Supervised Classification Algorithms

The extracted features and the original spectral bands are used as inputs of advanced supervised classification algorithms, which are required to solve complex classification problems such as cloud masking. The detection of clouds can be considered as a two-class classification problem. In these problems, we are given a set of ℓ labeled (training) samples $\{\mathbf{x}_i, y_i\}_{i=1}^{\ell}$, where $\mathbf{x}_i \in \mathbb{R}^d$ is defined in an input space $\mathcal{X}$, and $y_i \in \{0, 1\}$ belongs to the observation (output) space ('cloudy' or 'cloud free'). In this paper, different classification methods are analyzed: classification trees (TREE) [12], support vector machines (SVMs) [13], and multilayer perceptron (MLP) neural networks [14].

The *CART algorithm* is a tree graph structure with a sequence of nodes that are partitioned or split into two branches by means of decision rules and each terminal node (leaf) is classified with the predicted value for that node. Pruning and cross-validation methods are usually employed on CART algorithms to avoid overfitting. In this work, a 10-fold cross-validation procedure is used to find the minimum-cost tree and to estimate the best level of pruning.

The *SVM binary classifier* is a statistical learning algorithm based on constructing a maximum margin separating hyperplane in a reproducing kernel Hilbert space. SVMs allow the use of a high number of input features as it combats the curse of dimensionality efficiently. In this work, we use a Gaussian RBF kernel, given by $K(\mathbf{x}_i, \mathbf{x}_j) = \exp\left(-\|\mathbf{x}_i - \mathbf{x}_j\|^2 / 2\sigma^2\right)$; and free SVM parameters are selected by following an 8-fold cross-validation procedure in the training set.

The *MLP neural network* [14], which has been a traditional approach for supervised cloud classification [9], is also included in the comparison. In all the cases, the neurons of the hidden layer present the hyperbolic tangent sigmoid activation function while the neuron of the output layer presents a linear output function in order to better analyze the distribution of the output values.

3. EXPERIMENTS & RESULTS

In order to decide which features are more relevant, the information of the spectral channels and the extracted features for cloud detection is analyzed in terms of classification accu-
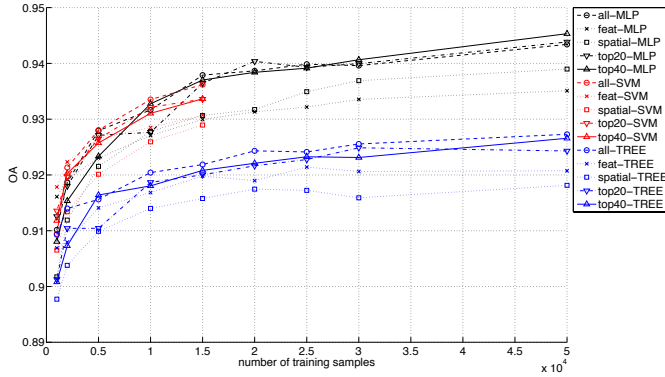
Fig. 2. Overall Accuracy (OA%) over the test sets for the analyzed methods (TREE, SVM and MLP). The number of input features (spectral, spatial, and all features) and training samples per class vary for each test set.

Fig. 3. Overall Accuracy (%) for the 54 validation images manually labeled in order to be used as reference (ground truth).

racy. In the following experiments, classifiers are trained with different numbers of training samples and different combinations of features. Figure 2 shows the Overall Accuracy (OA) for each number of features, i.e. the classification accuracy of the TREEs, MLPs and SVMs for the sets of selected features.

Several conclusions can be extracted from Fig. 2. Classification trees are very efficient classification algorithms but provide the less accurate detection results for all cases. SVMs provide excellent results when few training samples are available but present a huge computational cost when the number of samples increases (no SVM models have been trained for more than 15000 training samples per class). Finally, MLP neural networks provide excellent cloud detection accuracy and the extracted spatio-spectral cloud features drastically improve results, obtaining the best results with the top 40 selected features.

In what follows, we focus only on MLPs, which have offered improved performance, using the top 40 selected features for Proba-V. Also, and given that accuracy is not improved too much with an increasing number of samples, we concentrate on using $50,000$ training samples per class in all cases, and classification accuracy is computed using $380,000$ test samples. Figure 3 shows the Overall Accuracy (OA%) for the 54 validation images that have been manually labeled in order to be used as reference (ground truth). In this figure, one can observe that in most images the cloud detection accuracy is higher than 90%. This confirms that the trained MLP provides an excellent generalization over the analyzed images. However, it is important to remark that we are using as reference cloud mask a 'ground truth' that has been manually generated. Hence, the risk of learning and reproduce the errors present in the ground truth there always exists.

In order to better analyze the type of errors that we are committing, Table 2 shows the overall Confusion Matrix for the 54 validation images manually labeled. One can observe that the proposed method provides a balanced number of false
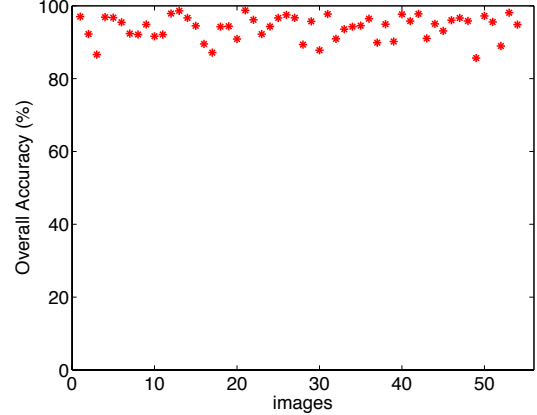
Table 2. Confusion Matrix for the 54 validation images that have been manually labeled in order to be used as reference (ground truth). True Negatives (TN), False Negatives (FN), True Positives (TP), False Positives (FP), Producer Accuracy (PA), User Accuracy (UA), and Overall Accuracy (OA).

	Manual Labels		
	Cloud-free	**Cloudy**	
Predicted Cloud-free	**TN**: 345×10^6	**FN**: 25×10^6	**PA**: 93%
Predicted Cloudy	**FP**: 9×10^6	**TP**: 191×10^6	**PA**: 95%
	UA: 97%	**UA**: 88%	**OA**: 93%

negatives (FN) and false positives (FP), although the number of FN is relatively higher. However, the average Overall Accuracy for all images is 93% and we can consider that the agreement between the predicted cloud masks and the generated ground truth is high enough.

Finally, an example of the resulting cloud mask is shown in Fig. 4. In this figure, the proposed cloud mask is benchmarked against the manually generated 'ground truth' for a Proba-V product. This image has been selected because it illustrates most common cloud detection problems (e.g. cloud borders, thin clouds, ice/snow covers). In these plots, pixels detected as cloudy pixels by the cloud detection method but labeled as cloud free in the 'ground truth' are plotted in *blue*, while discrepancy pixels classified as cloud free but marked as clouds in the 'ground truth' are shown in *orange*. The pixels detected as cloud-free by the algorithm that are labeled as cloudy in the 'ground truth' (*orange*) match snow and glaciers over mountains in the South Island of New Zealand, and some parts of the coastline. Therefore, one can assume that these differences are not errors but cloudy pixels, and that proposed method recognizes these difficult cases, although it is true that some thin cloud borders are also detected as cloud-free. However, in these cases a spatial growing of the cloud mask should easily improve the results.
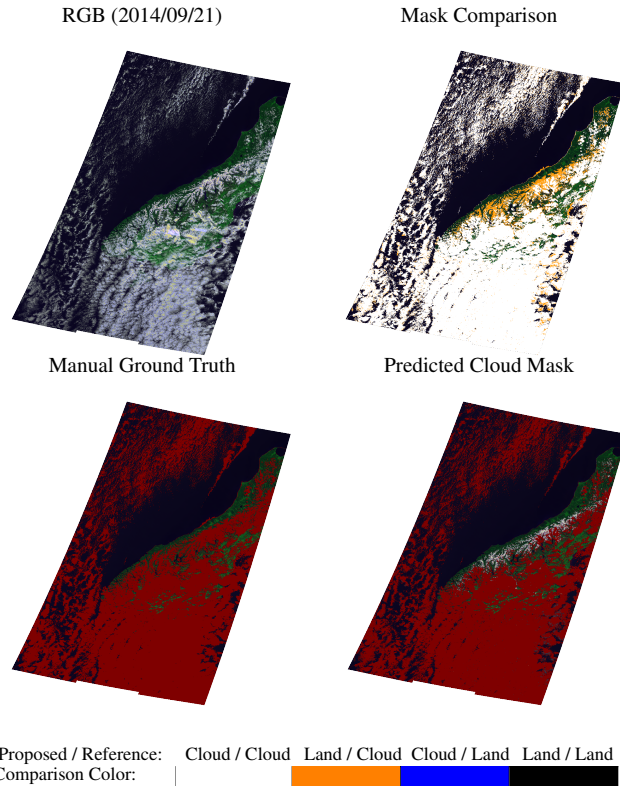
RGB (2014/09/21) Mask Comparison

Manual Ground Truth Predicted Cloud Mask

Proposed / Reference: Cloud / Cloud Land / Cloud Cloud / Land Land / Land
Comparison Color:

Fig. 4. Cloud detection example showing the RGB false color composite, the manually generated ground truth, the cloud mask obtained with the final MLP classifier, and the comparison of the 'manual ground truth' with the 'Proba-V Cloud Flag': discrepancies are shown in blue when proposed method detects cloud and in orange when pixels are classified as cloud-free.

4. CONCLUSIONS

In this paper, a methodology that faces the problem of accurately identifying the location of clouds in Proba-V images is described. The cloud masking algorithm is based on simple spatio-spectral physical features, which are intended to increase separability between clouds and ground covers, and are extracted from the converted top of atmosphere reflectance in order to reduce dependence on illumination and geometric acquisition conditions. A supervised classification is carried out based on the selected extracted features and selected training samples covering most relevant image conditions, background surfaces, and cloud types. In particular, several machine learning methods have been trained using different sets of input features and different sets of training samples in order to select the best empirical model. The final implemented method is based on artificial neural networks trained with manually labeled real data. The performance of the method has been tested on a large number of real images and on scenes presenting most critical cloud detection problems, and results show an accurate discrimination of thin clouds, cloud borders, and bright surfaces.

5. REFERENCES

[1] Wouter Dierckx, Sindy Sterckx, Iskander Benhadj, Stefan Livens, Geert Duhoux, Tanja Van Achteren, Michael Francois, Karim Mellab, and Gilbert Saint, "PROBA-V mission for global vegetation monitoring: standard products and image quality," *International Journal of Remote Sensing*, vol. 35, no. 7, pp. 2589–2614, 2014.

[2] M. Wang and W. Shi, "Cloud Masking for Ocean Color Data Processing in the Coastal Regions," *IEEE Trans. on Geoscience and Remote Sensing*, vol. 44, no. 11, pp. 3196–3105, Nov 2006.

[3] L. Gómez-Chova, G. Camps-Valls, J. Calpe, L. Guanter, and J. Moreno, "Cloud-screening algorithm for ENVISAT/MERIS multispectral images," *IEEE Trans. on Geoscience and Remote Sensing*, vol. 45, no. 12, Part 2, pp. 4105–4118, Dec 2007.

[4] L. Gómez-Chova, G. Camps-Valls, J. Muñoz-Marí, and J. Calpe-Maravilla, "Semi-supervised Cloud Screening with Laplacian SVM," in *IEEE Int. Geoscience and Remote Sensing Symposium, IGARSS'2007*, Jul 2007, pp. 1521–1524.

[5] G. Lisens, P. Kempencers, F. Fierens, and J. Van Rensbergen, "Development of cloud, snow, and shadow masking algorithms for VEGETATION imagery," in *IGARSS 2000. IEEE 2000 International Geoscience and Remote Sensing Symposium.*, 2000, vol. 2, pp. 834–836.

[6] E.L.A. Wolters, E. Swinnen, I. Benhadj, and W. Dierckx, "PROBA-V cloud detection evaluation and proposed modification," Tech. Rep. Technical Note, 17/7/2015, QWG, 2015.

[7] F. Murtagh, D. Barreto, and J. Marcello, "Decision Boundaries Using Bayes Factors: The Case of Cloud Masks," *IEEE Trans. on Geoscience and Remote Sensing*, vol. 41, no. 12, pp. 2952–2958, Dec 2003.

[8] A. Ghosh, N.R. Pal, and J. Das, "A fuzzy rule based approach to cloud cover estimation," *Remote Sensing of Environment*, vol. 100, pp. 531–549, 2006.

[9] José Antonio Torres Arriaza, Francisco Guindos Rojas, Mercedes Peralta López, and Manuel Cantón, "An Automatic Cloud-Masking System Using Backpro. Neural Nets for AVHRR Scenes," *IEEE Trans. on Geoscience and Remote Sensing*, vol. 41, no. 4, pp. 826–831, Apr 2003.

[10] L. Gómez-Chova, G. Camps-Valls, L. Bruzzone, and J. Calpe-Maravilla, "Mean map kernel methods for semisupervised cloud classification," *IEEE Trans. on Geoscience and Remote Sensing*, vol. 48, no. 1, pp. 207–220, Jan. 2010.

[11] Linlu Mei, Marco Vountas, Luis Gómez-Chova, Vladimir Rozanov, Malte Jäger, Wolfhardt Lotz, John P. Burrows, and Rainer Hollmann, "A cloud masking algorithm for the XBAER aerosol retrieval using MERIS data," *Remote Sensing of Environment*, vol. In Press, 2017.

[12] Leo Breiman, Jerome Friedman, Charles J. Stone, and R. A. Olshen, *Classification and Regression Trees*, Chapman & Hall/CRC, Monterey, CA, January 1984.

[13] B. Schölkopf and A. Smola, *Learning with Kernels – Support Vector Machines, Regularization, Optimization and Beyond*, MIT Press Series, Cambridge, MA, USA, 2002.

[14] S. Haykin, *Neural Networks: A Comprehensive Foundation*, Prentice Hall, Englewood Cliffs, NJ, USA, 1999.