Combining Fuzzy C-Mean and Normalized Convolution for Cloud Detection in IR Images

Anna Anzalone[1], Francesco Isgrò[2], and Domenico Tegolo[3]

[1] INAF - Istituto di Astrofisica e Fisica Cosmica, Palermo, Italy
[2] Dipartimento di Scienze Fisiche, Università degli Studi di Napoli Federico II, Napoli, Italy
[3] Dipartimento di Matematica ed Applicazioni, Università di Palermo, Palermo, Italy
anna.anzalone@ifc.inaf.it, francesco.isgro@unina.it, domenico.tegolo@unipa.it

Abstract. An important task for the cloud monitoring in several frameworks is providing maps of the cloud coverage. In this paper we present a method to detect cloudy pixels for images taken from ground by an infra-red camera. The method is a three-steps algorithm mainly based on a Fuzzy C-Mean clustering, that works on a feature space derived from the original image and the output of the reconstructed image obtained via normalized convolution. Experiments, run on several infra-red images acquired under different conditions, show that the cloud maps returned are satisfactory.

Keywords: Cloudiness mask, fuzzy set, infra-red images.

1 Introduction

Automated cloud detection is a challenging issue, crucial for cloud monitoring. Clouds play an important role in several fields such as the global climate change, weather forecast, climate modeling. For example the global change of the Sea Surface Temperature (SST) due to the greenhouse effect, is a study where it is necessary filtering out cloudy pixels from the satellite data, in order to avoid cloud contamination in the measurements of the temperature. For this purpose auxiliary cloud masks are used that indicate wether the single pixel is affected or not by clouds. Cloud detection process in the images from space, should take into account that cloud appearance is similar to other entities and also changes depending on the region, hence the discrimination is not a straightforward task. It is difficult to discriminate ice or snow from clouds in the polar regions, because of their similar reflectiveness in the visible and small contrast in the infra-red wavelengths. Cloud and volcanic ashes in volcanic areas, cloud and fire, cloud and dust over the deserts etc. [1]. Moreover in both space and ground observations, it must be taken into account that edges in cloud images are generally smoothed and sharp outlines are hard to be detected. The recent literature proposes methods for clear sky determination based on the idea that brightness temperature of cloudy pixels show different relationships and properties from those expected for

V. Di Gesù, S.K. Pal, and A. Petrosino (Eds.): WILF 2009, LNAI 5571, pp. 140–147, 2009.

clear conditions. A set of combined tests are applied on images from space and generally infra-red and visible bands are used. In [2] a pixel is labeled as cloudy if all tests indicate the presence of cloud. Different approaches exploit features non related to the physics of the clouds but to the spatial/temporal relationships between pixels of the same or consecutive frames [3] or determined by machine learning processes using Bayesian classification [4], decision trees [1], or support vector machines [5]. It is important to take into account that validation of a cloud mask is a very difficult issue. A way to proceed for quality estimation is assessing the agreement with the human analysis, combined with comparisons against masks detected with different algorithms.

In this paper we present an algorithm for detecting clouds from infra-red images. The algorithm is based on the Fuzzy C-Mean clustering method, and it is structured in three steps: Fuzzy C-mean clustering on the whole image, Gaussian classification on the detected clusters, Fuzzy C-mean clustering only on one of the two final clusters. The features used for clustering are the original grey-level and the value of the same image reconstructed using the normalized convolution.

The paper is structured as follows. Next section describes the method, together with a short description of the techniques used. Section 3 discusses the assessment of the method on a data set of infra-red images including different kinds of clouds. Section 4 is left to the final remarks.

2 Description of the Method

The proposed method for cloud mask detection is mainly based on three steps: Fuzzy C-mean clustering on the whole image, Gaussian classification on the detected clusters, Fuzzy C-mean clustering only on one of the two final clusters. The first step is a coarse segmentation of the image in three clusters representing sky, cloud or other scene objects. These three clusters have been used to include the most representative entities of a general cloudy scene. But in our experiments the sky and the other non-cloud elements (sky in the following) were joined because this paper focusses the attention only on cloud/non-cloud segmentation. The second step analyzes the outliers of each clusters and reassigns them by means of the normal distribution. Finally the last step is applied on the sky cluster to identify the misclassified cloud pixels, typically the border ones.

The representation of information has a relevant role in the image analysis, some different features have been considered to identify the map of cloud coverage (Grey Level, Mean and STD, Gradient, Local Histogram, Entropy,...), and an exhaustive analysis has been done to identify the best set of features. In the end the grey level feature resulted good for a first rough segmentation of the IR image, but the final mask needed to be improved by means of the support of other different features. To the scope a transformation of the grey level image based on normalized convolution, has been adopted.

Normalized Convolution: The Normalized Convolution is a method for signal analysis that takes into account uncertainties in signal values and at the same

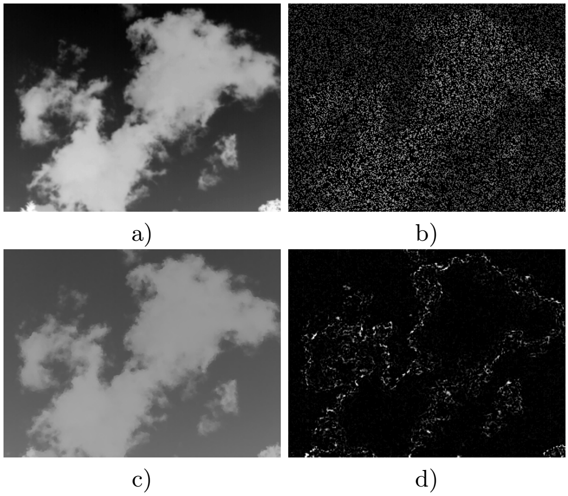

Fig. 1. Results of the normalized convolution algorithm on an IR cloud image. a) Original image, b) 10% Sampled image, c) Reconstructed image d) Difference between the original image and the reconstructed one.

time allows spatial localization of the possible unlimited functions by means their analysis.

The equation $U(\epsilon) = \sum_X a(x)B(x) \odot c(\epsilon - x)T(\epsilon - x)$ expresses the general formulation of convolution, where $\odot$, in the standard convolution, is the scalar multiplication. While the normalized convolution of aB and cT can be formulated as $U_N = \{aB\widehat{\odot}cT\}_N = N^{-1}D$ where D=$\{aB\widehat{\odot}cT\}$ and N=$\{aB \odot B * \widehat{\cdot}c\}$. Summation is an example to produce D and N over the corresponding indexes, and to give an example for a we can use the following function:

$$a = \begin{cases} r^{-\alpha} \cos^{\beta}(\frac{\pi r}{2r_{max}}) & r < r_{max} \\ 0 & \text{otherwise} \end{cases}$$

where r is the distance between the center and the nearest pixel, α and β are two positive integers. [6,7]. In figure 1 an example of the normalized convolution applied on an infra-red cloud image is shown.

The normalized convolution step extracts a sample of the pixels from the IR image (figure 1.b) and reconstructs the original image (figure 1.c). Dissimilarities among the input image and the reconstructed image result especially in the border pixels (figure 1.d). From the thermographic point of view, in these points there is a big variety of temperatures with large mutual exchanges and the uncertainty to assign these points to the sky or to the cloud sets, grows. The normalized convolution is a good support for the next step that tries to reassign those points. Therefore the original and the reconstructed grey levels are the only two characteristics considered by the next step to cluster the clouds. Note that the normalized convolution grey level is used as feature just to improve the border pixel classification.

A Fuzzy C-mean algorithm for three clusters has been chosen for the first step of the segmentation.

Fuzzy C-Mean: Standard segmentation is based on the use of attributes or features to distinguish different objects inside an image. Such characteristics are extracted during the low level vision phase to characterize afterwards objects in the high level vision step. In image data acquisition, all pixels of each frame are acquired in synchronous way. Let is k the number of observations of different elements, they can be grouped in N-dimensional vectors $z_k = [z_{1k}, z_{2k}, ...z_{nk}]^T, z_k R^n$ We can define a set of N observations as follows:

$$Z = \begin{pmatrix} z_{11} & z_{12} & \cdots & z_{1N} \\ z_{21} & z_{22} & \cdots & z_{2N} \\ \vdots & \vdots & \vdots & \vdots \\ z_{n1} & z_{n2} & \cdots & z_{nN} \end{pmatrix} \tag{1}$$

In the case of dynamic systems the matrix Z can contain sample of signals or different scale of signal. In our case two columns have been considered at different level of signal. Aim of the clustering is to segment the data in different classes. K-means clustering, is one of the simplest unsupervised classification algorithm, it can be arranged to segment objects that appear in the images [8]. Defined the number of clusters, the procedure follows a simple way to classify the data-set. The algorithm partitions a set of N vectors $X = \{x_j, j = 1..N\}$ into C classes $c_i, i = 1, .., C$. It finds a starting cluster centre for each class, named cluster centroid, then an objective function of dissimilarity has to be minimized [9]. If an Euclidean function is considered, the function $P = \sum_{i=1}^{c} \left(\sum_{k, x_k \in c_i} \parallel x_k - v_i \parallel^2 \right)$ must be minimized where v_i is the centroid of the cluster v_i. A binary matrix $U = (u_{ij})$ defines a membership matrix as

$$u_{ij} = \begin{cases} 1 & \text{if} \parallel x_j - v_i \parallel^2 \leq \parallel x_j - v_k \parallel^2, \forall k \neq i \\ 0 & \text{otherwise} \end{cases}$$

where $v_i = \frac{\sum_{x_j \in v_i j=1}^{N} X_j}{\|c_i\|}$. Outcome of this step is a rough segmentation of the image according to the selected features and the defined clusters. The next two steps refine the retrieved clustering improving the classification of the outliers and resuming the likely misclassified cloudy pixels. In fact the Gaussian classification step identifies some few outliers present in each cluster and switch them to the correct cluster. Then the Fuzzy C-mean algorithm, assuming that the cloud cluster is correct, extracts from the sky cluster some points that could be included in the other one.

3 Experiments

The method described in the previous section can be summarized as follows:

Norm_Conv: Computes the Normalized Convolution for a 10% pixels of the original image;

Fuzzy_Im: Runs the Fuzzy C-Mean algorithm on the whole image to segment sky and cloud areas;

Gauss_Im: Applies a Gaussian function to include outlier pixels to the correct cluster;

Fuzzy_Sky: Runs Fuzzy C-Mean algorithm only on the sky cluster, to move wrongly classified pixels into the cloud cluster.

The step **Norm_Conv** requires some parameters where windows, on which to apply the Gaussian function for the convolution, and radius, are the most important ones. Their values have been set to 2 and 4 based on experimental tests. Moreover the method works on 10% uniform random selection of pixels of the input image. The step **Fuzzy_Im** segments the image in sky and cloud pixels, according to a set of parameters: maximum number of iterations, minimum amount of improvement, and number of clusters present in the data set. We have defined 14 as maximum number of iterations, e^{-5} as minimum amount of improvement and 3 as cluster number. Cloudy and non cloudy pixels characterize the step **Gauss_Im**. Mean and standard deviation are calculated for each cluster, then two gaussian probability values are evaluated for each pixel. Their maximum value allows to reassign the pixel to another cluster leaving the outliers moving from a cluster to another one. In the step **Fuzzy_Sky** has been applied a Fuzzy C-Mean on the non cloudy cluster with the cluster parameter equal to 2 and the other options defined equal to the ones in the step Fuzzy_Im.

3.1 Data

In order to test the algorithm efficiency, we selected some different infra-red cloud images and figure 2 shows some of them. Cloud images acquired in the infra-red

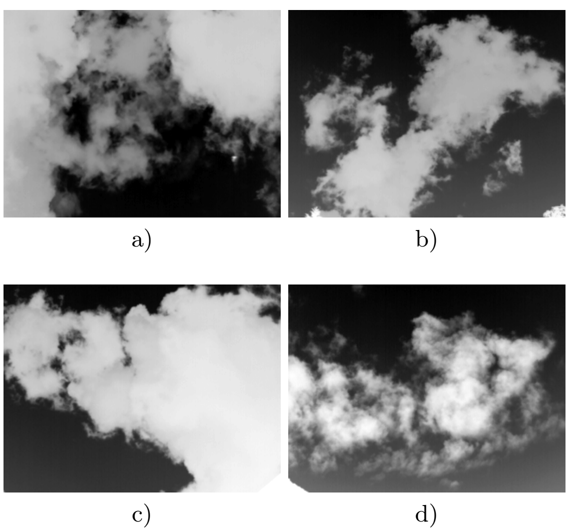

Fig. 2. Some examples of infrared cloud images

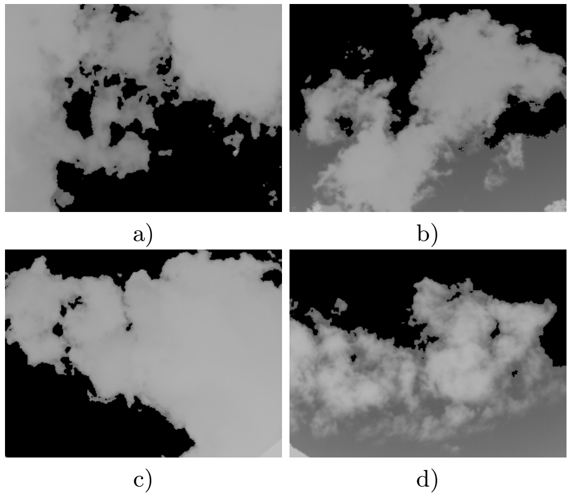

Fig. 3. Results after the first step

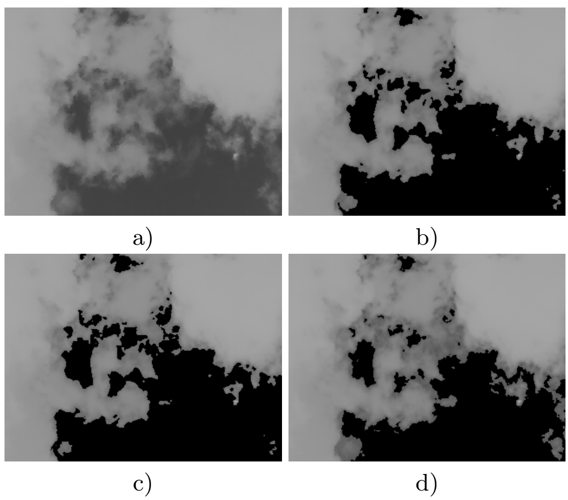

Fig. 4. Intermediate and final results for the image in figure 2 a). a) Normalized convolution step, b) First Fuzzy C-mean on whole image, c) Gaussian step; d) Second Fuzzy C-mean step on the sky cluster.

spectrum were selected from the archives of the Department of Mathematics and Application (DMA), University of Palermo, they include different kind of clouds such as cirrus, strato-cumulus, cumulonimbus,etc. The image sequences were acquired by the FLIR S-65 thermal-camera with a 18mm lens, spectral range 7.5-13μm and 320×240 pixels. The sequences have been processed by a FLIR software to extract single frames. At present our database includes a

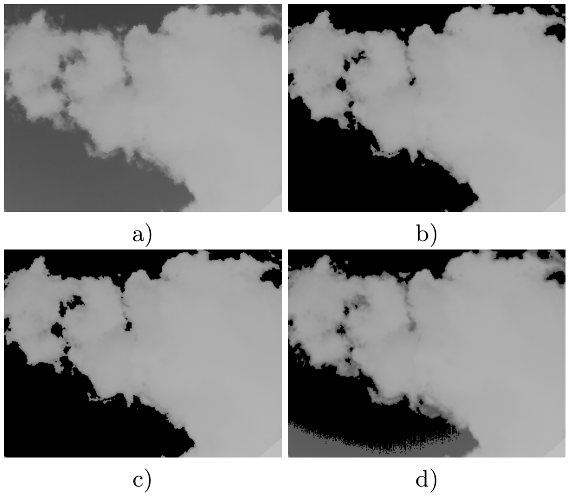

Fig. 5. Intermediate and final results for the image in figure 2 c). a) Normalized convolution step, b) First Fuzzy C-mean step on whole image, c) Gaussian step; d) Second Fuzzy C-mean step on the sky cluster.

large amount of sequences and many frames have been analyzed to evaluate the goodness of the method.

3.2 Results

Figure 3 depicts the images proposed in figure 2 after the first step of the method. Although the result of this step gives a reasonable classification of the pixels, we can observe that some pixels of cloud are lost and some other pixels are wrongly assigned as cloud, around the edges of the image. The background or sky pixels will be eliminated from the cloud cluster by the next step.

Figures 4 and 5 show the output of each step of the method for two portions of sky present in figure 2 a) and c) : sub-image a) shows the first step (Fuzzy C-mean algorithm), sub-images b) and c) illustrate the middle steps and finally the sub-image d) displays the cloudiness mask.

4 Conclusion

In this paper we tackled the problem of cloud segmentation from infra-red images, proposing a method that combines the normalized convolution with the Fuzzy C-mean clustering algorithm. Cloudiness masks were computed on infrared images acquired from ground including mainly two entities: sky and clouds. Furthermore the method was assessed considering several kind of clouds with different configurations. The quality of the results is satisfactory, although we could also give a qualitative validation. In the near future we plan to validate the algorithm on different kind of images from different sensors and also on images from space where different entities should be taken into account such as land, snow, sea etc. Finally it needs to be tested against other methods.

References

1. Shiffman, S., Nemani, R.: Evaluation of decision trees for cloud detection from AVHRR data. In: IEEE International Geoscience and Remote Sensing Symposium, pp. 5610–5613. IEEE Press, Los Alamitos (2005)
2. Ackerman, S., et al.: Discriminating clear sky from clouds with MODIS. J. Geophysical Research. 103, 141–157 (1998)
3. Seiz, G., Baltisavias, E.P., Gruen, A.: Cloud Mapping from the Ground:Use of Photogrammetric Methods. J. Photogrammetric Engineering and Remote Sensing, 941–951 (2002)
4. Murtagh, F., Barreto, D., Marcello, J.: Decision Boundaries Using Bayes Factors:The Case of Cloud Masks. J. IEEE Trans. on Geoscience and Remote Sensing 14, 2052–2958 (2003)
5. Lee, Y., Wahba, G., Ackerman, S.A.: Cloud classification of satellite radiance data by multicategory support vector machines. J. of Atmospheric and Oceanic Technology 21(2), 159–169 (2004)
6. Knutsson, H., Westin, C.F.: Normalized and differential convolution: Methods for interpolation and filtering of incomplete and uncertain data. In: Proceedings of Computer Vision and Pattern Recognition 1993, pp. 515–523 (1993)
7. Knutsson, H., Westin, C.F., Westelius, C.J.: Filtering of uncertain irregularly sampled multidimensional data. In: Conference Record of The Twenty-Seventh Asilomar Conference on Signals, Systems and Computers 1993, vol. 2, pp. 1301–1309 (1993)
8. Lee, J., Liu, R.: A Fuzzy Clustering Algorithm Based on Fuzzy Distance Norms for Asynchronously Sampled Data. In: 11th IEEE International Conference on Computational Science and Engineering, CSC 2008, pp. 361–368 (2008)
9. El-Melegy, M., Zanaty, E.A., Abd-Elhafiez, W.M., Farag, A.: On Cluster Validity Indexes in Fuzzy and Hard Clustering Algorithms for Image Segmentation. In: IEEE International Conference on Image Processing, vol. 6, pp. 5–8 (2007)