HIGH LEVEL SEMANTIC LAND COVER CLASSIFICATION OF MULTITEMPORAL SAR IMAGES USING SYNERGIC PIXEL-BASED AND OBJECT-BASED METHODS

Donato Amitrano, Raffaella Guida, Pasquale Iervolino

Surrey Space Centre, University of Surrey, Guildford, UK

ABSTRACT

Land cover mapping is one of the classic applications of synthetic aperture radar remote sensing. However, despite of the algorithmic progress in classification techniques, the semantic content of available maps does remain unchanged, with only a few macro-classes (like water, forest, urban, and bare soil) being discriminated in the majority of the works from past years. In this paper, a methodology to extract a higher level semantics from synthetic aperture radar images is presented. It is based on coupling pixel-based clustering with object-based image analysis and contextual information. Preliminary results have been produced from multitemporal SAR datasets over a forest area in Colombia. They demonstrate that the synergic exploitation of pixel and object information can provide higher quality land cover results and more information to map users.

Index Terms— synthetic aperture radar, land cover classification, high-level semantics, multitemporal, object-based image analysis

1. INTRODUCTION

The open access data policy applied by the European Space Agency (ESA) to the Copernicus mission is rapidly revolutionizing the remote sensing downstream sector, both at industrial and academic level. Such an abundance of data is greatly boosting the downstream market [1] and is posing new challenges to the scientific community.

Among them, the improvement of available land cover maps is probably one of the most urgent. In particular, when applied to synthetic aperture radar (SAR) datasets, classification is, for instance, hard to perform, as testified by the wide literature published on the topic in the last years. Indeed, most of the efforts concentrate on the development of new algorithms and techniques able to handle large bulks of data, eventually in cloud environment (like the Google Earth Engine [2]) and exploiting deep architectures [3]. However, a more critical analysis of the recent literature on the topic reveals that, despite of the algorithmic progress, the semantic level of the output maps has not really changed in the last 20 years [4], [5]. This is due to the fact that, generally, the pixel level is the privileged source of information. Therefore, the challenge is the development of new algorithms able to enrich the semantic level of land cover maps.

In this work, a new algorithm for multitemporal SAR land cover mapping is presented. It exploits both pixel-based and object-based techniques to extract a high number of highlevel classes. Novelties are introduced both in the processing phase and in the output quality. A synergy between pixelbased and object-based methods is exploited to improve the classification output through the enhancement of the semantic label attached to each class. In other words, a more specific verbal attribute with respect to the classic "Water", "Urban", "Forest", etc. is provided as output to better explain the meaning of some identified patterns and their interconnection with other scene features.

The work is organized as follows. The general methodology is discussed in Section 2. Preliminary experimental results are presented in Section 3. Conclusions are drawn at the end of the work.

2. METHODOLOGY

The general workflow for high-level semantic land cover classification is depicted in Figure 1. A multitemporal stack of Sentinel-1 SAR ground range detected (GRD) images is used as input and processed for the generation of temporal change-detection color composites as described in [6]. Actually, the original chain was designed for ingesting complex products, since the interferometric coherence was used as one of the temporal channels. Therefore, the processing has been adapted here to fit with detected images. In particular, the coherence information, useful to enhance the presence of urban areas, is replaced by a texture measure, the data range, having a comparable information content. The data range is defined as follows

$$d_r = max(W) - min(W), \tag{1}$$

where max(W) and min(W) represent the maximum and the minimum backscattering value in a moving window, respectively. The data range is basically an edge detector. It is expected to be high in presence of high-texture areas (like urban areas) and low where the texture is typically negligible (water surfaces, forests, etc.).

Following the workflow reported in Figure 1, the change-detection RGB composite is then processed for semantic clustering as discussed in [7]. This technique allows



Fig. 1. Workflow for the generation of high-level semantic land cover maps.

for automatically enriching the output cluster map with a basic semantics, i.e. each cluster is labelled with a verbal attribute related to its color. As explained in [7], this attribute can be exploited to pre-select an over-dimensioned group of clusters sharing the backscattering characteristics which are expected to originate from a certain scene pattern. This preclassification map is then refined through a feature-adaptive object-based reasoning aiming at selecting only the image segments having the geometric and/or contextual characteristics relevant to a specific class. This is useful to split macro-clusters, grouping more image features, into more specific classes. As an example, water surfaces are expected to have low backscattering values. However, the class "Water" is very generic since it includes at least three features like rivers, lakes, and sea surfaces. They can be identified if object information is considered. In [7], a complete discussion about the extraction of small reservoirs in semiarid environment has been provided. Here, that framework is generalized and specified for several classes and features, whose list is provided in Table 1.

The main goal is to further decompose the macro-clusters identifiable by pixel-based clustering by using geometric and contextual information. This way, as an example, the "Water" class can be split in "River", "River trunk", and "Lake". These features, indeed, share the same scattering attribute, being characterized by low reflectivity typical of water surfaces. However, they are clearly distinguishable if the geometry of the relevant image segments is analyzed. Rivers are elongated objects, while lakes are compact. In image processing, these concepts can be expressed as follows. The elongatedness is given by [8]

$$F_E = \frac{A}{W^2},\tag{2}$$

where *A* is the object area and *W* the number of cycles needed to completely shrink the image segment. The compactness is defined as follows [9] $F_c = \frac{4\pi A}{P^2},\tag{3}$

where P is the object's perimeter. This measure expresses how much the object is shaped like a circle.

The scattering and geometric properties of the objects are combined in a fuzzy classification system [10] able to automatically assign the output class based on the likelihood of each segment with low backscattering to belong to one of the two possible classes, i.e. "River" (high elongatedness) and "Lake" (high compactness). Objects identified as "River" but not fully satisfying the elongatedness criterion are downgraded to "River trunk". These clusters are likely to represent small river portions disconnected from the principal cluster due to the presence of dominant dry soil and/or because of dimensions at the limit of the sensor resolution.

Once the class "River" has been identified, it can be used to better specify the spatial arrangement of the macro-class "Bare soil" within the scene. Indeed, this is typically a not structured class, i.e. bare soils rarely have a well-defined shape and/or a precise collocation within the scene. However, when a bare soil cluster is adjacent to a river cluster, it can be labeled as "Sediment". Moreover, small areas completely enclosed in a "Sediment" cluster and having a mixed backscattering (i.e. not clearly associable to one of the macrogroups defined in Table 1) can be referred as "Mixed sediments."

Moving away from the river, bare soils can be classified as "Bare soil river proximity" and "Bare soil" depending on whether they fall or not within a properly defined buffer zone around the river itself. This allows the map user to understand a further spatial relation within the scene, since areas in proximity of rivers are subject to different problems and/or regulations.

Following similar reasoning, it is possible to define the 18 classes listed in Table 1 starting from the two other macrogroups (vegetation and urban) typically retrievable using

ID	Class name	Macro-group	Explanation
1	Shadow	Low scatterers/Water	Low reflectivity backslope surfaces, not structured
2	Layover	High scatterers/Urban	High reflectivity foreslope surfaces, typically structured
3	River	Low scatterers/Water	Highly elongated, low reflectivity
4	River trunk	Low scatterers/water	Moderately elongated, low reflectivity, small area
5	Lake	Low scatterers/water	Highly compact, low reflectivity
6	Sediments	Soil	Low reflectivity, not structured, adjacent to rivers
7	Mixed sediments	Soil	Mixed reflectivity, not structured, surrounded by sediments
8	Soil, river proximity	Soil	Low reflectivity, not structured, close to rivers
9	Forest	Vegetation	High texture, medium reflectivity
10	Mixed forest	Vegetation	Mixed reflectivity, surrounded by forest, not structured
11	Sloped forest	High scatterers/Urban	High reflectivity foreslope surfaces adjacent to dominant forest
12	Bare soil	Soil	Low reflectivity, not structured
13	Dense urban	High scatterers/Urban	High reflectivity big structured clusters
14	Village	High scatterers/Urban	High reflectivity small clusters, not structured
15	Peri-urban	High scatterers/Urban	High reflectivity small clusters close to structured urban areas
16	Green urban	Vegetation	Vegetation clusters totally surrounded by urban areas
17	Mixed urban	High scatterers/Urban	Unidentified clusters surrounded by dense urban areas
18	Isolated built-up	High scatterers/Urban	Isolated, structured high reflectivity clusters

Table 1. High-level semantic classes defined through the proposed workflow with their explanation.



Fig. 2. High-level semantic land cover mapping results. (a) Input RGB change-detection product. (b) Semantic clustering: each cluster is associated to a low-level semantic color label. (c) Output of the proposed workflow, 18 classes as reported in Table 1. (d) Support vector machine output, 4 macro-classes (water, forest, soil, urban).

pixel-based processing. The implemented algorithm is not fully described for brevity. However, the details provided in the last column of Table 1 should help the reader in figuring out the information layers involved in the specification of each class. As for the classes previously described, they are combined in a fuzzy system for retrieving the high-level class best fitting with the geometric, scattering, and contextual properties of the selected pre-classification mask.

3. PRELIMINARY EXPERIMENTAL RESULTS

In Figure 2, the output of the proposed workflow for highlevel semantic land cover mapping is reported. In particular, in Figure 2a, the input RGB change detection product is shown. The correspondent low-level semantic clustering is depicted in Figure 2b. The reader should note that the appearance of the RGB product and of its clustered version is very similar, and this allows for an immediate semantic transferring between the two products, as discussed in [7]. In Figure 2c, the 18-classes classification map resulted from the proposed workflow is shown. In the following, this map will be referred to as HLS map. Finally, in Figure 2d, the output of a four-class (water, urban, forest, and soil) supervised support vector machine (SVM) is reported for comparison.

To this end, the 18-classes of the HLS map have been grouped in the four macro-classes reported in Table 1. A ground truth of about 500 pixels for the "Water" class, 6000 pixels for the "Soil" and the "Urban" classes, and 20000 pixels for the "Forest" class has been manually retrieved via photo-interpretation of the patch shown in Figure 1.

In Table 2, the confusion matrix for the SVM classification is reported. The overall accuracy with respect to the retrieved ground truth is 61.3%. The kappa coefficient is 0.41.

In Table 3, the confusion matrix for the HSL classification is reported. The overall accuracy with respect

to the retrieved ground truth is 88.2%. The kappa coefficient is 0.79.

	Water	Soil	Urban	Forest
Unclass.	0.00	0.00	0.00	0.00
Water	72.9	0.20	0.24	1.30
Soil	25.1	77.2	20.7	28.3
Urban	0.00	0.00	30.4	34.1
Forest	2.00	22.6	48.6	36.2

Table 2. Confusion matrix for the SVM classification.

Table 3.	Confusion	matrix for	the HLS	classification.
1 uoie 5.	Comusion	mault in	i une muo	ciubbilicution.

	Water	Soil	Urban	Forest
Unclass.	0.00	0.18	0.00	3.13
Water	96.0	0.24	0.00	0.00
Soil	3.41	86.9	1.06	9.02
Urban	0.00	0.00	97.7	1.88
Forest	0.60	12.7	1.27	86.0

The HLS classification exhibits much higher values for all the considered quality indicators if compared with SVM classification. In particular, the latter is penalized by huge interclass confusion for the classes "Forest" and "Urban" which are highly affected by layover phenomena in the original SAR image. Working with objects and contextual information in HLS classification allows for the mitigation of this problem. Layover can be effectively eroded by considering the spatial arrangements of the clusters, since layover areas actually covered by forest are usually enclosed by or adjacent to dominant "Forest" clusters.

The same applies for the "Urban" class. This is typically very heterogeneous, with several different land covers falling in it. Using pixel-based approaches, the built-up is mostly identified. However, this is only one of the features constituting an urban agglomerate. In classic approaches, it is usual to find some "Forest" spots within the cities, which is a nonsense. Using objects, it is possible to include green areas within the "Urban" class which, in the HLS classification, shows a lower interclass confusion, as reported in Table 3.

4. CONCLUSIONS

A new framework for high-level semantic land cover mapping has been proposed to improve the quality of the output of SAR images classification. The framework exploits the synergy of pixel-based and object-based methods to specify the macro-clusters usually defined in the established SAR literature in more detailed classes better defining the spatial arrangement and the context of the scene features.

The proposed workflow has been applied to a dataset concerning a forest area in Colombia. The obtained preliminary results, when compared with literature classification procedures, show the superiority of the joint exploitation of pixel-based and object-based image analysis. The proposed methodology is fully unsupervised and is expected to be a promising solution for the generation of high-quality land cover maps

5. ACKNOWLEDGMENTS

The research presented in this paper is supported by the UK Space Agency within the grant "UK-IPP Projects – Space-Enabled Monitoring of Illegal Gold Mines"

6. REFERENCES

[1] European Space Agency, "Copernicus in Action -Fostering User Uptake of EO Services through the Copernicus Masters and the Space App Camps," 2016.

[2] N. Gorelick, M. Hancher, M. Dixon, S. Ilyushchenko, D. Thau, and R. Moore, "Google Earth Engine: Planetary-scale geospatial analysis for everyone," *Remote Sens. Environ.*, vol. 202, pp. 18–27, 2017.

[3] J. Geng, J. Fan, H. Wang, X. Ma, B. Li, and F. Chen, "High-Resolution SAR Image Classification via Deep Convolutional Autoencoders," *IEEE Geosci. Remote Sens. Lett.*, vol. 12, no. 11, pp. 2351–2355, 2015.

[4] L. Bruzzone, M. Marconcini, U. Wegmüller, and A. Wiesmann, "An Advanced System for the Automatic Classification of Multitemporal SAR Images," *IEEE Trans. Geosci. Remote Sens.*, vol. 42, no. 6, pp. 1321–1334, 2004.

[5] K. Kayabol and J. Zerubia, "Unsupervised amplitude and texture classification of SAR images with multinomial latent model," *IEEE Trans. Image Process.*, vol. 22, no. 2, pp. 561–572, 2013.

[6] D. Amitrano, G. Di Martino, A. Iodice, D. Riccio, and G. Ruello, "A New Framework for SAR Multitemporal Data RGB Representation: Rationale and Products," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 1, pp. 117–133, 2015.

[7] D. Amitrano, F. Cecinati, G. Di Martino, A. Iodice, P.-P. Mathieu, D. Riccio, and G. Ruello, "Feature Extraction From Multitemporal SAR Images Using Selforganizing Map Clustering and Object-Based Image Analysis," *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.*, vol. 11, no. 5, pp. 1556–1570, 2018.

[8] M. Nagao and T. Matsuyama, *A Structural Analysis of Complex Aerial Photographs*. New York: Plenum Press, 1980.

[9] E. P. Cox, "A Method of Assigning Numerical and Percentage Values to the Degree of Roundness of Sand Grains," *J. Paleontol.*, vol. 1, no. 3, pp. 179–183, 1927.

[10] D. Amitrano, G. Di Martino, A. Iodice, D. Riccio, and G. Ruello, "Unsupervised Rapid Flood Mapping Using Sentinel-1 GRD SAR Images," *IEEE Trans. Geosci. Remote Sens.*, vol. 56, no. 6, pp. 3290–3299, 2018.