Urban areas enhancement in multitemporal SAR RGB images using adaptive coherence window and texture information

Donato Amitrano *Student Member, IEEE,* Veronica Belfiore, Francesca Cecinati, Gerardo Di Martino, *Member, IEEE,* Antonio Iodice, *Senior Member, IEEE,* Pierre-Philippe Mathieu, Stefano Medagli, Davod Poreh, Daniele Riccio, *Fellow, IEEE,* Giuseppe Ruello, *Member, IEEE*

***Abstract*—In this paper, we present a technique for improving the representation of built-up features in model-based multitemporal SAR RGB composites. The proposed technique exploits the MAP3 framework to generate an *a priori* information which is used to implement an adaptive selection of the coherence window size. Image texture is used to support the coherence information in case of decorrelation. The coherence information, powered by texture analysis and combined with backscattering amplitude, provides an unique representation of built-up features, which allows for an immediate detection of urban agglomerates by human operators and is an advantaged starting point for urban area extraction algorithms.**

***Index Terms*—multitemporal sar, fuzzy logic, data representation, urban areas**

I. Introduction

The first contact with Earth observation data is realized by visual inspection, through which images are understood [1]. However, this is not an easy task, especially dealing with synthetic aperture radar (SAR) data, whose interpretation requires a technical expertise typically not held by multidisciplinary users.

As stated in [2], "a good knowledge representation design is the most important part of solving the understanding problem". Therefore, the remote sensing community started to think about new process model, in which the machine is a support for the operator in taking decisions [1], [3] rather than an executor of an algorithm completely unknown to the analyst. Clearly, the realization of an effective human-machine interaction, also takes place through a comfortable data visualization. Recently, some of the authors introduced the MAP3 framework [4] to build two classes of multitemporal RGB composites [4], [5]. The products have the aim i) to lower the expertise level required for managing SAR data, providing products easily interpretable and processable with simple algorithms [6], and ii) to shorten the distance between Level-1 and Level-2 products in a user-oriented environment, in which the fundamental requirements of repeatability, processing automation, and ease of data interpretation must be satisfied.

The main characteristic of these products, named as Level-1α [4] and Level-1β [5], is that the association color-object, being physical-based and guided by electromagnetic scattering models, does not depend on the scene. In this work, we focus in particular on the built-up feature. This class of objects is characterized by precise temporal properties: high (and stable) backscattering and high interferometric coherence, even when computed with a long temporal baseline. These properties make built-up features appear white in Level-1α images, which are obtained combining the backscattering amplitude of two acquisitions and their interferometric coherence (see [4] for details).

Coherence is estimated in a moving window, whose typical dimensions are of several meters in order to avoid bias [7]. However, this choice is not optimal for all the scene targets. Moreover, decorrelation can occur due to several causes, such as orbital instability, baseline length, shadowing, feature shape and so on. The lack of coherence contribution makes the built-up class appear in cyan (thanks to the contribution of the amplitude bands), thus introducing an alteration of the expected semantic.

In this work, we present a technique that aims at improving the information content of the inter-

Donato Amitrano, Veronica Belfiore, Gerardo Di Martino, Stefano Medagli, Antonio Iodice, Davod Poreh, Daniele Riccio and Giuseppe Ruello are with the Department of Electrical Engineering and Information Technology, University of Napoli Federico II, Napoli, Italy.

Francesca Cecinati and Pierre-Philippe Mathieu are with the European Space Agency, ESA ESRIN, Frascati (Rome), Italy

ferometric coherence band. This technique exploits an input Level-1α product for generating *a priori* information that is used for i) adapting the coherence window dimension to the scene target and ii) introducing a texture measure which is used to identify the built-up feature class when it does not response to the variation of the coherence window dimension.

This work was designed under the aegis of the MODISTA project, and in collaboration with by Ansaldo STS, which is one of the largest worldwide players in the market of trains and rail infrastructures. The objective of the project is the satellite monitoring of railway infrastructures, including railroads, stations and all the objects placed in their proximity, such as trellises or lamp posts. Railways are particularly critical in the response to the interferometric coherence. In fact, their structure, consisting of a small trihedral surrounded by stochastic scatterers (soil, stones, vegetation), typically causes an underestimation of the interferometric coherence. Therefore, in this case the introduction of texture can help in the restoration of the scene semantic.

In Section II we introduce the coherence feedback system for the MAP3 framework. The results of the feedback application are presented in Section III. In Section IV, we face a building extraction application using the enhanced Level-1α product and compare it with two literature machine learning methods. The assessment of the performed experiments is provided in Section V. Conclusions are drawn at the end of the work.

II. Enhanced MAP3 Framework

As known, the interferometric coherence is computed by the relation:

$$\gamma = \frac{E\left[I_1 \cdot I_2^*\right]}{\sqrt{E\left[\left|I_1\right|^2\right] \cdot E\left[\left|I_2\right|^2\right]}} \tag{1}$$

where $E[\cdot]$ and $*$ indicate the mathematical expectation and the complex conjugation operations, and I_1 and I_2 are the master and slave images, respectively.

In theory, the results of Equation (1) are obtained by averaging a large number of images acquired simultaneously. Obviously, this procedure is not possible. Therefore, in practical situations, under the assumption of ergodicity, it is possible to exchange the average of several images acquired simultaneously with a spatial average in a limited area surrounding the considered pixel [7]. Therefore, the maximum likelihood estimator of the coherence in a window including N pixels is given by [8]

$$|\hat{\gamma}| = \frac{\left|\sum_{n=1}^{N} I_1^{(n)} I_2^{*(n)}\right|}{\sqrt{\sum_{n=1}^{N}\left|I_1^{(n)}\right|^2 \sum_{n=1}^{N}\left|I_2^{(n)}\right|^2}}. \tag{2}$$

The pdf of the coherence magnitude estimator can be written as a function of the magnitude of the true coherence $|\gamma|$ and the number of independent samples $L > 2$ [9]

$$\begin{aligned}&\text{pdf}\left(|\hat{\gamma}|, |\gamma|, L\right) = 2\left(L-1\right)\left(1-|\gamma|^2\right)^L |\hat{\gamma}| \cdot \\ &\left(1-|\hat{\gamma}|^2\right)^{L-2} {}_2F_1\left(L, L, 1, |\gamma|^2, |\hat{\gamma}|^2\right),\end{aligned} \tag{3}$$

where ${}_2F_1$ is the hypergeometric function. The expectation for $|\hat{\gamma}|$ is given by [7]

$$\begin{aligned}&E\left[|\hat{\gamma}|\right] = \frac{\Gamma\left(L\right)\Gamma\left(3/2\right)}{\Gamma\left(L+1/2\right)} \cdot \\ &{}_3F_2\left(3/2, L, L; L+1/2, 1; |\hat{\gamma}|^2\right)\left(1-|\hat{\gamma}|^2\right)^L,\end{aligned} \tag{4}$$

where Γ is the gamma function and ${}_3F2$ is the generalized hypergeometric function. The expectation expressed by Equation (4) is plotted in Fig. 1. From this graph, it arises that the estimate is biased towards higher values for low coherence and/or when the estimation window is small [10].

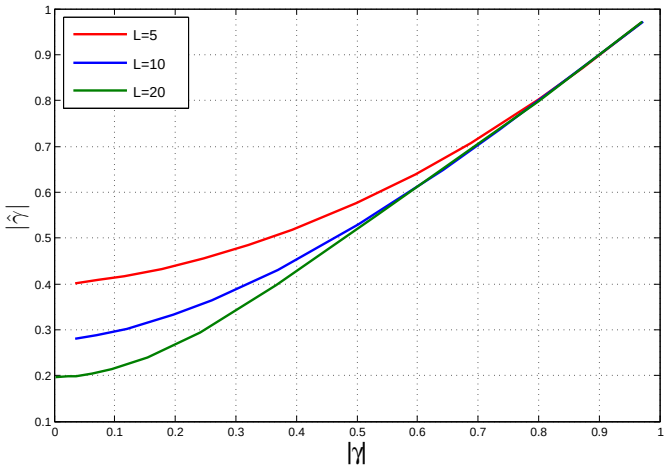


Fig. 1: The expectation value $|\hat{\gamma}|$ as function of the true coherence $|\gamma|$ for various L.

Equation (2) requires the selection of the mean window dimension, which determines the perfor-

mance of the estimator with respect to the scene objects [11]. As an example, rough surfaces (like the sea) exhibit a stochastic and non-stationary backscattering, which is expected to produce incoherent signals. Thus, as explained above, a small coherence window could produce a noisy coherence. Conversely, man-made structures, having a deterministic stable scattering, typically exhibit high coherence values. Especially if the urban texture is not dense, a large computation window could include features like vegetation, shadows, or roads together with buildings, thus causing a decrease in the resulting coherence, together with a poor resolution of the computed map.

In Fig. 2a we show a 3-meters resolution Level-1α product [4] of the city of Castel Volturno (Italy). The product is composed as follows: on the blue band, an image acquired on December 2009 (reference image) is loaded; the green band depicts an image acquired on August 2010 (test image); the red band is reserved for the interferometric coherence between the two images computed using an 11-pixel side mean window. For further details about Level-1α processing can be found in [4].

(a) (b)

Fig. 2: Castel Volturno: Level-1α product (reference image December 2009, test image August 2010) computed setting the coherence window to (a) 11 pixels and (b) 3 pixels.

About eight months passed between the two acquisitions. Therefore, we expect that only stable features, like buildings, keep high values of the interferometric coherence. Being these features also highly reflective, urban structures should be represented in white color. However, since the interferometric coherence is computed using a mean window (i.e. the interferometric coherence map has a coarser resolution than the intensity products), stable targets are surrounded by a red "crown", which helps the human photo-interpreter in their detection.

The product depicted in Fig. 2a is consistent with the above considerations. Anyway, it is reasonable to think that a mean window of about 30 meters is too large for representing at best the details of the urban areas. In fact, the reader may have the impression that white/red pixels are a bit sparse for being representative of a dense urban area. This means that the choice of coherence window dimension was not optimal since it did not allow to fully exploit the characteristics of the estimator defined in Equation (1).

In Fig. 2b we repeated the same experiment setting the coherence window to three by three pixels. In this case, it is clear that the image has an unacceptable granularity, which is physically inconsistent, besides being visually unpleasant. In fact, as an example, wide portions of the sea surface exhibit high coherence values. In this case the estimate on stochastic targets is affected by a bias which increases as the coherence window dimension decreases [7]. Anyway, it is also true that the urban area is better represented using a smaller window, since it is possible to appreciate that more details arise compared to the product depicted in Fig. 2a. Therefore, a strategy for adapting the coherence window to the scene target is needed.

To this end, we propose a feedback system, whose rationale is depicted in Fig. 3. The system is structured as follows:

- Level-1α product analysis: the characteristics of intensity and coherence of the MAP3 output are evaluated;
- Adaptive coherence window generation: an adaptive coherence window is implemented based on the above analysis; in such way the coherence estimated by Equation (1) is optimized for the considered target;
- New coherence map: the final coherence map is assembled and given as input in the quantization process for the generation of the refined Level-1α products.

In the following, we provide details about the implementation of the proposed feedback.

A. *Fuzzification*

The coherence feedback system is based on the fuzzification of Level-1α's coherence and intensity bands. We modeled these variables using three fuzzy sets with verbal attributes of "low" (Z-type),

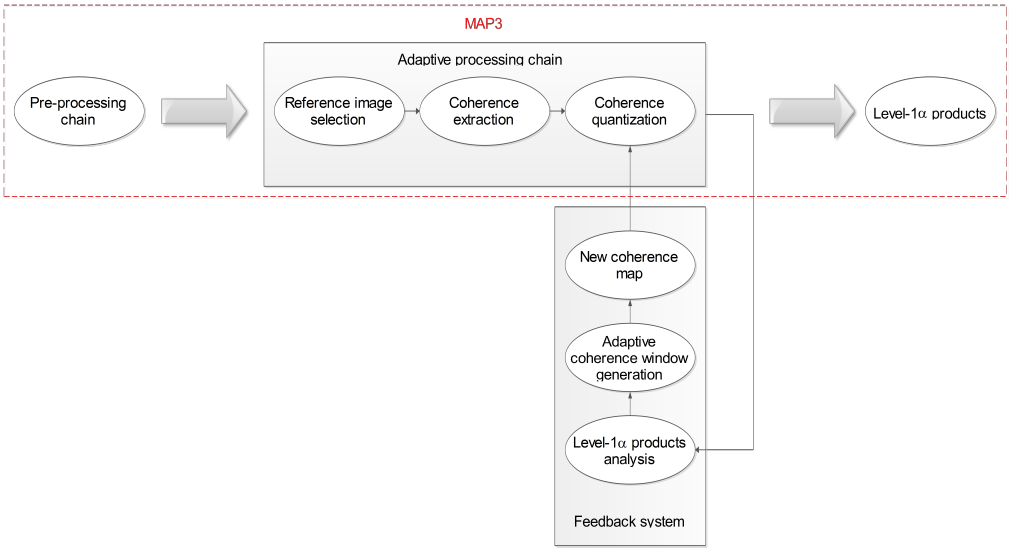


Fig. 3: Coherence feedback system for the MAP3 framework.

"medium" (π-type) and "high" (S-type). In particular, the following expression was adopted [12]:

$$S(DN,a,b,c) = \\ = \begin{cases} 0, & DN \le a \\ 2\left\{(DN-a)/(c-a)\right\}^2, & a < DN \le b \\ 1-2\left\{(DN-c)/(c-a)\right\}^2, & b < DN \le c \\ 1, & DN > c, \end{cases} \tag{5}$$

in which DN means digital number and a, c and $b = (a+c)/2$ are the parameters that model the S-function (see TABLE I for details). The Z-type fuzzy set is obtained from the S-type one being $Z(DN,a,b,c) = 1 - S(DN,a,b,c)$. Finally, the π-type function is built by combining a S-type and a Z-type function sharing the parameters c and a, respectively, as shown in TABLE I. The obtained fuzzy sets are plotted in Fig. 4.

TABLE I: Parameters used for modeling the adopted fuzzy set for Level-1α's coherence and intensity bands and for the coherence window.

	Level-1α			Coherence window		
Function type	a	b	c	a	b	c
S-type	160	190	220	21	36	51
Z-type	0	60	120	0	2.5	5
π-type (S-part)	40	100	160	3	7	11
π-type (Z-part)	160	185	210	11	26	31

The purpose of this system is to adapt the dimension of the coherence window as a function of the phase-stability and reflectivity characteristics of the target. Essentially, we aim at reducing the mean window as the coherence and the intensity responses increase. To this end, the system output (i.e. the coherence window size) is fuzzified as well as the characteristics of the input Level-1α products. In particular, we considered three types of windows, "small", "medium" and "large", which have been modeled similarly with a S-rule, a π-rule, and a Z-rule, respectively.

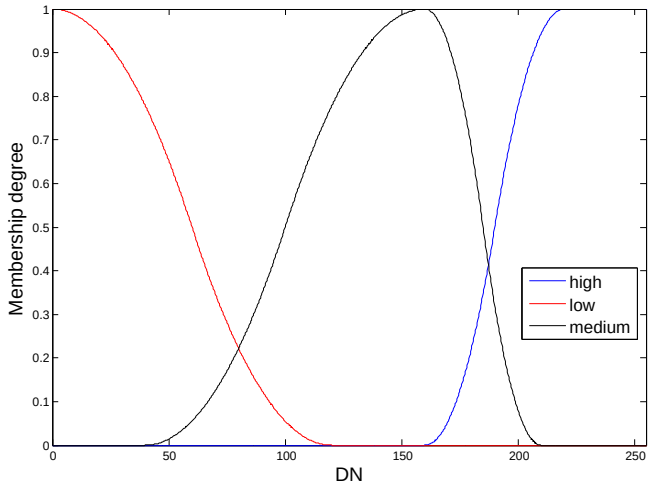


Fig. 4: Fuzzy set adopted for the fuzzification of the input Level-1α product.

The assignment of the coherence window category for each image target is now in order. The rationale is quite simple: the more stable and reflective the target, the smaller the coherence window. In TABLE II, we report the adopted rules for the

assignment of the fuzzy categories “small” and “medium” to the coherence window. All the other cases are reserved to the “large” window attribute.

TABLE II: Adopted rules for the assignment of the coherence window attribute.

Window	Coherence	Intensity 1	Intensity 2
Small	High	High	High
	Medium	High	High
	High	Medium	High
	High	High	Medium
Medium	Medium	Medium	Medium
	Low	High	High
	Medium	Medium	High
	Medium	High	Medium
	High	Medium	Medium

The selected fuzzy sets for the coherence window are depicted in Fig. 5. They have been partitioned as a function of the membership degree. As an example, the fuzzy set “small” is subdivided in three parts. The first part corresponds to persistent scatterers (PSs), and is reserved to pixels with the highest membership within this fuzzy set, which are likely to be the more reflective and coherent in the input Level-1α product. As the membership of “small” window decreases, the window dimension increases to three and five pixels, since the correspondent targets seem further away to behave as PSs. Similar reasoning can be performed for the “medium” window fuzzy set. As for the “large” window fuzzy set, we do not define any partition, since all the scatterers belonging to it are expected to be stochastic.

The coherence window attribute (“small”, “medium” or “large”) and, as a consequence, the dimension of the coherence window according to the aforementioned partitions of the fuzzy sets, is assigned by computing the maximum membership degree after the application of the rules reported in TABLE II.

It is remarkable that an ambiguity arises concerning the “medium” coherence window. In fact, due to the bell-shape of this fuzzy-set, two windows dimensions correspond to each membership degree, one for the S-part of the π and one its Z-part. However, these windows correspond to targets with very different characteristics. In fact, on the S-part of the π, we expect to have targets with medium-high characteristics of reflectivity and coherence. On the contrary, on the the Z-part of the π, we expect to find targets more likely to be incoherent. Therefore, we require that targets which lie in the S-part of the “medium” window fuzzy set must have “high” reflectivity in both the images which compose the Level-1α product. Otherwise, they are placed on the Z-part of the π, on the side of the largest windows associated to this fuzzy set.

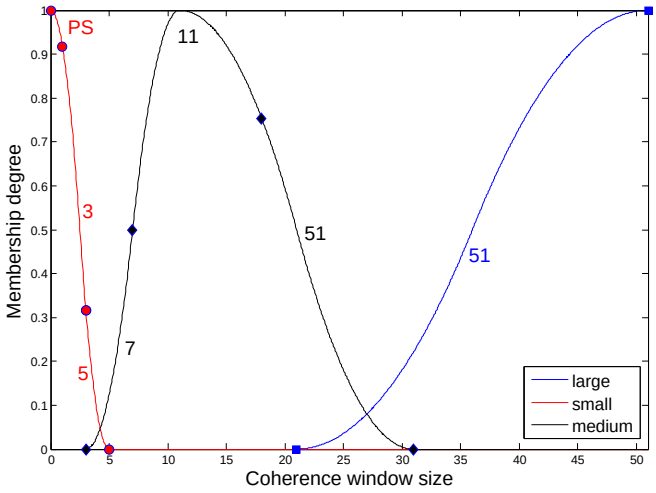
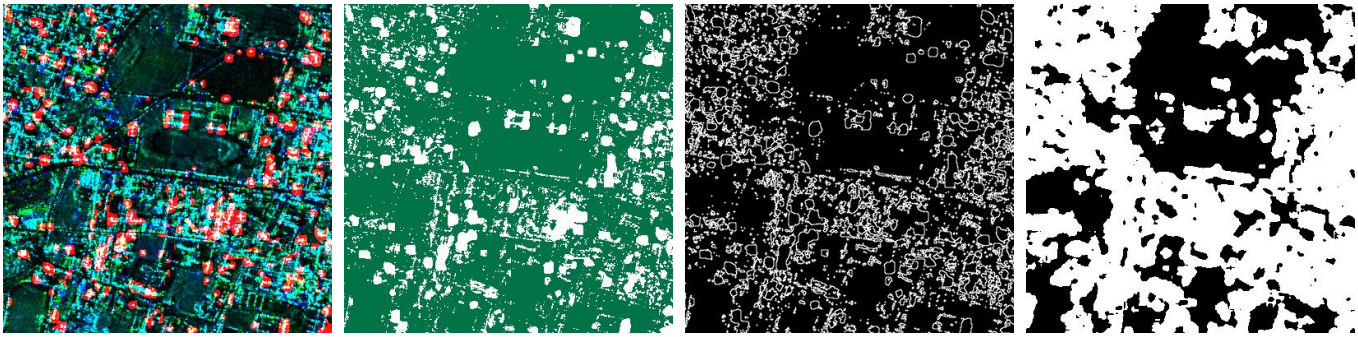


Fig. 5: Fuzzy set adopted for the fuzzification of the coherence window. The values labeling the curves indicates the dimension of the window used for that piece of curve. PS stands for “permanent scatterer”; in this case the coherence is not calculated but assumes a (high) default value.

B. Use of texture

Coherence response of targets depends on several factors. Decorrelation can occur, also on stable targets, due to imaging geometry (incidence angle), shadowing, misalignment with respect to the flight direction or the shape of the building. As an example, a squared-shape building with smooth walls is more likely to have a more coherent response than one with an irregular shape or a pitched roof. Therefore, it is possible that some buildings do not exhibit a coherent response after the reduction of the coherence window.

In Level-1α imagery, built-up features are represented in white color, due to the high contribution of both amplitude and interferometric channels. However, as explained above, decorrelation could cause this feature to appear in cyan. Here, we want to introduce a texture measure for enhancing the red band informative content. This way, it is possible to restore the semantic the user expects on built-up features, making this feature class appear white.



(a) (b) (c) (d)

Fig. 6: San Prisco (Italy): (a) Level-1α product, (b) two-categories k-means clustering, (c) edge map and (d) Nagao-Matsuyama texture.

The principle we exploit is the convergence of evidences [13]. Until now, we have assumed that the built-up class is characterized by high amplitude (in both the dates that constitute the Level-1α bi-temporal composite) and high interferometric coherence. The introduction of a fourth evidence, i.e. the texture, allows to slacken the requirements on the interferometric coherence and to consider as built-up also the objects that, even after the feedback application, do not exhibit a high response to the coherence estimator but are highly reflective (in both Level-1α's acquisitions) and located in areas characterized by high texture.

The texture measure we use was introduced by Nagao and Matsuyama in [14]. To obtain it, we first applied a k-means clustering to the input Level-1α product (see Fig. 6b and Fig. 6a, respectively). A coarse clustering is sufficient for our purposes. In fact, as shown in Fig. 6b, we just separated the white, red and cyan color (grouped in the white class in the classified map) from all the others (green class). The white class can be roughly associated with built-up features.

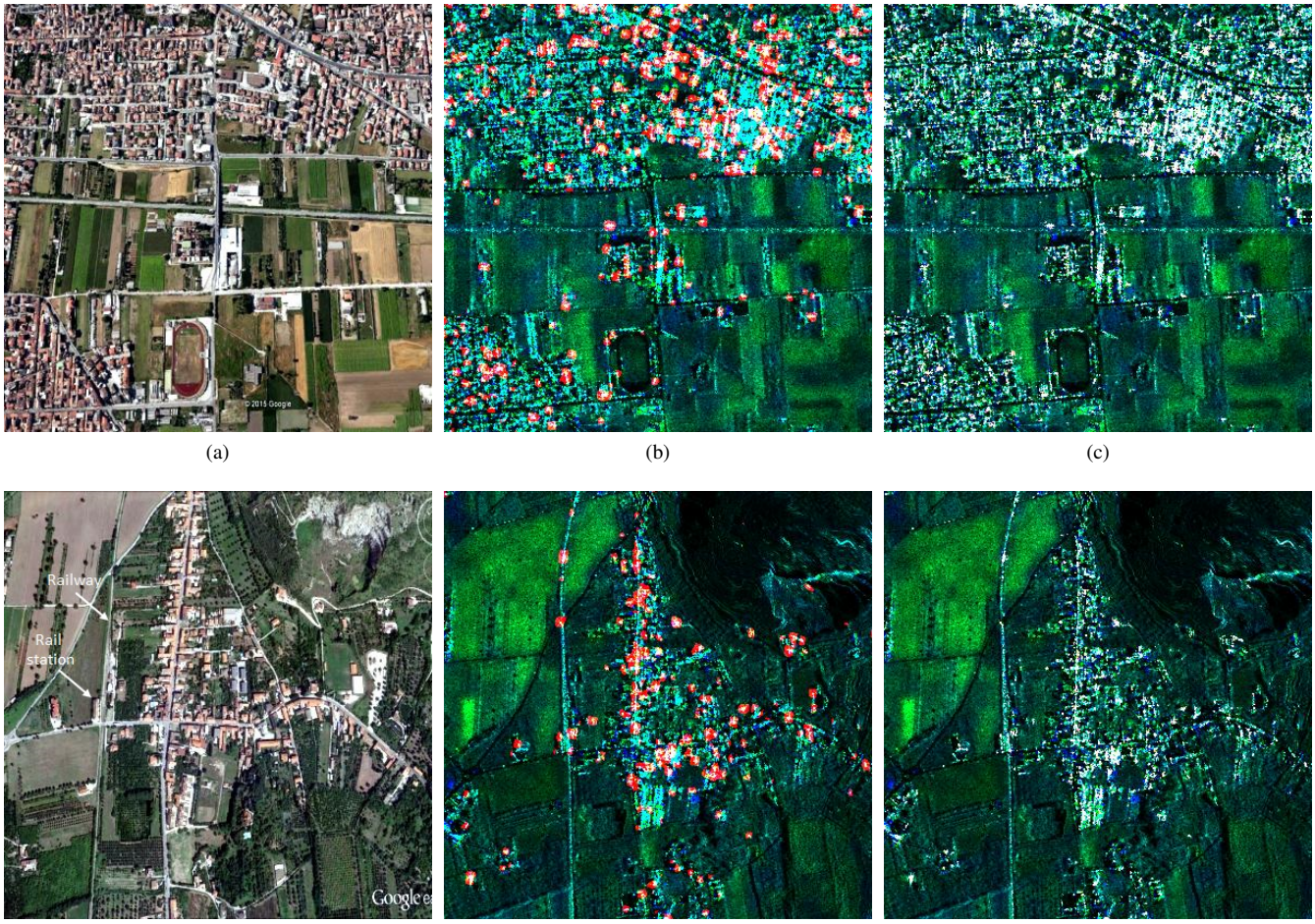
The cluster map is used for contour extraction through the application of the second Laplacian operator [15] (see Fig. 6c). Finally, the Nagao-Matsuyama texture can be computed. It is a very simple rule: a moving window of dimension N is applied to the contour map; if in a window fall at least $2N + 1$ border pixels, then the central pixel of the window is classified as high texture area. Otherwise, it is classified as a low texture area. The result of the application of this rule to the computed edge map is depicted in Fig. 6d.

The texture evidence is activated (for pixels classified as high texture areas) on objects characterized by high backscattering in both Level-1α acquisitions and low coherence. Obviously, the information conveyed by the red band for these targets does not concern anymore their phase stability.

III. Experimental results

The proposed feedback aims at enhancing the built-up features in Level-1α images in order to make the urban area extraction/mapping process easier. This is a current problem in the remote sensing community, which provided several studies about this topic. Reference [16] proposed a semi-automatic solution for mapping urban areas using backscattering, statistical information, and information fusion techniques. In [17], a multi-resolution and multi-sensor data fusion technique for mapping urban areas is presented. Reference [18] proposed a novel method for the automatic detection of building footprints from a single VHR SAR image. Marin *et al.* [19] introduced a new approach to building change detection in multitemporal VHR SAR images based on backscattering variability. A semi-automatic segmentation-based tool for urban area interpretation in SAR images was proposed in [20]. Reference [21] presented a new model combining amplitude SAR data and textural information into a Markov random field model to address the problem of classifying images of urban areas. Reference [22] proposed an interactive framework exploiting Markov random fields for the classification of multitemporal SAR data.

The objective of the feedback introduced in Section II is to make Level-1α images suitable to be processed with a simple rule, in a end-user-oriented framework, for extracting the built-up feature. It is based on the convergence of four evidences:



(a) (b) (c)

(d) (e) (f)

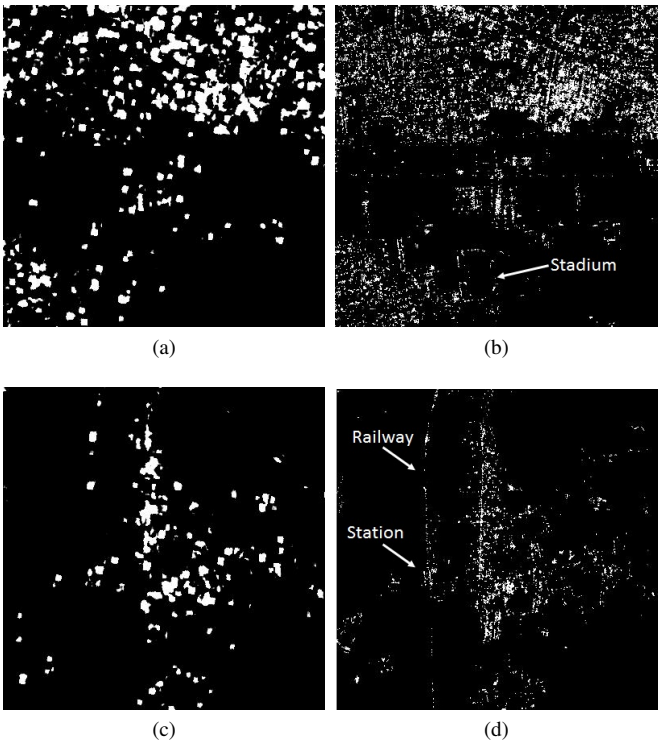
Fig. 7: (a) Google Earth view of Macerata Campania stadium with its correspondent Level-1α products (b) before and (c) after the feedback application. (e) Google Earth view of Sant'Angelo in Formis rail station with its correspondent Level-1α products (d) before and (f) after the feedback application.

the amplitudes of the two acquisition composing the input Level-1α product, their interferometric coherence and the texture. The first three evidences can be considered strong, since high characteristic of reflectivity and coherence are almost an exclusive property of built-up features. Texture, instead, is assumed to be a weak evidence, because in SAR imagery high texture areas are not necessarily urban areas. However, the convergence of two strong evidences (i.e. the high backscattering in the two Level-1α's acquisitions) and one weak evidence (i.e. the texture) make us confident that the target we are considering is man-made.

In Fig. 7a we show a Google Earth view of Macerata Campania, a small city in southern Italy. In Fig. 7b and Fig. 7c the correspondent Level-1α products before (coherence window set to 11 pixels) and after the feedback application, respectively, are depicted. Qualitatively, the reader should appreciate as the feedback system improves the representation of the built-up feature class. In fact, in Fig. 7c, more bright targets are visible than in Fig. 7b. Moreover, the red "crown" surrounding coherent targets is practically disappeared. This means that the resolution of the coherence map is higher.

In Fig. 7d we show a Google Earth view of the city of Sant'Angelo in Formis (Italy), with its rail station and railway at the left of the residential area (see annotations on the picture). In Fig. 7b and Fig. 7c the correspondent Level-1α products before (coherence window set to 11 pixels) and after the feedback application, respectively, are depicted. The same considerations made for the Macerata Campania scene can be made: the feedback application allows to retrieve a number of bright targets, i.e. more buildings, in the residential area, more details on the railway and improves the resolution of the coherence map.

These claims will be much more evident if Fig. 8 is considered. In particular, in Fig. 8a and Fig. 8b we reported the red band (i.e. the one representing the coherence information) of the products displayed in Fig. 7b and Fig. 7c. In Fig. 8c and Fig. 8d the red band of the products displayed in Fig. 7e and Fig. 7f are depicted. All the maps in Fig. 8 have been obtained through linear quantization of the input coherence product between the interval $[0.45, 0.6]$.



(a) (b) (c) (d)

Fig. 8: Macerata Campania: coherence map (a) before and (b) after the feedback application corresponding with Level-1α products depicted in Fig. 7b and Fig. 7c, respectively. Sant'Angelo in Formis: coherence map (c) before and (d) after the feedback application corresponding with Level-1α products depicted in Fig. 7e and Fig. 7f, respectively.

It is remarkable that more bright points appear in Fig. 8b and Fig. 8d (after the feedback application) than in the maps depicted in Fig. 8a and Fig. 8c obtained by applying Equation (1) with fixed 11 pixels window. This means that the urban area is better characterized. In fact, as an example, in Fig. 8b, the shape of the stadium is recognizable (see the annotation on the picture) while in Fig. 8a it is not.

In Fig. 8d (after the feedback application) it is remarkable that more details of the railway station arise than in Fig. 8c. Moreover, the shape of the railway is now visible at the left of the residential area (see annotation on Fig. 8d).

These experiments should qualitatively convince the reader of the effectiveness of the proposed algorithm. A quantitative assessment is provided in the following section.

IV. Feature extraction

We tested the effectiveness of the proposed methodology with a feature extraction application, i.e. building extraction. The technique we used is a simple, user-oriented band product [23]. The obtained results are then compared with those given by some literature technique based on machine learning. In particular, support vector machines (SVMs) and self-organizing maps (SOMs) will be used to classify data. In such way, we will prove that the enhancement of the informative content of the input Level-1α product, in combination with a simple band product, provides results fully comparable with those achievable using more and more complicated techniques, both supervised (SVMs) and unsupervised (SOMs).

A. Building extraction using the Building Index

The feedback application, enhancing the informative content of the red band of Level-1α imagery, allows for an effective feature extraction using a simple band product. The Building Index (BI) has been proposed in [23], and it is here reported for the ease of the reader:

$$BI = \frac{RGB}{255^3}, \tag{6}$$

where R, G and B are the values of the red, green and blue bands of the input Level-1α product, respectively.

Equation (6) has been applied to the input Level-1α product before and after the feedback application. In both cases a threshold of 0.1 has been adopted for maps binarization.

The results of this activity are presented in Section V. They will be compared with a ground truth extracted from the Urban Atlas of the European Environmental Agency.

B. Buildings extraction using self-organizing maps

Self-organizing maps (SOMs) are a machine learning technique introduced by Kohonen [24], which are exploited for the classification of the most diverse data types in several applications [25]–[27].

SOMs have been widely exploited in remote sensing applications, in combination both with SAR and multispectral/hyperspectral data [5], [28]–[33]. The flexibility of the tool allowed for such a widespread diffusion.

A SOM is a network composed by a user-defined number of nodes, connected with (usually) a rectangular structure. They are randomly initialized and trained with sample vectors randomly taken from the data. Each time a training element is presented to the network, the most similar node is detected and identified as Best Matching Unit (BMU). The BMU and its neighborhood, defined by a radius are updated to become more similar to the presented training element. This operation is repeated for all the training vectors and for several cycles. As the cycle number increases, the neighbour of the BMU decreases; in such way the SOM becomes stable (usually a high number of cycle is needed) and the obtained nodes can be used to classify data.

Here, we used 1000 training vectors and 200 epochs for performing buildings extraction using a 3×3 SOM. The input was a couple of SAR images, i.e. the same amount of information used for extracting features using the BI. After classification, the clusters representative of building features were selected. The obtained results are discussed in Section V.

C. Buildings extraction using support vector machines

In the last decade, support vector machines (SVMs) have been extensively introduced in the statistical learning theory domain for regression and classification problems [34]. In this methodology, the optimal separation surface between classes are used to identify the most representative training samples, which are called "support vectors". The quality of the support vectors and of the classification result depends on many parameters such as number of training samples, kernel, penalty parameter. Therefore, some expertise and a tuning phase are necessary to use this tool.

SVMs have been widely exploited for the classification and segmentation of remote sensing images [31], [35], [36]. Here, we ran a SVM with radial basis kernel on the stack composed by the two intensity SAR images composing the Level-1α product for identifying buildings. The obtained results are discussed in Section V.

V. Assessment

The Urban Atlas was used for a quantitative assessment of the experiments above illustrated. It is a land cover map in which the density of urban areas is measured with respect to the soil sealing (SL), i.e. the covering of the ground by an impermeable material. In particular, here we are particularly interested in five urban categories of the urban atlas: "Continuous urban fabric" (SL $> 80\%$)", "Discontinuous Dense Urban Fabric" (SL $50 - 80\%$), "Discontinuous Medium Density Urban Fabric" (SL $30-50\%$), "Discontinuous Low Density Urban Fabric" (SL $10 - 30\%$) and "Discontinuous Very Low Density Urban Fabric" (SL $< 10\%$).

The binary maps retrieved using the BI, the SOM, and the SVM were compared with the polygons provided by the Urban Atlas. In particular, we compute the zonal statistics for each class, obtaining the total number of building pixels that fall in the polygons belonging to that class. Therefore, for each class, the following equation was computed

$$N^i = \sum_{k^i} n_k^i, \tag{7}$$

where N^i is the total number of building pixels for the $i-th$ class, k^i is the index of the polygon belonging to the $i-th$ class, and n_k^i is the number of building pixels which fall in the $k-th$ polygon of the $i-th$ class.

The Urban Atlas class list can be found in TABLE III together with the SL (when specified) and the total area of the classes relatively to the study area.

Once N^i has been computed for each class, the percentage of soil (on average) covered by built-up pixelsfor the $i-th$ class is obtained by relation

$$\bar{SL}^i = 100r\frac{N^i}{A^i}, \tag{8}$$

where r is the map cell size and A^i the total area covered by polygons belonging to the $i-th$ class.

In Fig. 9 a picture of the study area containing both the input Level-1α product and the urban atlas layer is shown. The results of the application of the above described assessment procedure are reported in TABLE III.

As expected, the percentage of soil covered by the built-up class significantly increased after the feedback application. In particular, the SL index passes:

TABLE III: Area covered by built-up pixels compared with Urban Atlas categories before and after the feedback application. In the table headings SL stands for soil sealing; BI for building index; SOM for self-organizing map; SVM for support vector machine.

Urban atlas categories	SL (%)	Area [km^2]	BI SL before (%)	BI SL after (%)	SL SOM (%)	SL SVM (%)
Continuous Urban Fabric	> 80	2.77	12.11	24.83	24.23	31.11
Discontinuous Dense Urban Fabric	50-80	5.22	9.6	20.54	20.09	26.09
Discontinuous Medium Density Urban Fabric	30-50	1.09	4.81	11.04	11	14.61
Discontinuous Low Density Urban Fabric	10-30	0.46	2.33	5.59	5.55	7.49
Discontinuous Very Low Density Urban Fabric	< 10	0.08	1.39	3.35	3.05	4.57
Agricultural, Semi-natural areas, Wetlands		42.1	0.13	0.7	0.73	1.02
Isolated structures		0.6	1.82	3.96	3.96	5.35
Industrial, commercial, public, military and private units		4.13	4.11	9.18	9.46	12.76
Other roads and associated land		25.6	0.54	1.32	1.23	1.64
Land without current use		0.2	1.33	3.56	4.3	6.06
Fast transit roads and associated land		0.41	0.14	0.5	0.52	0.71
Railways and associated land		0.43	1.54	4.61	3.5	6.44
Mineral extraction and dump sites		0.32	0.2	10.28	7.89	11.87
Construction sites		0.04	5.26	7.62	7.12	9.42
Green urban areas		0.09	4.03	10.21	9.31	12.06
Sports and leisure facilities		0.34	4.28	8.4	8.59	11.02
Forests		6.37	0.06	2.32	2.17	2.92
Water bodies		1.16	0.09	0.2	0.59	0.94

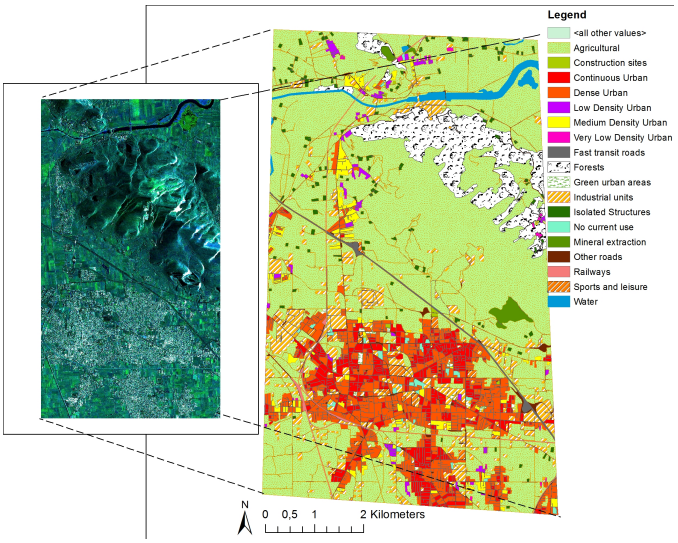


Fig. 9: Level-1α product of the study area with the correspondent urban atlas layer.

from 12.11% to 24.83% for the class "Continuous Urban Fabric", from 9.6% to 20.54 for the class "Discontinuous Dense Urban Fabric", from 4.81% to 11.04% for the class "Discontinuous Medium Density Urban Fabric", from 2.33% to 5.59% for the class "Discontinuous Low Density Urban Fabric" and from 1.39% to 3.35% for the class "Discontinuous Very Low Density Urban Fabric". It is remarkable that, even after the feedback, this values are very far from the percentages indicated in the urban atlas class description (see TABLE III). This can be explained. In fact, consulting the urban atlas mapping guide, we know that the macro-class "Urban Fabric" (i.e. the one that contains all those above mentioned) is formed by "built-up areas and their associated land, such as gardens, parks, planted areas and non-surfaced public areas and the infrastructure" [37]. Therefore: i) in the reference urban atlas class, other land covers are included beyond built-up; ii) some of this land cover (roads, sidewalks, car parks) do not have the fundamental property of high backscattering we use for the detection; iii) shadowing effects, influencing backscattering, partially prevent the feature detection using Equation (6). The last phenomenon is more severe as the density of the urban area increases. That's why the more dense the urban area, the more distant the soil sealing index retrieved through Equation (8) on the Level-1α product compared to the one indicated in the urban atlas.

Concerning the MODISTA project, it is remarkable that the proposed algorithm brings significant benefits in the representation and detection of the class "Railways and associated land". In fact, in this case, the percentage of land covered by built-up features increased from 1.54% before the feedback

to 4.61% after its application.

The results obtained using the SOM are reported in the sixth column of TABLE III. They are comparable with those obtained by application of Equation (6) on the enhanced Level-1α product, thus confirming the reliability and the effectiveness of the proposed representation to be exploited for the extraction of the feature of interest using a simple and user-oriented technique [6].

SVM showed the highest percentage of soil sealing for all the considered categories. However, this is not necessarily an index of better detection since data concerning false alarms are not available. Indeed, using this technique, their amount is expected to be not negligible. This can be assessed qualitatively considering Fig. 10. The arrows point to some areas in which false alarms in the SVM detection have been found by visual inspection. A Google Earth patch of these areas is also reported, together with the close-up taken from the Level-1α product itself. Moreover, a close up from a classification difference map is attached. In this representation, points detected by both building index and SVM are depicted in green. Points detected only by the SVM are depicted in red. Points detected only by the building index are depicted in yellow. From the analysis of this pictures, it arises that SVM classifies as buildings also changing natural areas (see patch 1 in Fig. 10). In the same way, buildings depicted in patch 2 and 3 of Fig. 10 are arranged by SVM in structures with some landscape pixels. Conversely, the classification based on the building index has a negligible contribution of natural surfaces pixels.

By analyzing the results reported in TABLE III it arises that, as the urban density decreases, the performances of all the examined solutions tend to be equivalent. In fact, if we consider an electromagnetically isolated building in a stationary landscape (see Fig. 11), the obtained masks are very similar, regardless of the technique adopted. Conversely, in denser urban areas, the SVM tends to aggregate in the building mask also pixels that do not exhibit a backscattering among the highest in the scene. This results in a higher soil sealing index, but, as explained above, also in a higher probability of false alarms. This behavior could also be due to a non-optimal selection of the SVM training samples. A more refined solution could be to exploit the building index for the automatic selection of the training samples for the SVM.

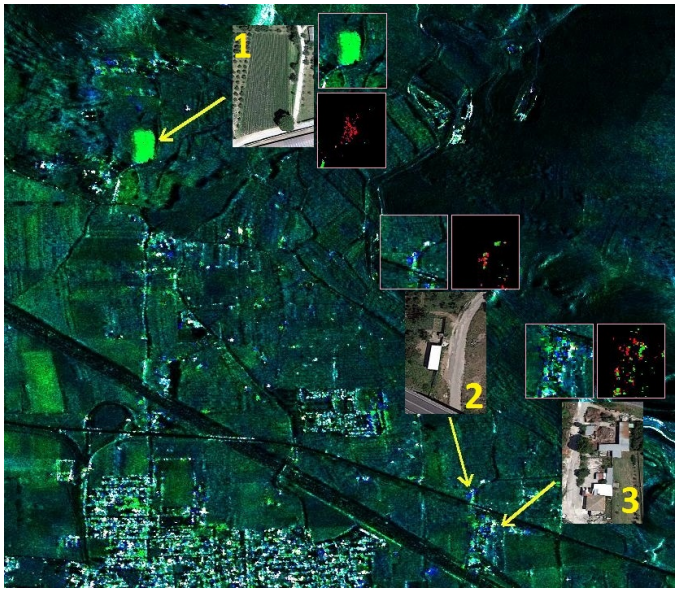
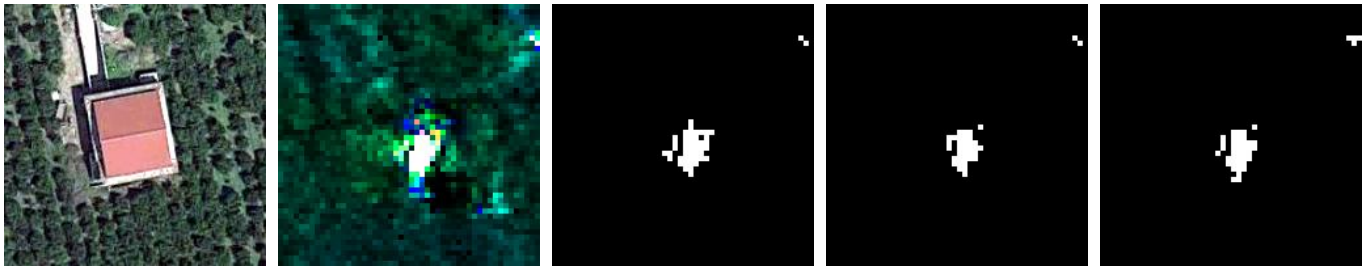


Fig. 10: Qualitative evaluation of false alarm using Google Earth patches and a classification difference map in which exclusive SVM detections are depicted in red, exclusive BI detections in yellow and SVM/BI detections in green. It arises that SVM (red pixels) tends to classify as buildings also changing landscape. Therefore a more significant probability of false alarms is expected than for BI detections.

As for the computational burden, the SVM ran in about 10 minutes on a scene of about 3000×5200 pixels using a 12 GB RAM machine with eight cores. SOM completed the job in more or less the same time. As for the proposed methodology, the feedback took about 5 minutes, while the building extraction process has a negligible impact on computational time. The proposed methodology introduces an advantage in computational time, that can be important when the amount of data to classify is high.

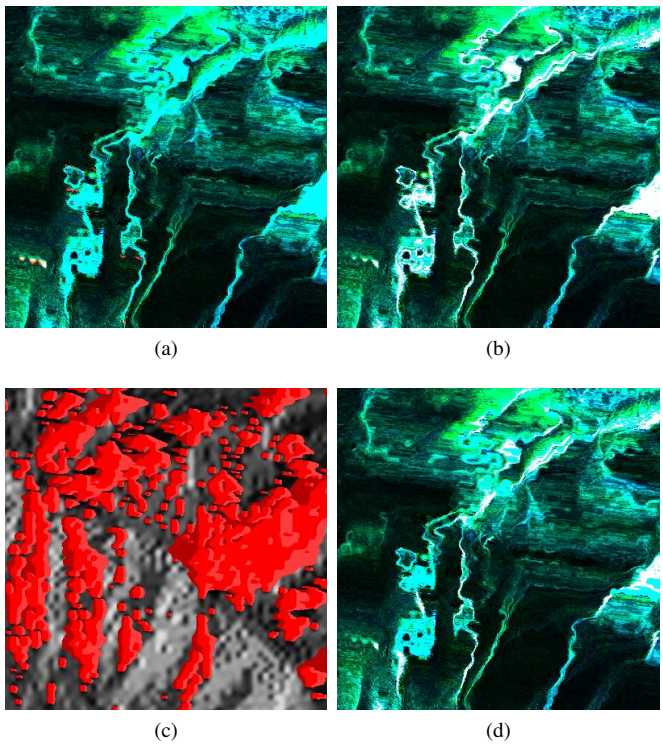
The last observation concerns layover. Using a very coarse clustering, it is likely that the texture evidence brings to an incorrect decision on layover features. This is more clear considering Fig. 12. In Fig. 12a we show the Level-1α product before the feedback. Here, layover is correctly represented, since we have no response of the red band (i.e. of the interferometric coherence). The texture evidence action makes this feature to appear as a bright target, thus introducing a confusion with the built-up class (see Fig. 12b). This problem can be solved using topographic corrections before building Level-1α products or by applying a layover mask in post-processing. We opted for this choice. We used



(a) (b) (c) (d) (e)

Fig. 11: Electromagnetically isolated building in a stable landscape: (a) Google Earth view, (b) Level-1α view. Building mask extracted (c) using the building index and the parameters reported in TABLE I, (d) using SOM and (e) using SVM.

a 20 meters resolution DEM for generating the layover mask depicted in Fig. 12c. Qualitatively, the masking operation applied to the Level-1α product after the feedback algorithm mitigates the confusion between layover (which should be displayed in cyan) and the built-up class, as showed in Fig. 12c.



(a) (b) (c) (d)

Fig. 12: Layover treatment: Level-1α product (a) before and (b) after the feedback application. (c) Layover mask. (d) Level-1α product after the application of feedback and layover masking.

Quantitatively, the influence of layover in the soil sealing index calculated through Equation (8) on the Level-1α product can be found in the class "Forests" of TABLE III. In the urban atlas layer, this class is mainly located in the upper left part of our study area (see Fig. 9, where the presence of a relevant topography makes layover to occur. The application of the proposed feedback causes the soil sealing to increase from 0.06% (before the feedback) to 2.2% due to layover features included in the computation. The application of the layover mask allows to reduce this value to 1.14%. Better results could be using a finer resolution DEM.

A. *Sensitivity to parameters*

The study of the sensitivity of the proposed method with respect to the parameters reported in TABLE I concerning the fuzzification of the input Level-1α product is presented.

As a general comment, if these parameters are increased, the curves plotted in Fig. 4 move to the right and the reflectivity requirements for the detection are reinforced. Vice-versa, if the parameters are decreased, the curves in Fig. 4 move to the left, slackening the requirements of reflectivity for the detection. Operatively, the positioning of the curves in the plan digital-number/fuzzy membership affects: i) the dimension of the window in which the coherence is calculated and ii) the activation of the texture evidence.

Indeed, the variation of coherence window dimension is expected to give negligible variations in the detection, provided that the scheme in TABLE I is maintained. In fact, if the variation of the parameters is small, a variation of category is likely to occur only for targets placed at the borders of the fuzzy sets. As an example, if a change from the category "small window" to "medium window" occurs, the coherence window dimension passes from 5 pixels to 7 pixels, therefore the variation of the estimated coherence value is expected to be small. In the same way, if we a have a change from the category "medium window" to "large window", this would affect targets whose characteristics are likely to be non-coherent; therefore, the passage from the 11

TABLE IV: Summary of the obtained results. The variation of the parameters reported in TABLE I for the fuzzification of the input Level-1α product does not affect significantly the result of the building extraction application.

Urban atlas category	BI SL (%)	BI (+10%) SL (%)	BI (+5%) SL (%)	BI (-10%) SL (%)	BI (-5%) SL (%)	SOM SL (%)	SVM SL (%)
Continuous Urban Fabric	24.83	23.75	24.72	28.23	26.98	24.23	31.11
Discontinuous Dense Urban Fabric	20.54	19.55	20.43	23.42	22.34	20.09	26.09
Discontinuous Medium Density Urban Fabric	11.04	10.09	10.55	12.29	11.69	11	14.61
Discontinuous Low Density Urban Fabric	5.59	4.9	5.13	5.95	5.65	5.55	7.49
Discontinuous Very Low Density Urban Fabric	3.35	2.79	2.89	3.49	3.21	3.05	4.57
Isolated structures	3.96	3.41	3.59	4.2	3.99	3.96	5.35
Railways and associated land	4.61	3.5	3.83	5.11	4.68	3.5	6.44

pixels window to the 51 pixels window would be inconsequential.

The most important effect of a variation of the fuzzy set is expected on the texture evidence. In fact, if the reflectivity attribute changes from "high" to "medium" even in only one acquisition, then the texture evidence is not activated. This can slightly affect the detection.

In order to give a measure of such variation, we repeated the whole processing chain varying the parameters reported in TABLE I concerning the fuzzification of the input Level-1α product of +5%, +10%, -5% and -10%. The result of this experiment are reported in TABLE IV, together with a recap of the previously obtained results for a comparison. For brevity, only the classes of the Urban Atlas most relevant with the building extraction application have been reported in the table.

As expected, a higher percentage of soil sealing is obtained decreasing the parameters reported in TABLE I. The lower the parameters, the higher the soil sealing. However, a higher probability of false alarms is expected, since the texture evidence is activated for less reflective targets. Vice-versa, a lower percentage of soil sealing is obtained by increasing those parameters. The higher the parameters, the lower the soil sealing, as well as the probability of false alarm.

Anyway, if the variation of the parameters is small (in the order of $\pm 5\%$), the detection rate is rather stable. In fact, fuzzy variables allow for an effective modeling of the uncertainty. This makes the technique scarcely sensitive to the parameters of TABLE I, even if significant variations $(\pm 10\%)$ of selected values are applied.

VI. Conclusions

In this paper, we introduced a technique for enhancing the information content of the red band of a Level-1α product. This band is usually reserved to the interferometric coherence information and therefore it is aimed at the enhancement of targets characterized by high phase stability over the time, such as built-up features. The proposed technique is devoted at enhancing the performances of the coherence estimator through an adaptive selection of the moving window dimension. To this end, the input Level-1α product is used to generate an *a priori* knowledge which is used to build the expert system for the choice of the coherence window dimension through fuzzy rules.

Texture evidence is used for supporting the enhancement of built-up features where the feedback fails due to decorrelation. We proposed the use of the Nagao-Matsuyama rule for improving the informative content of the red band of the refined Level-1α product in areas characterized by high backscattering and texture.

We assessed the performance of the proposed algorithm qualitatively and quantitatively. Qualitatively, the visual experience of the operator is enhanced by the feedback application. This makes easier to detect the urban area through the enhancement of its built-up features. Quantitatively, we used the Urban Atlas layer for comparing the soil sealing measured on the Level-1α product (before and after the feedback) with data provided in the reference layer description. We found that the feedback application significantly increased the number of detected pixels belonging to the built-up class for all the relevant classes. In particular, it is appreciable the improvement obtained on railways and their associated land. the obtained results, in terms of

measured soil sealing, are congruent with the data provided by the Urban Atlas, compatibly with SAR imaging characteristics and classes definition.

The obtained results were compared with those retrieved using two machine learning techniques such as SOMs and SVMs. The comparison showed that the proposed methodology, provided the enhancement of the informative content of the input Level-1α product, can represent a good and user-oriented alternative to more refined techniques. In fact, a simple band product can be understood and managed even by non-expert users, while SOM and SVM require a higher expertise and a certain confidence with machine learning to be used.

ACKNOWLEDGMENTS

The authors thank the Italian Aerospace Research Center (CIRA) for providing the COSMO-SKyMed dataset of Castel Volturno. This study was done in collaboration with Ansaldo STS under the aegis of the MODISTA project.

REFERENCES

[1] M. Datcu and K. Seidel, "Human-Centered Concepts for Exploration and Understanding of Earth Observation Images," *IEEE Trans. Geosci. Remote Sens.*, vol. 43, no. 3, pp. 52–59, 2005.

[2] M. Sonka, V. Hlavac, and R. Boyle, *Image Processing, Analysis and Machine Vision*. Pacific Grove, CA, 93950, USA: Brooks/Cole Publishing Company, 1999.

[3] V. Madhok and D. A. Landgrebe, "A Process Model for Remote Sensing Data Analysis," *IEEE Trans. Geosci. Remote Sens.*, vol. 40, no. 3, pp. 680–686, 2002.

[4] 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.

[5] D. Amitrano, F. Cecinati, G. Di Martino, A. Iodice, D. Riccio, and G. Ruello, "Sentinel-1 Multitemporal SAR Products," in *IEEE Int. Geosci. Remote Sens. Symp.*, 2015, pp. 3973–3976.

[6] D. Amitrano, G. Di Martino, A. Iodice, D. Riccio, and G. Ruello, "An end-user-oriented framework for the classification of multitemporal SAR images," *Int. J. Remote Sens.*, vol. 37, no. 1, pp. 248–261, 2016.

[7] R. F. Hanssen, *Radar Interferometry - Data Interpretation and Error Analysis*. Dordrecht: Kluwer Academic Publishers, 2001.

[8] M. S. Seymour and I. G. Cumming, "Maximum likelyhood estimation for sar interferometry," in *IEEE Int. Geosci. Remote Sens. Symp-*, 1994, pp. 2272–2275.

[9] R. Touzi and A. Lopes, "Statistics of the Stokes Parameters and of the Complex Coherence Parameters in One-Look and Multilook Speckle Fields," *IEEE Trans. Geosci. Remote Sens.*, vol. 34, no. 2, pp. 519–531, 1996.

[10] R. J. A. Tough, D. Blacknell, and S. Quegan, "A statistical description of polarimetric and interferometric synthetic aperture radar," *Proc. R. Soc. Lond. A*, vol. 449, pp. 567–589, 1995.

[11] M. Arrigoni, D. D'Aria, and A. Monti Guarnieri, "Space-adaptive coherence estimation," in *ESA FRINGE Workshop*, 2005.

[12] O. J. Tobias and R. Seara, "Image segmentation by histogram thresholding using fuzzy sets," *IEEE Trans. Image Process.*, vol. 11, no. 12, pp. 1457–1465, 2002.

[13] T. Matsuyama and V. S.-H. Hwang, *SIGMA - A Knowledge-Based Aerial Image Understanding System*. New York: Plenum Press, 1990.

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

[15] D. Marr and E. Hildreth, "Theory of edge detection," *Proceedings of the Royal Society B*, vol. 207, no. 1167, pp. 187–217, 1980.

[16] H. Chaabouni-Chouayakh and M. Datcu, "Backscattering and Statistical Information Fusion for Urban Area Mapping Using TerraSAR-X Data," *IEEE J. Sel. Topics Appl. Earh Obs.*, vol. 3, no. 4, pp. 718–730, 2010.

[17] A. Santinig and P. Gamba, "Combining SAR-Based and Multispectral-Based Extractions to Map Urban Areas at Multiple Spatial Resolutions," *IEEE Geosci. Remote Sens. Mag.*, vol. 3, no. 3, pp. 100–112, 2015.

[18] A. Ferro, D. Brunner, and L. Bruzzone, "Automatic Detection and Reconstruction of Building Radar Footprints From Single VHR SAR Images," *IEEE Trans. Geosci. Remote Sens.*, vol. 51, no. 2, pp. 935–952, 2013.

[19] C. Marin, F. Bovolo, and L. Bruzzone, "Building Change Detection in Multitemporal Very High Resolution SAR Images," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 3, pp. 2664–2682, 2015.

[20] H. Chaabouni-Chouayakh and M. Datcu, "Coarse-to-Fine Approach for Urban Area Interpretation Using TerraSAR-X Data," *IEEE Geosci. Remote Sens. Lett.*, vol. 7, no. 1, pp. 78–82, 2010.

[21] A. Voisin, V. A. Krylov, G. Moser, S. B. Serpico, and J. Zerubia, "Classication of Very High Resolution SAR Images of Urban Areas Using Copulas and Texture in a Hierarchical Markov Random Field Model," *IEEE Geosci. Remote Sens. Lett.*, vol. 10, no. 1, pp. 96–100, 2013.

[22] R. Gaetano, D. Amitrano, G. Masi, G. Poggi, A. Verdoliva, G. Ruello, and G. Scarpa, "Exploration of Multitemporal COSMO-SkyMed Data Via Tree-Structured MRF Segmentation," *IEEE J. Sel. Topics Appl. Earth Observ.*, vol. 7, no. 7, pp. 2763–2775, 2014.

[23] D. Amitrano, G. Di Martino, A. Iodice, D. Riccio, and G. Ruello, "Urban Areas Enhancement in Multitemporal SAR RGB Images Through a Feedback System," in *IEEE Joint Urban Remote Sens. Event*, 2015.

[24] T. Kohonen, *Self-Organizing Maps*. Berlin, Heidelberg: Springer-Verlag, 2001.

[25] P. Törönen, M. Kolehmainen, G. Wong, and E. Castrén, "Analysis of gene expression data using self-organizing maps," *FEBS Lett.*, vol. 45, no. 2, pp. 142–146, 1999.

[26] P. T. Pearson and C. I. Cooper, "Using Self Organizing Maps to Analyze Demographics and Swing State Voting in the 2008 U.S. Presidential Election," in *Artificial Neural Networks in Pattern Recognition*, ser. Lecture Notes in Computer Science, N. Mana, F. Schwenker, and E. Trentin, Eds. Springer Berlin Heidelberg, 2012, vol. 7477, pp. 201–212.

[27] B. C. Hewitson and R. G. Crane, "Self-organizing maps: applications to synoptic climatology," *Climate Res.*, vol. 22, pp. 13–26, 2002.

[28] H. Yuan, C. F. Van Der Wiele, and S. Khorram, "An Automated Artificial Neural Network System for Land Use/Land Cover Classification from Landsat TM Imagery," *Remote Sens.*, vol. 1, pp. 243–265, 2009.

[29] Y. Ito and S. Omatu, "Polarimetric SAR data classification using competitive neural networks," *Int. J. Remote Sens.*, vol. 19, no. 14, pp. 2265–2684, 1998.

[30] S. Skakun, "A Neural Network Approach to Flood Mapping Using Satellite Imagery," *Computing and Informatics*, vol. 29, pp. 1013–1024, 2010.

[31] S. Patra and L. Bruzzone, "A Novel SOM-SVM-Based Active Learning Technique for Remote Sensing Image Classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 11, pp. 606–616, 2014.

[32] M. L. Gonçalves, M. L. A. Netto, J. A. F. Costa, and J. Zulio Júnior, "An unsupervised method of classifying remotely sensed images using Kohonen self-organizing maps and agglomerative hierarchical clustering methods," *Int. J. Remote Sens.*, vol. 29, no. 11, pp. 3171–3207, 2008.

[33] Y. Hara, R. G. Atkins, R. T. Shin, J. A. Kong, S. H. Yueh, and R. Kwok, "Application of neural networks for sea ice classification in polarimetric SAR images," *IEEE Trans. Geosci. Remote Sens.*, vol. 33, no. 3, pp. 740–748, 1995.

[34] C. J. Burges, "A tutorial on support vector machines for pattern recognition," in *Data mining and knowledge discovery*, U. Fayyad, Ed. Norwell, MA, U.S.A.: Kluwer Academic Publishers, 1998.

[35] A. Farag, R. Mohamed, and A. El-Baz, "A unified framework for map estimation in remote sensing image segmentation," *IEEE Trans. Geosci. Remote Sens.*, vol. 43, no. 7, pp. 1617–1634, 2005.

[36] G. Camps-Valls and L. Bruzzone, "Kernel-Based Methods for Hyperspectral Image Classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 43, no. 6, pp. 1351–1362, 2005.

[37] European Environment Agency, "Mapping Guide for a European Urban Atlas," European Commission, Tech. Rep., 2011.