ABSTRACT

Monitoring urban areas at a regional scale, and even at a global scale, has become an increasingly important topic in the last decades. New Earth Observation (EO) platforms provide remotely-sensed imagery that allows the monitoring of urban expansion with the required spatial and temporal resolutions. Currently, we can take advantage of the combination of sensors present in the ESA ENVISAT platform, which offers simultaneous acquisition of MERIS (MEdium Resolution Imaging Spectrometer) and ASAR (Advanced Synthetic Aperture Radar) sensors.

The aim of this work is to demonstrate the capabilities of MERIS and ASAR data to map urban areas at a regional scale. Since MERIS FR works with a 300 m pixel size, this application focuses on obtaining high accuracy at a regional scale, providing the basis for a low-cost highly automated pan-European service dedicated to identifying and monitoring urban areas, for instance, of EU25. For this purpose, a two-stage classification scheme based on an unsupervised approach is proposed, which allows us to introduce supervised information about the class of interest without an additional sample labelling.

Key words: remote sensing, urban area monitoring, multispectral-SAR data fusion, partially supervised classification, hierarchical clustering.

1. INTRODUCTION

Monitoring urban areas at a regional scale, and even at a global scale, has become an increasingly important topic in the last decades in order to keep track of the loss of natural areas due to urban development. Earth Observation (EO) using remotely-sensed imagery is a relatively new tool to monitor urban growth. The combined use of optical and Synthetic Aperture Radar (SAR) data, interferometric SAR processing, and development of new pattern recognition techniques, are expected to enable an accurate detection of the urban tissue. Moreover, the multi-temporal data acquisition allows the monitoring of urban expansion [1].

The information retrieved by SAR and optical sensors differs to a great extent one from another. On the one hand, a multispectral image allows the reconstruction of the energy radiated by the Earth’s surface throughout the visible and infrared ranges of the electromagnetic spectrum [2]. On the other hand, SAR complex images provide a measurement of the changes that microwaves suffer in amplitude and phase when they interact with the Earth’s surface [3]. Therefore, SAR provides information that could not be obtained from optical sensors: SAR images are related to the dielectric properties of the target returning, are sensitive to microtextures, and microwave radiation penetrates the soil and canopies depending on frequency and polarization. In repeat-pass SAR interferometry, the interferometric phase measures distance to the targets, and the correlation between the complex SAR image pair, known as coherence, measures the temporal stability. In addition, SAR images overcome two of the main drawbacks of optical sensors: they are illumination-independent since they come from active sensors, and weather-independent since the used wavelength is almost unaffected by clouds, fog, rain, etc. However, the coherent nature of the microwave signal gives rise to a phenomenon called speckle, which can be modelled as multiplicative random noise [4], and tends to mask the back-scattering characteristics of the observed objects (granular noise). As a consequence, the coherent image statistics of SAR data will be dominated by the presence of speckle. In this context, images of different nature, and hence with different properties, can be used to develop an urban methodology properly designed for the optimal exploitation of SAR and multispectral images.

Moreover, in this key operational domain, it is not interesting to obtain an exhaustive map accounting for all the thematic classes present in an area, but to
produce an automatic classification of ‘Urban/Non-Urban’. This binary information is of great help and extended use for European policy bodies such EEA or DG-Regio in different aspects such as environmental monitoring, soil protection and regional cohesion policies and spatial planning. From this practical point of view, several methods have been proposed using the denomination of partially supervised or partially unsupervised classification. Most of the partially supervised methods [5, 6, 7, 8, 9] are based on the Expectation-Maximization (EM) algorithm and they assume that the user knows the probability density function (pdf) of the class of interest in the basis of a previous image of the same area or a specific training set of the class of interest. But this information is not always available, and sometimes the user only has some notions about the characteristics that presents the class of interest.

In particular, this paper analyzes the joint use of the MEduim Resolution Imaging Spectrometer (MERIS) and the Advanced Synthetic Aperture Radar (ASAR) images to characterize and monitor urban areas. For this purpose, a two-stage classification scheme based on an one-class partially supervised approach is proposed. The method use as inputs a set features extracted from both data in order to optimize the ‘Urban/Non-Urban’ separability in the input space, and it allows us to introduce supervised information (at least in one reliable feature) about the class of interest without an additional sample labeling.

2. MATERIAL

Data used in this work were obtained under framework of the ESA Category-1 project (C1P-ID2489) titled “Development of an Specialized Classification System for Urban Monitoring at Regional Scale Based on ASAR and MERIS data”. The test site was Naples (Italy), where images from ASAR and MERIS sensors were acquired in 2003 (Fig. 1). In particular, images used in this work are three multi-temporal ASAR Single Complex Look (SCL) images acquired at June, July, and August; and two MERIS Full Resolution (FR) Level 1b images acquired at July and August. The ASAR pairs were selected with small perpendicular baselines in order to obtain the interferometric coherence from each complex SAR image pair. In addition, the CORINE Land Cover 2000 (CLC2000) classification map is used as a reference to estimate the classification accuracy over the whole test area (2000×3000 pixels). The nomenclature of the CORINE land cover classes [10] presents three levels, which are subdivided in 5, 15, and 44 headings respectively. However, ‘Urban’ is a very heterogeneous land cover class and was not possible to define a ground truth using only one heading. The CLC2000 Level 1 is divided in: Artificial surfaces; Agricultural areas; Forest and semi natural areas; Wetlands; and Water bodies. The most suited heading to define urban areas is Artificial surfaces but in this work, when building the ground truth, we excluded one of its subclasses (Artificial non-agricultural vegetated areas) in order to consider only man-made artificial surfaces (Fig. 1).
3. METHODOLOGY

3.1. Preprocessing of MERIS/ASAR data

Since images come from different sensors, the first step is to perform a specific processing and conditioning of optical and SAR data, and to co-register all images. ESA ASAR SCL products were geocoded with the Envisat ASAR Geocoding System (EGEO) of German Aerospace Center (DLR) [11]:

- Slant range (time) to ground range (spatial) projection was performed in order to obtain a five-azimuth-looks interferograms and amplitude images, in which we increased the number of looks up to five by averaging five times in the azimuth direction in order to have approximately square pixels and to reduce speckle.

- Images were Enhanced Ellipsoid Corrected (EEC) using the "GLOBE" DEM in order to correct the topographic effects in SAR images over rugged terrains in two different ways: (i) subtracting the topographic phase of the interferometric phase in the coherence estimation; and (ii) normalizing the backscattering intensity for the true pixel size to reduce slope effects.

- Resulting amplitude images and a Lay-er&Shadow mask were provided with a resolution of 12.5 m.

The interferometric coherence was computed using the ESA BEST software, and coregistered and geocoded with the same parameters than the amplitude images. Finally, processed images were compared with a SRTM-X SAR amplitude grid, obtaining an horizontal registration error (RMSE) of about 20m, which potentially enables good urban classification ability (Fig. 2). On the other hand, MERIS images were projected in the same projection than the ASAR images using the BEAM software (up-scaling from 300m to 12.5m using a nearest-neighbor interpolation), and a refinement of the coregistration of both sensors was performed using ground control points (GCP). Additionally, MERIS FR Level 1b products contain TOA radiance. Therefore, in order to obtain correctly surface reflectance, images were atmospherically corrected using the method presented in [12].

3.2. Feature extraction in urban monitoring

The proposed classification method is mainly unsupervised so natural data clusters might not be related to the underlying and potential discriminative information between ‘Urban/Non-Urban’ areas. If we include features that are not related to this information, classification results could be wrong or misleading. Therefore, analysis of the optical and SAR data characteristics over urban areas, and definition and extraction of discriminant features are mandatory when using unsupervised or knowledge-based methods.

- Due to the reflectivity and angular structure of buildings, bridges, and other man-made objects, these targets tend to behave as corner reflectors and show up as bright spots in a SAR image. Therefore, in the SAR intensity images, the urban areas are characterized by very bright reflections (given by the presence of planes at 90 angles) which are maintained in time and under varying angles [13].

- In repeat-pass interferometry, when the satellite is on an exact repeat orbit–baseline zero–, and the phase contribution due to topography (DEM) is also removed, the absolute value of coherence provides information on measurement stability over time. Therefore, since urban areas do not have strong and fast changes, they should lead to a high coherence [14].

- The information contained in multispectral images makes possible the characterization and
identification of the observed materials from their spectral signature. In particular, urban areas present no chlorophyll absorption, like soil and dry organic matter, with under-developed red edge, and absorption features of iron bearing minerals [15].

A previous analysis, attending to the SAR and optical data information content over urban areas, was performed in [16, 17] to find a set of representative features to our problem. In this study, the most relevant and robust features were (Fig. 3):

1. A spatially filtered coherence map ($Co$): computed from the interferometric pairs of Single Complex Look SAR images by application-oriented spatial filtering based on previous knowledge and optimized to differentiate urban areas.

2. A spatially filtered intensity ($In$): obtained from the multi-temporal log-transformed SAR intensities filtered to reduce the speckle and maximize the value of local areas with high values.

3. The Normalized Difference Vegetation Index ($NDVI$): obtained from MERIS bands 8 and 13.

In addition, we can take advantage of a priori knowledge about the multispectral and SAR data characteristics. Before classification was conducted, we exploited the multispectral information by masking covers without any relation with our study like water bodies (using ratios between blue and NIR bands due to strong water absorption over 800 nm). Later, during the classification, we masked cloudy areas where multispectral information is not reliable and only SAR data might be used (cloud mask based on the brightness and whiteness of the spectrum and the atmospheric absorption of $O_2$ and $H_2O$ corrected by the DEM [18]). On the other hand, we used the SAR Layover&Shadow mask to identify areas where SAR information is not reliable and only multispectral data might be used.

### 3.3. Partially Supervised Hierarchical Clustering

From the unsupervised point of view, agglomerative hierarchical clustering should be an intuitive solution to our problem since it starts from all the data samples and iteratively joins the most similar samples until only two clusters remain (‘Urban/Non-Urban’). However, it suffers some drawbacks: the computational cost is higher; it does not work well if the classes are overlapped, and of course, it can not detect the class of interest by itself. In order to alleviate these limitations we have adopted a two-stage classification scheme, proposed by the authors in [19], which allows the operator to introduce some supervised information in order to define the class of interest without an additional sample labeling (the scheme plotted in Fig. 4 gives an intuitive idea of the partially supervised hierarchical clustering).

- The first stage of the process consists in an initial clustering of the image using the Self-Organizing Map (SOM) algorithm [20] in order...
to obtain a model containing all different clusters appearing in the scene. The objective of this stage is to found a reduced number of prototype vectors (units of SOM) that represent the data set. The hierarchical algorithm can be applied over this set of prototypes with an important reduction of noise and computational cost.

- The second stage consists of the partially supervised hierarchical joint of clusters, where we modify the employed criterion of similarity by introducing fuzzy membership functions (FMF) that make use of a priori information about the class of interest. In each step, the most similar clusters that fulfill the restrictions imposed by the FMF are joined, and a binary label is assigned to this new cluster.

At the end of the process, we find two clusters: our Class of Interest (‘Urban’) and the remaining classes. The two-stage clustering process assigns the most accurate possible set of labels to the prototypes of SOM. Therefore, our classifier has generated a non-linear boundary between the Class of Interest and the rest of the image. The main assumption of this method is that the user knows the range or the threshold values, of at least one feature, for the class of interest, which is not a hard assumption in common remote sensing applications. This algorithm is specially well suited to urban monitoring since it fulfills all requisites of this application: (i) urban monitoring is a clear example of the single class of interest problem, (ii) urban areas change from place to place so it is difficult to know the pdf of the ‘urban class’, and (iii) the multi-source approach –SAR and multispectral– requires the use of a non-parametric approach to perform the data fusion at the feature level.

4. RESULTS

4.1. Classification Results

We compare our one-class partially supervised approach in four different scenarios, i.e. to use different sets of features in order to test the dependence of the urban classification accuracy on the input dimension when including features less related with the ‘Urban/Non-Urban’ differences. The four scenarios are defined by the employed set of features used in each case: (i) the NDVI; (ii) the coherence map Co; (iii) NDVI and Co; and (iv) all the available features (NDVI, Co, the multitemporal SAR intensities In, and the MERIS bands). In addition, for each one of these scenarios, we include information about the urban class only in one or two reliable features in order to see the usefulness of the proposed approach: (i) the NDVI; (ii) the coherence map Co; and (iii) both NDVI and Co. In all cases, the supervised information consists in threshold values based on previous knowledge about the problem: NDVI < 0.3 since urban areas do not present vegetation trends (neither chlorophyll absorption or high NIR reflectance); and Co > 0.3 since urban areas present stability over time that leads to a high coherence.

Table 1 summarizes the overall classification accuracy (CA) when different sets of features are used. When considering only the two most significant extracted features, NDVI provides better CA than coherence. When adding more features, less related to information characteristic of urban areas, results should get worse with an unsupervised method. However, the partially supervised method (constraints imposed by the FMFs) finds the class of interest, and takes advantage of the discriminative content of the other features. In particular, the classification accuracy provided by the method rises up to 81% when including supervised information only for the NDVI (first row) or the Co (second row).

This fact can be explained because, in this work,
we combine two different data sources with different spatial resolutions with a stacked vector approach. Therefore, depending on the feature used to include some supervised information, there is a trade off between the quality of the feature and the original scale of its source. In Fig. 5, analyzing the classification map generated from NDVI and from Co, one can see different classification patterns that are driven by their different original scales: the NDVI-based classification shows clear square boundaries since original MERIS pixel size is 300 m; while the Co-based classification presents a noisy pattern due to its coherent nature. For this reason, by comparing separately NDVI and Co, NDVI has an a priori advantage over Co since implicitly takes into account the spatial distribution of urban class (large areas rather than isolated pixels). On the other hand, when using the information of NDVI or Co together with all the features, Co shows its better urban discriminative capabilities. In the following section, we perform a thoughtful analysis of the classification accuracy dependence on the spatial resolution.

4.2. Dependence of the Classification Accuracy on the Spatial Resolution

ASAR can achieve 12.5 m pixel size whereas MERIS products will be available at two spatial resolutions: Full Resolution (FR) with a resolution at sub-satellite point of 300 m and Reduced Resolution (RR) with a resolution at sub-satellite point of 1200 m. In this context, a main objective is to investigate the relationship between accuracy and scale (minimum mapping cell size) that allows getting an acceptable accuracy in the final map. The Full Resolution (FR) mode of MERIS provides a low resolution (300 m) for common urban applications, but the objective in this application is finding urban agglomerations rather than performing a detailed classification of different urban subclasses. We could afford to reach high accuracy at the expense of resolution because misclassification of one isolated building is not important but all urban agglomerations must be detected. In order to perform this analysis, we reduced the resolution of the images in two different ways (Fig. 6). On the one hand, we reduced the sensor images using the mean method, where the mean of the local sensor values provides the new pixel (average of the pixels in a spatial window). On the other hand, we reduced the ground truth (GT) using the vote method, in which the most probable local label is the new pixel label. In order to test the classification accuracy obtained with the new images, and by taking into account the information loss during the spatial resolution reduction, we computed the CA using both the original GT at the higher resolution and the series of downscaled GT at four increasing pixel sizes (Fig. 6).

Following the procedure described in section 4.1, we obtain the CA using four different sets of input features (NDVI, Co, NDVT&Co, and all available features) including information about the urban class only in the NDVI and Co. In Fig. 7 we represent CAs as a function of the spatial resolution (37.5, 150, 300, and 1000 m) computing the CA for both the downscaled GT (left) and the original full resolution GT (right). The following results can be extracted:
Figure 7. Classification accuracy (CA) as a function of the spatial resolution using four different sets of features as input and including information about the urban class only in the NDVI and Co. Left: CA computed from the downsampled GT. Right: CA computed from the original full-resolution GT (right).

- **Downscaled Ground Truth.** In all cases, when simulating reduced resolutions, classification accuracy increases at the expense of spatial details. The most significant improvement corresponds to the coherence since when downscaling the coherence by a local average the noise is reduced. In the case of the NDVI, note that the CA does not increase for pixel sizes lower than 300 m since it is the actual MERIS resolution.

- **Original Ground Truth.** In this case, all spatial details of the full resolution ground truth are taken into account, thus worsening results at the higher pixel size (1000 m). This is due to the fact that many errors are committed at boundaries between classes or at small areas and isolated pixels. However, good accuracies are obtained up to 300 m.

These results suggest that, when downscaling the features at 300 m, classification accuracy shows a good trade-off between the scale and the urban discrimination, thus MERIS FR would be valid for urban monitoring at regional scale. Note that, in both cases, results do not improve drastically when using more features. In fact, since the most relevant features are included first, at the optimum resolution of 300 m, better results are obtained using the NDVI and Co, rather than all features. This effect could be explained since we use a partially supervised method and features unrelated with urban areas can worsen results. It is worth noting that these results, extracted from the resolution analysis, can depend on the type of urban area, since the urban texture information could be different from one place to another, changing the optimum urban cell (optimum pixel size).

5. CONCLUSIONS

In this paper, we have analyzed the capabilities of ENVISAT MERIS and ASAR data to map urban areas. In a first stage, extraction of good features is performed in order to optimize the ‘Urban/Non-Urban’ separability in the input space. Later, we use a partially supervised classification technique that faces the problem of identifying only a class of interest accurately. The proposed method improves the classification accuracy of the class of interest using the minimum supervised information, and without an additional sample labelling (in our case simple threshold values for NDVI and Coherence). This avoids losing the discrimination of the class of interest when dealing with a higher number of unknown features (which could help or not).

The method has yielded good performance in the classification of urban and non-urban areas using MERIS and ASAR images of Naples. The best produced classification maps were obtained by using supervised information only in the coherence map, and presented a 84% classification accuracy. Finally, resolution of MERIS is a critical issue depending on the application requirements. Results have shown that MERIS FR can be used together with ASAR data to improve the classification accuracy without reducing drastically the classification effective spatial resolution. However, MERIS RR should not be capable of giving products at local or regional scale (< 1 km). Further work should consider including the spatial information in the classification process, and propose a data fusion approach to improve the combination of sensors of different resolutions.
ACKNOWLEDGMENTS

This paper has been partially supported by the Ministerio de Educaciòn y Ciencia of Spain, CICYT, under the project DATASAT, and by the European Space Agency (ESA) under ESA Category-1 project (C1P-ID2489) titled “Development of an Specialized Classification System for Urban Monitoring at Regional Scale Based on ASAR and MERIS data”. The authors wish to thank Dr. Martin Huber of the DLR by his support with the ASAR SCL geocoding.

REFERENCES


