Discriminating urban environments using multiscale texture and multiple SAR images

Article *in* International Journal of Remote Sensing · September 2006

DOI: 10.1080/01431160600557572

2 authors:

Fabio Dell'Acqua
University of Pavia

Paolo Gamba
University of Pavia

Discriminating urban environments using multi-scale texture and multiple SAR images

F. Dell'Acqua, and Paolo Gamba

The authors are with Dipartimento di Elettronica, Università di Pavia, Via Ferrata, 1, 1-27100 Pavia, Italy. This work is an enlarged version of the paper "Discriminating urban environments using multi-scale texture and multiple SAR images", presented at WARSD'03. The research was partially funded by ASI, contract I/R/177/02, and the Canadian Space Agency, under the Data Research User program.

Abstract

In this work the use of multi-scale textural features for urban satellite SAR image characterization is introduced. The multi-scale nature of urban environments requires indeed that no unique scale of analysis is considered. So, an accurate texture-based discrimination of land use/land cover classes needs for instance the computation of multi-scale textural features for a wide range of parameters of the co-occurrence algorithm.

The technique proposed in this paper shows how to reduce the full multi-scale feature set to a subset, the most suitable for classification using a fuzzy ARTMAP neural network. This is done by analyzing the relevance of each feature for this particular classifier by means of the Histogram Distance Index.

We validate the procedure by providing results of the classification of many satellite SAR data of the same urban test site. Results are encouraging. They show the potential of this technique for automatic extraction of the best texture and scale subset, suitable for efficient urban mapping using SAR satellite data.

Keywords

urban remote sensing, co-occurrence matrix, texture measures, fuzzy ARTMAP.

I. Introduction

The characterization of urban environments by means of satellite data is becoming more and more challenging because of the available ground spatial resolution of the sensors. Not always, however, a very high resolution is the solution to data interpretation problems. As a matter of fact, scales of urban features may be very different, and even for classification purposes, they depend on the level of land cover/land use one is seeking to discriminate (Woodkcock and Strahler, 1987). It is thus of paramount importance to find the best combination of data sets, feature manipulation techniques and interpretation/classification procedure for a given scale of interest. In this sense, there have been in recent literature a number of interesting papers on the use of different satellite images for different purposes: from very high resolution Ikonos and Quickbird images for precise building characterization and Digital Elevation Models (DEMs) definition (Ganas *et al.*, 2002), to SPOT and IRS data for discrimination of urban residential/industrial areas and tracking of urban sprawl (Shaban and Dikshit, 2002), to Landsat and ASTER data for multispectral land cover characterization and vegetation analysis (Small, 2001), finally to NOAA AVHRR thermal bands for detection of urban heat island effect (Gallo *et al.*, 1993).

In parallel, a more recent and somehow less known literature is trying to exploit the information in SAR images at the current coarse resolution, as it has been and is provided by ERS, RADARSAT-1 and ENVISAT satellites as well as the Shuttle Imaging Radar missions (from SRL-1 to SRTM). Radar data are known to be affected by problems associated with the coarseness of their spatial resolution and the side-looking nature of the sensor. There are, however, a number of satellites with SAR sensors planned to be operative in the next few years with fine resolution radars, from RADARSAT-2 to TerraSAR-X to the Cosmo/SkyMed constellation, while orientation-dependent problems may be overcome by combining different views of the same area (Dell'Acqua *et al.*, 2003), which is common today and will be more and more common with the new sensors and their short (less than 12 hours) revisit time.

On the other hand, one of the most interesting aspects of SAR data is that they provide information about scatterers in an area, and this has been used for many applications in the field of urban subsidence monitoring by differential interferometry (Ferretti *et al.*, 2000). This information is also important to discriminate among different parts of the same urban areas by recognizing different scattering patterns. This approach has been proposed in Dell'Acqua and Gamba (2003), where single date and multi-temporal ERS data have been analyzed to detect the city center, the residential areas, and the outer sparsely built zones. The idea is based on the use of co-occurrence matrix texture measures. A similar method has been proposed in Dekker (2001) using other textures for the same purpose. This work is aimed at a further investigation in the same field, i.e. the possibility to use different textures and/or different scales to improve the classification results.

The issue of scales in remote sensing images is indeed extremely interesting. Key references for this topic are Quattrochi and Goodchild (1997) and Marceau and Hay (1999). In particular, the latter paper introduces a very interesting connection among remote sensing and geographical entities. It is shown that the issue of scales in remotely sensed images may be seen as a modifiable areal unit problem. In this framework, the proposed methods may be reduced to optimal spatial zoning systems, approaches based on geographical entities, algorithms based on spatial statistics and procedures for analyzing relationships among results, variables and scales. Recently, the same problem has been addressed in Atkinson and Aplin (2004), showing that no unique scale is possible for characterizing objects within a remotely sensed image.

Since our problem is mainly to define which are the most useful textures for urban environment discrimination, we are mostly interested in the use of the scale factor in classification and interpretation algorithms. Among them, in Townsend and Justice (1988) the Fourier Transform has been employed to investigate the spatial frequency content and therefore the scale of an image. A similar approach has been proposed more recently in Chen and Blong (2003), but using a wavelet decomposition. The authors show that there is a function of the change in the statistics of the wavelet coefficients that has a maximum in correspondence with the scale of the features in the image. The methodology is quicker than FT, but also rougher, due to the logarithmic spatial resolution levels of the wavelet transform. A precise analysis of the scales for different classes in the same scene has been proposed also in Marceau *et al.* (1994). The authors suppose that the window width giving the minimal variance in the neighborhood of each pixel for the largest number of bands may be used as scale of the class to which the pixel belongs. It was found that, even for the same class, there are differences in the recognized scale, and this suggest the complexity of capturing this value. The approach is similar to the one in Woodcock and Strahler (1987), since the neighbourhood is used to compute local scale information through grey level variance, but calling for a multi-scale analysis. The need for multiple scales suggested different approaches. In Pesaresi and Benediktsson (2001), spatial features referring to very different scales are extracted using morphological operators. Alternatively, it was proposed in Hay *et al.* (2001), where one single image at fine spatial resolution is used to capture the objects that emerge at their characteristic scale.

In this paper we want to use multiple scale co-occurrence textural features to discriminate among different urban environments, following the approach delineated in Dell'Acqua and Gamba (2003). the enhancement of this paper is the use of multiple scales, with a training methodology to pick the more interesting ones for the particular classification problem.

II. Co-occurrence textural features for urban area characterization

Textural features in SAR imagery may help in highlighting the spatial patterns of backscatterers. A supervised clustering of these features reveals where buildings and other man-made objects gather in a similar way. So, residential areas with isolated scattering elements are quite different from town centers with many crowded backscatterers or even business district areas, where a few large buildings cause higher peaks in radar response.

Since the computation of co-occurrence measures has been carried out in a standard way, we will not discuss here the algorithm (Haralick *et al.*, 1988). We simply recall that it computes for each pixel of the original data a co-occurrence matrix, whose elements contain the probability of a joint occurrence of the two values i and j for a pixel pair in a window around the location of interest. Pairs are characterized by their distance and direction. The (square) window is instead defined by its width. Finally, the values are quantized and so the last parameter we need to introduce is the number of quantization levels.

It is easy to understand that many of these parameters are related to scale issues. The only independent one is the number of quantization levels, or equivalently the number of bits required to represent them. This number affects of course the computational load of the co-occurrence algorithm. So, the smaller, the faster. On the other hand, the fewer the bits, the larger the information lost. For optical images, for example, a linear quantization and a reduction to 64 levels (6 bits) has been shown to speed up the computation with negligible effects on segmentation results (Clausi, 2002). SAR images in urban areas have instead a grey level histogram with long tails because of strong scatterers, like some buildings are. They cause very bright spikes in the data, due to the dihedral backscattering mechanisms (Franceschetti *et al.*, 2002) and produce very large data values. A plain linear quantization does not work properly, and eight bits are likely to be more useful than 6. So, original amplitude floating point SAR data may be reduced to integers with 256 levels before texture extraction. This is done by means of a 2% clipping of the histogram, before applying a linear stretch.

Once the number of levels and the quantization law have been decided, the other parameters of the co-occurrence algorithm need to be carefully considered. In general, the distance and direction are pointed out as the most critical values. Direction is especially important when there is anisotropy in the texture we want to characterize (Barber and LeDrew, 1991). The distance provides an important way to discriminate among textures made up by identical, but differently spaced, elements. So, there are different evaluations of the best distance values, depending on the particular image, and a range of distances between 1 and 10 pixels is considered. Finally, and differently from these parameters, the window width is usually neglected. In Clausi (2002), for instance, there is a long dissertation about the best values for the previous parameters, but nothing is stated on window width. The reason is that in many image segmentation problems there is a need to care about it only in the boundary areas. Segments are considered as internally homogeneous.

In remote sensing images, and especially in satellite images at coarse resolution, we need to change this point of view. Only very compact patterns are visible, and sometimes we barely see them. For points inside the urban boundaries, distances longer than 1 pixel (or $\sqrt{2}$ pixels for diagonal directions) should be discarded. As for the direction, it is true that any remote sensing data has an anisotropic value distribution. This is a result of the anisotropy of artificial environments and the side-looking nature of the SAR sensor. So, the usual approach (Clausi, 2002) is to combine multiple directions and compute the mean value of the textural features in these directions. For urban areas, however, the anisotropy of each part of the scene builds up, in coarse resolution satellite images, to a general isotropy.

These comments bring us to the last parameter, the co-occurrence window width. Its value may be neglected in segmentation problems, when clear textures are to be separated. When large textured areas are to be discriminated, the larger the window used for co-occurrence computation, the more the statistical outliers in the data are filtered out. However, this cause problems on the boundaries, and depends on the dimensions of the textured areas. So, the choice of a single, large value for the window dimension is valid only for some, very particular, test images. In the test images used in this research, however, no clear texture segment can be seen. We must face a continuously and slowly changing pattern, whose boundaries are somehow loosely defined. These boundaries are more sharply delineated in very high resolution imagery, where the pixels are comparable in size with the basic elements of the urban landscape, buildings and trees. This explains the different choice in this work with respect to Karathanassi *et al.* (2000) and Zhang *et al.* (2003). This also stresses why the window width should be considered as the most important parameter in the following. In other words, in this work, the only procedure parameter that depends on the spatial scale of the urban environment is this width. It may be considered as global urban *spatial scale* of the patterns we are searching or as the *scale of stability* for the statistics that describe these patterns.

A. Textures and scales

One of the trivial results of Dell'Acqua and Gamba (2003) is that the *best* scale of texture measures (in our meaning of "window width") is related to the mean block dimension in the studied urban area. This is of course a very rough simplification of the problem, since each environment has its own "optimal" scale (Marceau *et al.*, 1994), and the value proposed in Dell'Acqua and Gamba (2003) is just the mean among all the scales of all the environments we are looking at. Different scales for different classes have to been considered when applying co-occurrence texture analysis to optical remote sensing images, as already shown in Marceau *et al.* (1990). So, the choice in Dell'Acqua and Gamba (2003) is a "sub-optimal" one and depends on the request to use one fixed scale for the whole scene.

The problem is that this value must be determined by a trial-and-error procedure and it does not work for any image of the same scene, even for a multi-temporal sequence of data from the same sensor. To exemplify, in Table I we show the overall accuracy value obtained for two different ASAR images on the same test area as a function of the window size. While it is clear that very different window sizes are not valuable for our classification, many values around our choice of a 21×21 window provide comparable or slightly better results. So, this value is just a mean of the best window sizes, that may be different from image to image, due to the radar noise or image orientation.

Another problem that affects the choice of the co-occurrence window width is that the block size is usually unknown, and may be different, for large towns, in different parts of the same urban area. Thus, we should consider, jointly or separately, different scales to prevent loss of information and reduce the misclassifications due to a wrong scale choice. A single scale for texture analysis reduces, indeed, the quality of the classification map at the border among different zones, since it discards information at other scales.

B. Multi-scale texture-based SAR segmentation

Unfortunately, if more scales are considered, a large number of bands should be analyzed. Starting for instance by a set of four texture measures Dell'Acqua and Gamba (2003), it is necessary to work on $4 \times n$ bands for n considered scales. This is a problem not only for the increased CPU-time required by the classifiers, but also for the size of the training set required by supervised procedures (Richards, 2003). A common way to reduce data dimensionality is to exploit feature reduction techniques, initially proposed for hyperspectral data, like Decision Boundary Feature Extraction (DBFE) or Discriminant Analysis Feature Extraction (DAFE) (Landgrebe, 2003).

We prefer a feature extraction procedure, based on the Histogram Distance Index (HDI) as a mean to detect the best feature subset. After this step, we are left with the most important textural features at different scales for the given data analysis problem, used as input to the classifier. To understand the procedure, we recall here the definition of HDI (Pesaresi, 2000), whose aim is to compare two probability density functions, $f(x)$ and $g(x)$, of the same random variable x, whose domain is labeled as D:

$$\text{HDI} = \left(\frac{1 - 2\sum_D \min\left(f(x), g(x)\right)}{\sum_D \left(f(x) + g(x)\right)} \right) \tag{1}$$

In this work, x is actually the vector $\vec{x} = (x_1, x_2, \ldots, x_M)$ containing the values of a subset of M out of N textural features, and D is the training set. Practically, HDI computes how much the clusters of training samples referring to different classes overlap in the multidimensional feature space, and no assumption outside of the range of values in the training set is done.

In this paper it is proposed to exploit HDI as a feature reduction tool in texture-base classification of satellite SAR images. Fig. 1 shows visually the sequence of processing steps. The method requires first to compute the co-occurrence matrix with 8 bit quantization, only one pixel distance in the 45° direction, but many window widths. The reasons behind these choices were explained in the previous section. Then, some textural features are extracted, i.e. mean, variance, entropy and dissimilarity, which have been shown in Dell'Acqua and Gamba (2003) as the most suited for urban SAR mapping. This large data set of textures is reduced using HDI. HDI values are computed for any subset of these features on a training set, and the "best" subset is determined by comparing the computed values and choosing the largest. Finally, we may classify the chosen feature set using the same training set in a supervised clustering procedure.

There are two main reasons why HDI is used for the feature reduction step. The first one is that HDI does not assume a precise model for the probability density functions $f(x)$ and $g(x)$. No normal distribution is considered, as in other separability indexes (Fukunaga, 1990). This property allows the algorithm to adapt to SAR data coming from different sources. SAR amplitude data have distributions strongly dependent on geometrical and radiometric properties of the sensors, and thus a flexible approach is mandatory.

The second reason is the tight connection between the methodology used by HDI to compute the separability among classes and the algorithm used by the subsequent classification step, which is a non-parametric classifier. In particular, the final classification step of the procedure in fig. 1 is based on a fuzzy ARTMAP neural network classifier developed in our group for multiband data analysis. The details of the classifiers may be found in Gamba and Dell'Acqua (2003) and are not reported here for sake of brevity. It may enough to say that fuzzy ARTMAP is a neural network based on the Adaptive Resonance Theory (ART) concepts (Mannan *et al.*, 1998). ART networks accept analog or digital inputs and try and find associations among them looking for the resonance of the input with one of the memories that they store. In the training phase, multi-band input data are associated with output classes, and build network memories. In the testing phase, for each input the most similar memory of the network is selected, and the corresponding output class is offered as a result.

In the fuzzy ARTMAP structure, the above mentioned memories are updated by new training patterns by means of a fuzzy AND operator. This fuzzy operator is essential a band-by-band *min* between the values of the training patterns, and translates the (normalized) input into a suitable internal set of ranges. Each range set delimits a *hypercube*, and one or more hypercubes are connected to a given output class. In a sense, the fuzzy ARTMP training phase builds a "decision tree" internal to the fuzzy ARTMAP structure, that provides a set of rules for the assignment of a new, unknown pattern to one of the known classes. The parameters of the ARTMAP network rule this mechanism, and determine the number and shape of these hypercubes. ρ parameters act as regularization factors in the segmentation of the hyperspace which is the internal data representation of the neural network, while β parameters act as resilience factors, determining how long a training input is going to influence a network memory. In other words, fuzzy ARTMAP neural networks translate input data into memories maintaining their finite support.

III. Results

The proposed methodology has been tested on a rather complete data set, comprising SAR images from most of current satellite sensors, all depicting the same urban area. Of course, this choice is an advantage, since the results that we are going to present and discuss in the following pages are comparable and allow a better understanding of the pros and cons of the sensors and their combination.

A. Test area and data set

For our experiments, we used SAR data on the test urban area of Pavia, Northern Italy. The data come from the ASAR sensor on board the ENVISAT satellite, from ERS-1/2 and from

RADARSAT-1. Dates, polarization, ground spatial resolution and ascending/descending modes for the complete data set are shown in Table II.

For the purpose of data analysis, a ground truth was built by visual interpretation of a very fine resolution optical satellite image of the same area, an Ikonos-2 panchromatic image, 1 m spatial ground resolution, recorded on July 2001. Hints for interpretation and for characterizing past dates came from the corresponding sector of the Technical Regional Map, which is a raster map with 50 cm resolution, based on aerial photogrammetric flights in the early ‘90. Since the data sets from ERS and ENVISAT have different spatial resolution than the RADARSAT-1 fine beam images, two copies of the same ground truth were provided, at 12.5 m and 7 m ground resolution, and all the SAR data, possibly after slant- to ground-range conversion, were coregistered to one of these two reference maps. The land use classes in the ground truth are vegetation (green), water (light blue), city center and high density urban areas (yellow), residential and medium density areas (red), suburban zones and sparse buildings (blue). The ground truth covers the whole urban area, and a part of its surroundings. Therefore, the pixel number for each class is different, ranging from 4,527 for the water to 99,804 for vegetation.

The training set was based on a small subset of the ground truth maps, with 150 to 200 pixel for each class, concentrated in areas whose land cover has not changed in the time period of the data. Samples of the area of interest depicted by the different sensors are provided in fig. 2, while in fig. 3 we show the ground truth and training set maps. Note that in all SAR images of fig. 2, despite the different backscattering behaviors, due to the viewing angles, it is always possible to visually discriminate the land use classes included in the ground truth.

B. ERS results

The ERS-1 image recorded on August 13, 1992 was used to analyze the effect of the parameters of the co-occurrence algorithm and to validate our choice of the window width as the most important parameter. To this aim, the fuzzy ARTMAP classifier was applied to a set of four textures (mean, variance, dissimilarity and entropy). the analysis was performed by varying only one parameter at a time, and evaluating the classification results in terms of overall and mean class producer’s accuracy; the reason for this latter selection of indexes will become apparent later in this paragraph. Moreover, since the classification results depend on the ρ and β parameters of the fuzzy ARTMAP network, many classifications were performed for a range of these parameters. In particular, ρ was changed from 0.5 to 0.7 with a step of 0.1, and β from 0.05 to 0.85 with a step of 0.05. For each parameter choice value the mean overall and producer's/user's accuracy values were computed, together with the corresponding standard deviation.

As argued in section II, distance and orientation (direction) do not cause significant differences in the results. Table III provides the two accuracy values used for performance monitoring. It shows, by considering four orientations (0°, 45°, 90°, and 135°) but the same distance (one pixel) and window width (21 pixels) that we have some changes in overall accuracy, but of very small extent with respect to mean producer's accuracy values. Very similar results, also shown in Table III, correspond to changes in the distance, from 1 to 3 to 5 pixel steps, with fixed direction (45°) and window width (21 pixels).

More interest deserves the fig. 4, where the kappa coefficient, as well as the overall and mean producer's accuracy values are plotted against a wide range of window widths. The bars in the plot have a length proportional to the standard deviation of accuracy values in our classification set, as explained above. The graphs show that apparently the best overall accuracy is obtained with very large width, around 35 or 37 pixels. These large window are characterized, however, by a high variability of the classification map accuracy, because of the errors in boundary areas. Instead, kappa coefficient and the mean producer's accuracy value have a different behavior, and their maximum is obtained for a larger range of window widths. In particular, around 23 to 25 pixels there are also the smaller standard deviation values, showing a good stability of the results.

The reason of this different behavior of the two overall and mean producer's accuracy values can be found in their nature. The mean producer's accuracy shows how much the land cover classes has been discriminated by the classifier, irrespectively of the number of pixels that belong to each class in the ground truth. The overall accuracy, instead, takes into account these numbers. So, since in our ground truth map the vegetation class is largely predominant, the overall accuracy mostly follows the producer's accuracy of this class. "Vegetation" means essentially "fields" around our urban test area. This explains why the best window width is larger than for urban classes. The mean size of these fields are larger than those of urban blocks.

So, limiting ourselves to urban land cover classes (this was the choice in Dell'Acqua and Gamba (2003)) or looking at the mean producer's accuracy (like we do here) we observe first that the window width is extremely important for a correct delineation of urban areas in the classification map. Moreover, fig. 4 confirms that a width in the range of 20 pixels is the preferred choice for SAR images with ERS-like spatial resolution.

C. ENVISAT results

After the discussion on co-occurrence parameters, it is still necessary to validate the multi-scale texture-based procedure. To this aim, we will use the ASAR images to show how the proposed procedure allows integrating different scales, and how we can manage to provide better results by using the HDI index for feature selection. To this aim, we compute the above mentioned eight textural features for the ASAR images recorded on December 8th, 2002, but for two different window widths, i. e. 21 and 25 pixels. These values have been found "a posteriori" as very good ones, as shown in Table I, but for different images this may not be true and, for this test, we assume we cannot determine "a priori" which is the best choice. We therefore consider 16 texture features (mean, entropy, variance, dissimilarity, contrast, correlation, second moment and homogeneity at the two scales) as input to the feature selection step, and compare all the subsets of four features, in order to find the most useful ones. The HDI values for all the subsets are shown in fig. 5. In this figure the three red lines identify three different combinations, representing small, medium, and large HDI values. As a matter of fact, the corresponding classifications provide similarly low, medium and high overall accuracy values, i. e. 21.3%, 45.6%, and 55.2%. The best result out of these three is depicted in fig. 6(a), while the corresponding confusion matrix is shown on top of Table IV.

Considering now the ascending image by ASAR, recorded on November 25th, 2002, we assume to have defined the set of mean, variance and dissimilarity as our required classification input and look for any combination of texture scales. We find that the set where all the measures are computed with a 25×25 window happens to be the best, immediately followed by the same texture set, but with the mean computed in a 21×21 window. Accordingly, the overall accuracy after classification are 55.4% and 55.0%, respectively. Of course, these values are unbearably low for any application. Their role here is just to prove experimentally the usefulness of the proposed procedure.

D. RADARSAT-1 results

The RADARSAT images allows studying the effect of the viewing angle on the proposed procedure. We are interested to understand how robust the method is with respect to different orientations of the SAR beam. We thus have had data collected on our urban test site, thanks to the Data Research User program of the Canadian Space Agency, corresponding to two fine beam modes, F2 ($39° - 41°$) and F4 ($43° - 46°$). This set was integrated with a standard beam mode image of the same area (S7, $45° - 49°$) and constitutes a sufficient test set for our purposes.

Applying the proposed procedure, and starting from the scale texture and scale set, different "optimal" subsets are obtained. A reason is that different viewing angles and/or spatial resolution stress different backscattering components. In turn, this changes the usefulness of input data sets for the subsequent fuzzy ARTMAP multi-band pattern classification. To simplify the scenario, and only to the aim of this section, overall and mean accuracy values are compared for the same texture subset. This subset is composed by mean, variance, entropy and dissimilarity, computed considering one pixel distance, 45° direction, and a fixed scale, corresponding to a 21×21 window for ERS data. The overall accuracy values as well as the producer's accuracy for the urban classes are plotted in fig. 7 against the beam angle, and show a strong dependence. As we expected, all the values increase with steeper angles, confirming that the observation of urban structures require incidence angles to be as small as possible. Moreover, we may observe that, as in all other results, the worst classification accuracy is found for the "residential area" class. This problem arises from the ambiguous definition of this urban environment, which has different characteristics in different parts of the same town.

E. Combination of more sensors

After the characterization of a single date/single sensor image, we applied the same procedure to multiple sensors data, seeking possible improvements in the results. We expect that a suitable combination of multi-temporal textural information is able to capture the urban environments to a better accuracy. Even for this situation the choice of the scales and images in our data set is driven by the HDI values, although in this case we found a looser correlation between the results of the classification process and the sequence of HDI values. This shows that the procedure depends, at least partially, on the characteristics of the sensors, and needs to be fine-tuned with this respect. To show an example of the results, in fig.6(b) we present the classification maps obtained by using the above mentioned textural features at the best scale, classifying jointly the 1992 ERS-1 image and the two 2002 ASAR images.

We note that the joint use of ENVISAT and ERS data is very successful, with an overall accuracy of 71.5%, much higher even than ERS data alone, which resulted in values around 62% (see above), and also higher than classifying jointly only the two ASAR images (65.4%). The big advantage of using ERS data comes perhaps from its better signal to noise ratio, but also from the multi-temporal pattern (ERS and ASAR data were sensed in different periods of the year). The result is that the residential area, where vegetation produces seasonal changes in the backscattered electromagnetic field, is much better recognized and the overall accuracy noteworthy increased. A look at Table IV, bottom matrix allows quantitatively understanding the point.

IV. Conclusions

The analysis proposed in this work shows that we may improve the characterization of an urban environment using multi-scale textural features extracted from satellite SAR data. We have shown a way to take into account these scales, the most relevant for the given task, in an automatic fashion. Moreover, the use of images coming from different sensors has been investigated.

Our future work will focus on a further refinement of the procedure, aimed at extracting with higher precision the boundaries of the urban areas. Moreover, we are considering a different ground truth more related to land use mapping. A future step will be to validate this new map. Finally, it is clearly mandatory to prove the usefulness of this procedure in view of its application to more urban test sites, exploiting the experience we gained on our data set.

Acknowledgments

The authors wish to thank G. Trianni and M. Tosi for evaluating the numerical results and for the help in preparing the figures and tables of this manuscript.

REFERENCES

ATKINSON, P. M., and APLIN, P., 2004, Spatial variation in land cover and choice of spatial resolution for remote sensing. *International Journal of Remote Sensing*, **18**(6), 1351-1364.

BARBER, D. G., and LEDREW, E. F., 2001.SAR sea ice discrimination using texture statistics: a multivariate approach. *Photogrammetric Engineering and Remote Sensing*, **57**(4), 385-395.

CHEN K., and BLONG, R., 2003. Identifying the characteristic scale of scene variation in fine spatial resolution imagery with wavelet transform-based sub-image statistics. *International Journal of Remote Sensing*, **24**(9), 1983-1989.

CLAUSI, D. A., 2002. An analysis of co-occurrence texture statistics as a function of grey level quantization,.*Canadian Journal of Remote Sensing*, **28**(1), 45-62.

DEKKER, R. J., 2001. Texture analysis of urban areas in ERS SAR imagery for map updating. In *1st IEEE/ISPRS Joint Workshop on Data Fusion and Remote Sensing over Urban Areas,* Rome (Italy), 8-9 Nov. 2001.

DELL'ACQUA, F., and GAMBA, P., 2003. Texture-based characterization of urban environments on satellite SAR images. *IEEE Trans. on Geoscience and Remote Sensing,* **41**(1), 153-159.

DELL'ACQUA, F., GAMBA, P., LISINI, G., 2003. Improvements to urban area characterization using multitemporal and multiangle SAR images. *IEEE Trans. on Geoscience and Remote Sensing*, **41**(9), 1996-2004.

FERRETTI, A. PRATI, C., and ROCCA, F., 2000. Nonlinear Subsidence Rate Estimation Using Permanent Scatterers in Differential SAR Interferometry. *IEEE Trans. on Geoscience and Remote Sensing,* **38**(5), 2002-2212.

FRANCESCHETTI, G., IODICE, A., and RICCIO, D., 2002. A canonical problem in electromagnetic backscattering from buildings. *IEEE Trans. on Geoscience and Remote Sensing,* **40**(8), 1787 -1801.

FUKUNAGA, K., 1990. *Introduction to Statistical Pattern recognition.* 2^{nd} Edition, (San Diego, CA: Academic Press).

GALLO, K. P., McNab, A. L., Karl, T. R., BROWN, J. F., HOOD, J. J., and TARPLEY J. D., 1993. The use of NOAA AVHRR data for assessment of the urban heat island effect. *Journal of Applied Meteorology,* **32**, 899-908.

GAMBA, P., and DELL'ACQUA, F., 2003. Improved multiband urban classification using a neuro-fuzzy classifier, *International Journal of Remote Sensing,* **24**(4), 827-834.

GANAS, A. LAGIOS, E., and TZANNETOS, N. 2002. An investigation into the spatial accuracy of the IKONOS 2 orthoimagery within an urban environment. *International Journal of Remote Sensing,* **23**(17), 513-3519.

HARALICK, R. M., SHAMMUGAM, K., and DINSTEIN, I., 1988. Textural features for image classification. *IEEE Trans. Systems, Manufact., Cybernet.*, **3**(6), 610-621.

HAY, G. J., MARCEAU, D. J., BOUCHARD, A., and DUBE', P. , 2001. A multiscale framework for landscape analysis: Object-specific upscaling. *Landscape Ecology,* **16**, 471-490.

KARATHANASSI, V., IOSSIFIDIS, Ch., and ROKOS, D., 2000."A texture-based classification method for classifying built areas according to their density," *International Journal of Remote Sensing*, **21**(9), 1807-1823.

LANDGREBE, D. A., 2003, *Signal Theory Methods in Multispectral Remote Sensing*. (Hoboken, NJ: John Wiley and Sons).

MANNAN, B., ROY, J., and RAY, K. A., 1998. Fuzzy ARTMAP supervised classification of multi-spectral remotely-sensed images *International Journal of Remote Sensing*, **19**(4), 767-774.

MARCEAU, D. J., GRATTON, D. J., FOURNIER, R. A., and FORTIN, J.-P., 1994. Remote sensing and the measurement of geographical entities. 2. The optimal spatial resolution. *Remote Sensing of Environment*, **49**, 105-117.

MARCEAU, D. J., and HAY, G. J., 1999. Remote sensing contributions to the scale issue. *Canadian Journal of Remote Sensing*, **25**(4), 357-366.

MARCEAU, D. J., HOWARTH, P. J., DUBOIS, J.-M. M., and GRATTON, D. J., 1990. Evaluation of the Grey-Level Cooccurrence Matrix (GLCM) Method for Land-Cover Classification Using SPOT Imagery. *IEEE Trans. Geoscience and Remote Sensing*, **28**(4), 513-519.

PESARESI, M., 2000. The remotely sensed city. Final report of the post-doctoral fellowship at the JRC Ispra, Italy.

PESARESI, M., and BENEDIKTSSON, J. A., 2001. A New Approach for the Morphological Segmentation of High-Resolution Satellite Imagery. *IEEE Trans. on Geoscience and Remote Sensing,* **39**(2), 309-320.

QUATTROCHI, D. A., and GOODCHILD M. F., editors, 1997, *Scale in remote sensing and GIS*, (New York: CRC Press).

RICHARDS, J., 2003, Progress in Hyperspectral Image Data Analysis. In *Tyrrenian Workshop on Remote Sensing*, Elba Island, Italy, Sept. 15-18, 2003, 401-416.

SHABAN, M. A., and DIKSHIT, O., 2002. Evaluation of the merging of SPOT multispectral and panchromatic data for classification of an urban environment, *International Journal of Remote Sensing,* **23**(2), 249-262.

SMALL, C., 2001. Estimation of urban vegetation abundance by spectral mixture analysis. *International Journal of Remote Sensing,* **22**(7), 1305-1334.

TOWNSEND, J. R. G., and JUSTICE, C. O., 1988. Selecting the spatial resolution of satellite sensors required for global monitoring of land transforms. *International Journal of Remote Sensing*, **9**, 187-236.

WOODCOCK, C. E., and STRAHLER, A. H., 1987. The scale factor in remote sensing. *Remote Sensing of Environment,* **21**, 311-332.

ZHANG, Q., WANG, J., GONG, P., and SHI, P., 2003. Study of urban spatial patterns from SPOT panchromatic imagery using textural analysis. *International Journal of Remote Sensing*, **24**(21), 4137-4160.

Figure Captions

- Fig. 1: The modified procedure for urban area characterization with multi-scale texture analysis.

- Fig. 2: Samples of SAR images for the test area: (a) ERS-1 (descending), (b) ENVISAT (ascending), (c) RADARSAT-1 (fine beam mode), (d) SIR-C (SRL-1 mission, VV-polarization).

- Fig. 3: Ground truth (a) and training set (b) for the test area.

- Fig. 4: Overall and mean producer's accuracy values for the classification of mean, variance, dissimilarity and entropy using a wide range of window width for co-occurrence matrix computation (ERS-1 image on August 13, 1992).

- Fig. 5: HDI values for multiscale feature sets of four texture measures. Three sets are highlighted and classification results compared in the text.

- Fig. 6: Classification maps for different SAR images and textural features (w is co-occurrence window width): (a) ASAR image (contrast, correlation, dissimilarity and entropy, $w = 25$); (b) combined ASAR ascending and descending with ERS-1 image (mean, variance, entropy and dissimilarity, $w = 21$).

- Fig. 7: Producer's accuracy values for the urban classes and overall accuracy values against the incidence angle for RADARSAT-1 scenes of Pavia.

TABLE I

OVERALL ACCURACY VALUES FOR ASAR IMAGE SEGMENTATION IN URBAN DENSITY ZONES USING DIFFERENT WINDOW SIZE (w) FOR COMPUTING THE CO-OCCURRENCE TEXTURAL FEATURES.

image	$w = 15$	$w = 19$	$w = 21$	$w = 23$	$w = 25$
25/11/02	58.7%	58.2%	58.4%	58.4%	58.7%
08/12/02	60.3%	55.2%	60.3%	60.3%	58.9%
both	57.9%	60.4%	65.4%	63.8%	64.2%

TABLE II

SAR IMAGES IN THE TEST DATA SET.

Satellite	Mode	Polarization	Orbit	Posting (m)	Date
ERS-1	=	VV	Descending	12.5	13/08/92
ERS-1	=	VV	Descending	12.5	22/10/92
ERS-1	=	VV	Descending	12.5	24/06/93
ERS-1	=	VV	Descending	12.5	11/11/93
ERS-1	=	VV	Descending	12.5	03/10/94
ERS-1	=	VV	Descending	12.5	09/11/94
ERS-2	=	VV	Descending	12.5	29/10/00
RADARSAT-1	S7	HH	Ascending	12.5	20/10/00
RADARSAT-1	F2	HH	Ascending	7	
RADARSAT-1	F4	HH	Ascending	7	
ASAR	IMP IS4	VV	Ascending	12.5	25/11/02
ASAR	IMP IS4	VV	Descending	12.5	08/12/02

TABLE III

OVERALL ACCURACY (OA) AND MEAN PRODUCER'S ACCURACY (MOA) VALUES FOR THE CLASSIFICATION OF MEAN, VARIANCE, DISSIMILARITY AND ENTROPY USING THE SAME WINDOW WIDTH A DIFFERENT ORIENTATIONS OR DISTANCES FOR CO-OCCURRENCE MATRIX COMPUTATION (ERS-1 IMAGE ON AUGUST 13, 1992).

orientation	$0°$	$45°$	$90°$	$135°$
OA	63.63%	63.20%	58.55%	53.39%
MOA	59.62%	59.68%	56.70%	55.82%
distance	1	3	5	
OA	63.20%	53.89%	56.74%	
MOA	59.68%	58.00%	59.24%	

TABLE IV

CONFUSION MATRIX FOR THE CLASSIFICATION OF ASAR AND ERS-1 IMAGES (DETAILS IN THE TEXT)

	Center	Residential	Suburban	Water	Vegetation	*Producer's acc.*
Center	8962	5171	569	107	936	56.9%
Residential	4623	17422	4661	552	6423	51.7%
Suburban	1038	14068	2905	2585	13120	8.6%
Water	2291	142	15	1861	218	41.1%
Vegetation	5086	17637	1497	3047	72537	72.7%
User's acc.	40.7%	32.0%	30.1%	22.8%	77.8%	OA = 55.2%

	Center	Residential	Suburban	Water	Vegetation	*Producer's acc.*
Center	8300	6270	1055	38	82	52.7%
Residential	2626	17943	11539	498	920	53.5%
Suburban	133	9293	18027	1503	3958	54.8%
Water	0	304	82	3136	1005	69.3%
Vegetation	36	5917	5932	2054	85865	86.0%
User's acc.	74.8%	45.2%	49.2%	43.4%	93.5%	OA = 71.5%