TEXTURE-BASED CHARACTERIZATION OF URBAN ENVIRONMENTS ON SATELLITE SAR IMAGES

F. Dell’Acqua, P. Gamba

Dipartimento di Elettronica, Univ. di Pavia, Italy

doi: 10.1109/TGRS.2002.807754

Abstract

We investigate the use of co-occurrence texture measures to provide information on different building densities inside a town structure. We try to improve the pixel-by-pixel classification of an urban area by considering texture measures as a means for block analysis and classification. We find some interesting hints concerning the optimal dimension of the window to be considered for texture measures, as well as the most useful measures. Moreover, we show that it is possible to use medium-resolution readily available satellite synthetic aperture radar images for a more refined urban analysis than previously shown.

1. Introduction

A detailed analysis of urban environments requires a wealth of information to be extracted from satellite or aerial data. In many situations, the available resolution (or cost) of existing sensors’ data does not afford fulfilling the requirements of the final users. For instance, only very high resolution satellites may provide the accuracy desired by most urban planners [1], [2]. In several situations, however, existing satellite images may be sufficient, possibly at the cost of a laborious processing. This is more likely to be true where the need is for a structural analysis as opposed to a detailed building-by-building characterization.

The sensor that conveys the greatest amount of information about structural as well as dielectric properties of the urban materials is the synthetic aperture radar (SAR), whose ground resolution for spaceborne sensors is currently around 10 m. So, there is an interest in the analysis of radar images of urban areas, given also SAR all-weather capabilities and, therefore, its usefulness for disaster monitoring. However, in urban areas, where objects tend to cluster and show almost no regular structure, the problems of radar imaging have long prevented SAR from being useful for urban characterization. In the future, high-resolution SAR sensors should be increasingly available, and future satellite platforms, like the COSMO/SkyMed ones [3], will hopefully provide higher resolution data at the same cheap cost of the actual coarser ones.

For the moment, however, we have to deal with existing satellite SAR and their limited ground resolution. So, working methodologies to overcome current limitations still need to be introduced. On the one hand, we may investigate if temporal redundancy may be able to recover the single data pixel-by-pixel classification errors and problems [4]. On the other hand, we are also interested in understanding if medium area (e.g., block) characterization is still possible and worth studying with these datasets. We think that this is a very interesting and open research field, and we give here some results of our investigation on the use of cooccurrence texture measures to provide information on different building density inside a town structure.
2. Satellite SAR dataset and urban test area

The dataset used in this research corresponds to subsamples of six ERS-1 images referring to the urban and suburban area around the town of Pavia, Northern Italy, and acquired between 1992 and 1994. These images were coregistered to a map at the same spatial resolution (12.5 m in both directions). The map was extracted from the Regional Technical Map of the Lombardia Region, suitably resampled to match the SAR data.

The same map may also be used to delineate a detailed ground truth of the test area by means of manual interpretation [shown in Fig. 1(a)]. As already mentioned in [4], however, not all the classes that a map may suggest can actually be discriminated in a SAR image. For instance, railroads and urban vegetation are very difficult to distinguish from streets and built-up areas, respectively. One way to accomplish this task is to take into account structural or relational elements of the landscape. Of course, the only way to have such an information, in turn, is to use some sort of spatial analysis. Similarly, we may expect that a pixel-by-pixel classification will not be able to discriminate between residential and industrial areas. This will instead be possible if we rely on some sort of building density information. Now the point is: to what extent can we extract this information from satellite SAR data?

Fig. 1. Different kinds of ground truth available for the urban study area (Pavia, Northern Italy). (a) pixel-by-pixel seven-classes ground truth. (b) Ground truth for building density, computed starting by the previous image, where the “city center” is in white, “residential areas” in gray, and “suburban areas” in light gray. (c) Same three environments, as they are recognized by manual interpretation of very high resolution satellite data (other light gray tones represent vegetation and water respectively, ignored in this application).
Just to answer to this question, we need a different kind of ground truth than the one in Fig. 1(a). In particular, we need an estimate of the mean building density in order to provide a rough distinction of the different urban environments in the same area. A possibility is depicted in Fig. 1(b), where we used the definitions usually adopted for the characterization of radio propagation in urban areas. Here the “building density” is the ratio between the area covered by buildings and the total area, computed in one square kilometre [5]. Moreover, we label as “suburban areas” (light gray) the portions of the scene where building density is less than 20%, “residential areas” (gray) those with a building density between 20% and 50%, and “city centre” (white) those where 50% building density is exceeded. Of course, this image, while quantitatively well defined, does not match our expected subdivision of the urban area, even when relaxing “city centre” to “high-density areas.” So, it may be useful to consider a third ground truth, extracted by manual interpretation from very high resolution satellite data. This map, shown in Fig. 1(c), represents a more subjective but also more user-friendly segmentation of the investigated area.

3. Urban density maps from single SAR images

Given the ground truth in Fig. 1(c), we would like to know if it can be extracted from our European Remote Sensing (ERS) dataset. To this aim, we need first of all a classifier, able to analyze multiband data, with possibly very different statistics. We, thus, decided to rely on a Fuzzy ARTMAP structure that has already shown superior performance in classifying very complex environments, like urban ones [6]. Moreover, this advantage has been especially verified on multiband data, coming from different sensors or from processing of the original records [7].
The neural classifier may be applied directly to the ERS images, one at a time or as a multitemporal sequence [4], retrieving the built-up area with a sufficiently high precision. In particular, we found in [4] that the buildings are correctly recognized up to 67.7%. So, we expect that we could extract from this classification a building density estimate to be compared with the one in Fig. 1(a). Unfortunately, this is not the case, because the problems inside the urban area, especially in high-density blocks, prevent the SAR image from providing a very detailed map of the buildings. So, while the overall analysis is sufficiently accurate, local estimates could be wrong, and the building density exceeds (sometimes dramatically) the ground truth value. This is shown in Fig. 2, where we show the classification map and the corresponding density map obtained on August 13, 1994, June 6, 1992, and October 3, 1993. The overestimate of the “city center” area is evident, because SAR response is more related to scatterer density than to building density, which is a point we will return to later. As a consequence, we found once more that it is impossible to characterize an urban environment with respect to different building density without taking into account any textural information, only on the basis of a pixel-by-pixel analysis.

<table>
<thead>
<tr>
<th>Feature</th>
<th>3</th>
<th>21</th>
<th>41</th>
<th>77</th>
</tr>
</thead>
<tbody>
<tr>
<td>Original data</td>
<td>33.64</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Contrast</td>
<td>21.16</td>
<td>30.80</td>
<td>19.84</td>
<td>25.58</td>
</tr>
<tr>
<td>Correlation</td>
<td>18.24</td>
<td>23.93</td>
<td>24.03</td>
<td>25.49</td>
</tr>
<tr>
<td>Dissimilarity</td>
<td>21.00</td>
<td>27.60</td>
<td>19.44</td>
<td>25.65</td>
</tr>
<tr>
<td>Entropy</td>
<td>30.16</td>
<td>36.60</td>
<td>28.29</td>
<td>26.71</td>
</tr>
<tr>
<td>Mean</td>
<td>31.82</td>
<td>32.17</td>
<td>34.96</td>
<td>31.00</td>
</tr>
<tr>
<td>Homogeneity</td>
<td>40.24</td>
<td>28.97</td>
<td>29.18</td>
<td>30.25</td>
</tr>
<tr>
<td>Second Moment</td>
<td>44.39</td>
<td>45.72</td>
<td>38.84</td>
<td>34.41</td>
</tr>
<tr>
<td>Variance</td>
<td>22.20</td>
<td>32.17</td>
<td>34.96</td>
<td>31.00</td>
</tr>
</tbody>
</table>

**TABLE I MEAN HDI VALUES FOR EACH TEXTURAL FEATURE ON TEST SAMPLES**

Strong texture parameter estimation techniques are available for SAR images, based on statistical noise models [8], [9]. In our situation, however, they require the noise model of urban areas. This topic is currently under investigation [10], and final results still are to be achieved. So, to the aim of evaluating if texture measures are able to provide urban density characterization, we used a model-free approach. In particular, we considered textural features computed by means of the cooccurrence matrix [11] and applied our neurofuzzy classifier to
many different textural combinations, computed with different window width. Since the dataset is large, and the number of experiments has been huge, we limit ourselves here to describe qualitatively the achievements, leaving to the final part of this section a more quantitative analysis.

First of all, we stress that we used only one-step diagonal cooccurrence matrix, because no preferred direction is present in the data. We computed the eight texture measures listed in Table I using many different window widths (from 3 to 77 pixels). In the table, however, we show only four values, which, at the SAR ground resolution and in a relatively small town like Pavia, correspond to building, block, and structural area dimensions.

In order to choose the best feature set for the classification, we considered six test areas (25×25 pixels wide), two for each of the considered building density class. To provide a quantitative assessment of texture measures' ability to discriminate the urban environments, we use in the test areas the histogram density index (HDI) to compare textural information [12]. HDI was first considered to assess feature discriminability in high-resolution visible images of urban areas. It is a measure of distance between two probability density functions $f(x)$ and $g(x)$ and is defined as:

$$\text{HDI} = \left(1 - 2 \sum_x \min(f(x), g(x)) / \sum_x (f(x) + g(x))\right) \cdot 100.$$  \hspace{1cm} (1)

For our purposes, $f(x)$ and $g(x)$ represent one of the textural features computed from the cooccurrence matrix on two different test areas. The more they are different, the higher the HDI, and the more useful is this feature to discriminate between the classes the test areas belong to. Since we have six test areas, we have 17 combinations: the final HDI is the mean among all these comparisons and represents to what extent a feature could be helpful to recognize urban environments by their building density.

In Table I, we provide the values of this “mean HDI” for each textural feature as a function of the window width. These values are computed considering only a subset of the SAR image sequence, i.e., the three ERS-1 images recorded on August 13, 1992, October 22, 1992, and June 24, 1993. The reported values suggest that block area dimension (21 pixels, nearly 250 m) is the best choice. Moreover, even if HDI values are always lower than 50%, mean, entropy, second moment, and variance seem to be among the best choices for our classification task. Mean value is a straightforward discriminating feature for homogeneous regions in SAR images [13]. Entropy and second moment are inversely correlated [14] and discriminate among differently periodical or uniform gray-level distributions, which is the case for residential versus suburban areas. Finally, variance accounts for the different scattering patterns of differently built-up areas.

In an attempt to better fulfill our requirement of building density recognition, we investigated also if the joint use of more texture measures could be useful, using up to four features. As a matter of fact, we found that two measures are sufficient to achieve mean HDI values around 90 (see Fig. 3). We note that suburban areas tend to be very similar to residential ones for many texture combinations. Moreover, we found that the best results with four textural features are achieved using dissimilarity, entropy, mean, and variance. This choice is consistent with values in Table I, because entropy and second moment are strongly correlated, and dissimilarity adds information about correlation (actually, lack of correlation) among neighboring pixels.
Fig. 3. HDI values for different combinations of textural features (window width fixed to 21).

These results can be exploited to extract density maps. In Fig. 4, we provide these maps for all of the six ERS images, showing the stability of the urban blocks depicted in the classifications. Even in different dates, we have almost the same characterization of the urban area, with many parts of the town clearly recognized.

Fig. 4. Fuzzy ARTMAP building density classifications for the six images of our dataset. Maps are obtained clustering dissimilarity, entropy, mean, and variance computed from the cooccurrence matrix with a window width of 21 pixels.

However, it is worth observing that there are still some differences between the best fuzzy ARTMAP results and the ground truth. These differences account for the relatively low overall accuracy values that we find; as a
matter of fact, no more than 60% overall accuracy was reported. One possible explanation is that the building density ground truth was realized by manually interpreting an Ikonos-2 image and investigating the building densities, while the SAR image is primarily a map of the scatterer density. It is, therefore, interesting to analyze the classification results for one of the ERS images and compare problematic areas with the aid of the same image of Pavia, acquired on July 2001. Since the town has not changed too much in the last ten years, the time difference is immaterial for our research.

We observe (Fig. 5) that indeed the original ground truth is not sufficiently precise, at least outside the city center. There are a number of situations where residential areas are constituted by many small houses very close one to the other, so that their scattering behavior is similar to the one of the large clustered buildings of the historical center. Moreover, the transition between the latter and the river banks, as correctly depicted by the classification maps, is smoother than considered in the original ground truth. Finally, even in the same historical center, there are areas where the buildings blocks are separated, and these are correctly considered as lower building density zones. That is why the original ground truth has undergone a change reflecting these considerations, which resulted in accuracy improvements up to 10% for the characterization of the urban classes.
Looking in more detail to the quantitative results, accuracy values for the urban density environments are presented in Table II. The first comment is that the central area and the densely built outer parts are correctly recognized, although with a few problems still open, especially in the residential areas, where the building density is usually underestimated. Visually, light grey areas replace grey ones, and sometimes very dense blocks within these patterns appear. Moreover, while the overall impression is almost the same for the classification maps, accuracy values do change for the same class in different images. These differences are partly due to coregistration issue, and partly to speckle noise, since we worked on unfiltered images. However, the overall accuracy is sufficiently stable, changing only a few points from date to date.

<table>
<thead>
<tr>
<th>ERS image</th>
<th>city center</th>
<th>residential</th>
<th>suburban</th>
</tr>
</thead>
<tbody>
<tr>
<td>Aug, 13, 1992</td>
<td>64.0</td>
<td>22.3</td>
<td>56.4</td>
</tr>
<tr>
<td>Oct. 22, 1992</td>
<td>23.0</td>
<td>35.2</td>
<td>47.3</td>
</tr>
<tr>
<td>June 24, 1993</td>
<td>37.9</td>
<td>30.5</td>
<td>51.1</td>
</tr>
<tr>
<td>Nov. 11, 1993</td>
<td>29.2</td>
<td>38.4</td>
<td>48.3</td>
</tr>
<tr>
<td>Oct. 3, 1994</td>
<td>60.6</td>
<td>25.6</td>
<td>44.5</td>
</tr>
<tr>
<td>Nov. 9, 1994</td>
<td>71.0</td>
<td>54.0</td>
<td>2.5</td>
</tr>
<tr>
<td>sum3a</td>
<td>49.1</td>
<td>35.0</td>
<td>54.5</td>
</tr>
<tr>
<td>sum3b</td>
<td>46.3</td>
<td>33.6</td>
<td>55.7</td>
</tr>
</tbody>
</table>

**TABLE II** CLASSIFICATION ACCURACY VALUES FOR THE URBAN ENVIRONMENTS

As a matter of fact, we may say that textural information from the cooccurrence matrix helps in discriminating the urban environments. Moreover, they reduce the building density overestimate experienced by classifying intensity data alone. Generally speaking, however, the above-mentioned accuracy values are still far from being
satisfying, suggesting that the textural features we are using may not be optimal to our task. Moreover, SAR texture effects depend on scattering patterns, while the ground truth in Fig. 1(c) is based more on land use patterns. Even if the two patterns are related, our results show that using only SAR images, it is difficult to retrieve the latter with very high detail.

One more reason for these differences may be the 21-pixel-wide window used to compute textural features. Indeed, the coarse-data resolution and the use of this relatively large window strongly limits our capability to delineate in detail the environment boundaries.

Finally, to evaluate to what extent the proposed technique could be applied to other urban environments, we investigated also the area of Milano, also imaged in our ERS dataset. While no detailed ground truth is available, in Fig. 6 we offer a comparison among the classification map and the SPOT 10-m image available through the National Imagery and Mapping Agency Web site, showing a sufficiently good agreement with respect to the salient features of the town structure.

4. Multitemporal urban density maps

To complete the discussion in Section III, we investigated also the advantages connected with the multitemporal aspect of the dataset. For direct pixel-by-pixel classification, the problem has already been investigated in [4], and the result is that the urban structure could be recognized by using more images, although the advantage, in terms of overall accuracy, is not dramatic. Therefore we expect the same: textural information should be stressed by the multilook effect that corresponds to the use of more than one image, and therefore the density map should provide higher (even if not much higher) accuracy values.

The multitemporal analysis could be done in two different ways, since we are considering static features, corresponding to structural characteristics of the urban area not changing in the time frame of our ERS images. So, we could merge more images to obtain less noisy data, extract the textural features, and then classify them, or simply jointly classify these features for the dates we are using. When the first three images of the sequence are considered and both the analyses are performed, we obtain the results shown in Table II, case “sum3a” and “sum3b,” respectively. We observe that both joint analyses provide a better characterization of the urban environments, although only to a limited extent. In particular, it seems that the joint use of three images produces a global refinement of the classification because each of the urban classes is better recognized, but not better than the best single date map.

This effect is stressed merging the whole sequence (six images), obtaining a classification accuracy of 47.0%, 34.8%, and 54.8% for the three urban classes, respectively. The reason could be that the despeckle effect makes punctual urban features more evident, improving the pixel-by-pixel classification results, but does not help textural features to better discriminate between urban environments. Texture measures take into account spatial relationships between pixel values, while speckle noise affects randomly isolated pixels. Inside urban areas, where uniform zones are rare, despeckling has almost no effect on texture analysis.
5. Conclusion

The results of the procedure proposed in this letter show that it is possible to extract some kind of information on urban environments from current satellite SAR images and classify them with respect to building density. The coarse resolution of ERS images does not prevent the possibility to characterize these areas.

In particular, cooccurrence measures computed with window width corresponding to the mean block dimension in the considered urban area allow representing crowded, residential, and suburban areas with sufficient precision and stability in the classification maps.

One interesting conclusion of this analysis is that current satellite SAR data may provide hints on how to interpret maps of urban areas. For instance, the use of a classified density map can help in refining a manually extracted ground truth.

Finally, the density maps have been compared with a high-resolution satellite image of the same area, confirming that the areas where difference between classifications and ground truth arise correspond to zones where the ambiguity in urban environment definitions are more evident.

Acknowledgements
This work was supported by ASI, the Italian Space Agency

References


