Comparing different textural approaches to extract human settlement from CBERS-2B data

Gianni Cristian Iannelli
Paolo Gamba
Fabio Dell'Acqua

University of Pavia, Department DIII
27100 - Pavia - PV, Italy
{giannicristian.iannelli,paolo.gamba,fabio.dellacqua}-at-unipv.it

Abstract. Although it is well-known that textural features have proved to be the most relevant ones to extract human settlements from high spatial resolution (HR) panchromatic remote sensing images, it is still to be investigated in details which is the most efficient one. This is especially true when looing for a methodology aiming at settlement extent detection and at the same time able to cope with various and different definitions of what a "human settlement" is. In this paper we compare two approaches developed for this task and already checked at the global level on HR and VHR SAR data, comparing their performances on two complete CBERS-2B panchromatic scenes. The processing chains, their results and the computational loads of two approaches, one based on the co-occurrence *contrast* feature and one on the occurrence *range* feature, are introduced, evaluated, and discussed. The relevant outcome of this research is that there is no clear winner, and less complex/less efficient approaches may still be able to provide suitable results. Detection accuracy and false positives (commission errors) varies according to the geographical scale considered and the features used to identify the settlement, such as their size, number of built-up stricture, built-up area density and so on.

Keywords: Urban remote sensing, CBERS, textural features, human settlements.

1. Introduction

The methodologies developed to extract human settlement extents and presented in technical urban remote sensing literature are mostly driven by the sensor typology and the spatial/spectral resolution of the input data. For coarse spatial resolutions, for instance, spectral information is considered as the most efficient way to discriminate between urban and rural environments, using for example nighttime lights (Elvidge at al. 2007), spectral indexes obtained by combining wideband reflectance values collected at different wavelengths (Phinn et al. 2002), or other spectral features (Forster, 1983). With higher spatial resolution data, textural features have been considered as equally or even more important (Glich, 2002), to the extent that some authors feel they are the only relevant features to be considered, like in Pesaresi et al. (2007). We follow this path and would like to use panchromatic HR data coming from CBERS-2B and analyze two approaches based on the co-occurrence (Haralick et al., 1988) or occurrence matrices.

With respect to existing literature, this paper aims at introducing and validating processing chains not only effective in extracting human settlement extents, but also flexible enough to accommodate for different definitions of what a human settlement is. Indeed, the definition of "human settlement", and therefore the issue to find consistent ground truth to validate approaches aimed at urban remote sensing in wide geographical areas, has been consistently one of the main points to be disputed in global urban remote sensing research. The approach followed in (Schneider et al. 2010), for instance, although very valuable and a the very basis of the MODIS 500 m data set, currently the most widely used global urban layer, limits validation to a huge datasets of large cities. It lacks instead, and on purpose, because of the resolution of the input data, a validation for small towns and villages: Accordingly, the methodology proposed for extracting that layer cannot be applied to higher spatial resolution data. The aim of this work is instead to compare two methodologies available for HR data, and see how they could be extended to match different ground truths.

2. Tools

As mentioned in the introduction, the processing chains discussed in this work are devoted to HR panchromatic data analysis, and aim at human settlement extent extraction. They exploit the important consideration that in this kind of images urban areas may be immediately recognized by a human interpreter based on their distinct spatial patterns, due to the artificial composition of structures and gaps between (mostly, but not only roads), in a more ordered fashion and with higher local contrast within these areas than in any natural environment. Accordingly, the two chains proposed in Section 2.1 and 2.2 are based on the extraction and exploitation of textural features, that are a powerful way to quantify these spatial relationships.

2.1 The "Urban Focus" processing chain

The first processing chain discussed in this work is based on the "Urban Focus" (UF) software tool developed in the framework of the BREC suite, Gamba et al. (2008). It is not limited however to the UF tool by itself, but includes a post-processing step aimed at reducing the uncertainties due to areas that are not built-up (in the sense that they do not include buildings, for instance), but still are part of the human settlement. Examples are urban parks, undeveloped areas within the urban boundary or construction sites, both within and around the main built-up area. The complete chain is depicted in Figure 1. It must be remarked that the data were resampled by a factor of two in order to fit the recommended resolution of the input image for the UF tool, in the range if 5 meters.

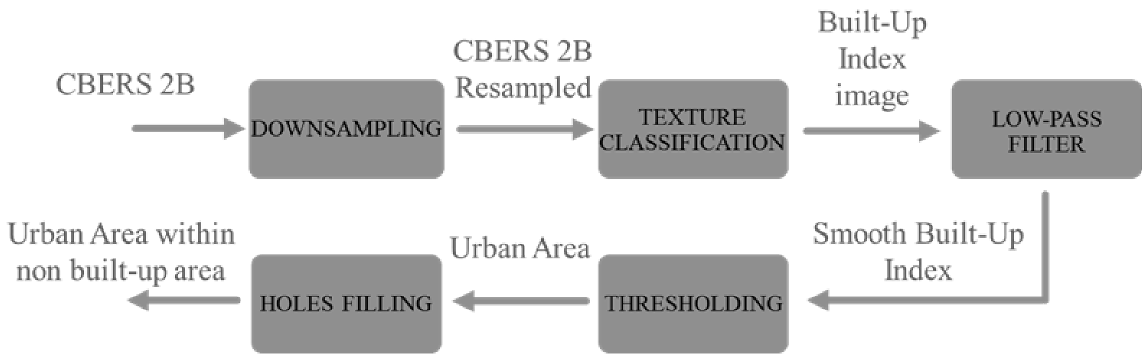

Figure 1. The first processing chain presented and compared in this paper, based on the UF tool.

UF implements a slightly generalized version of the PanTeX index, proposed in Pesaresi et al. (2007), with improvements discussed in Gamba et al. (2008). It detects 'built-up areas', under the assumption (validated a posteriori by the excellent extraction results) that buildings and other built-up artificial structures have a strong contrast with their background, while natural environments are not that contrasted. Additionally, urban features have no peculiar direction, and thus exhibit an anisotropic behavior.

To this aim, UF exploits the contrast co-occurrence textural feature computed from the Grey Level Co-Occurrence Matrix (GLCM). The GLCM matrix is a n by n matrix containing the relative frequencies with which two pixels linked by a spatial relation (displacement vector) occur on a local domain of the image, one with gray level i and the other with gray level j, with $i,j \in [0..n\text{-}1]$, and n is the number of gray-levels of the image. Contrast is thus computed as:

$$\text{CON} = \sum_{i=1}^{Ng} * \sum_{j=1}^{Ng} (i-j)^2 * P_{i,j} \qquad (1)$$

where n has been substituted by N_g, and $P_{i,j}$ is the (i,j)-th entry of the co-occurrence matrix.

After computing the contrast in different directions and combining them into the UF index according to their minimum value, the processing procedure performs additional steps, aimed at correcting the differences between built-up areas (the output of this first detection step) and human settlements. Specifically, under the assumption that urban areas include both the built-up structures and the gaps between them, the following steps (see again Figure 1) are implemented.

1. ***Low-pass filtering***: an average moving kernel of size 31x31 pixel was used, to smooth the built-up feature extraction in the scale of 100 meters; the net effect is a graceful degradation of built-up area both within the human settlements and in the urban/rural fringe. An additional result of this step is the reduction of isolated false positives caused by small pixel agglomerates in addition to the extraction of more homogeneous areas.
2. ***Thresholding***: a threshold is applied to the smoothed image in order to get the instances representing human settlements. This threshold must be selected carefully, and the selection of its value is discussed in details in the results section.
3. ***"Hole" filling***: the extracted areas detected in the previous step could include "holes", that is non built-up areas such as parks, airports, urban water bodies (e.g., ponds). This step reduces these false negative by including them into the final human settlement extents. Specifically, the algorithm deletes instances (that represent urban areas) based on its number of pixel. It counts the number of pixel for every image object and, if lower than a defined threshold, deletes it. The instances have value '255' and the background has value '0'. What happens it's that there are image objects with value '255' (urban area) and inside them image objects that have value '0' that are parts of the background (non built-up areas). The method used, firstly, swaps the classes values (negative of the image) and so, the background assumes value '255' and the instances representing urban areas assume value '0'. Consequently, in the new image there are: a huge instance (that was the background) and others small instances (representing the non-built up area) with value '255', the new background composed by the image-objects representing the urban area. The middle step is to apply the algorithm applying a very high threshold (the maximum limit is the huge instance) in order to delete the non built-up area in the new parts of background; the result is an image with just the huge instance with value '255'. In the end the method swaps again the classes and the final result is an image which has image objects representing the urban area without holes. The threshold value is not critical because the histogram of the background has a large gap between the small image-objects (representing the non built-up area inside the human settlements) and the huge instance representing the effective background (non urban area).

2.2 The *range* processing chain

The second processing chain is based on the *range* textural feature. This feature is computed starting from the occurrence (as opposed to co-occurrence) matrix, and is calculated as the difference between the maximum and the minimum value of the reflectance in a 5x5 kernel moving over the image. A graphical representation of the chain, including the range extraction and the following post-processing steps, is provided in Figure 2.

As for the first chain, the computation of the textural feature is only the first step of the chain. The other steps are similar to those proposed in the previous chain, although the values of the threshold may be different, as once more discussed in the result section.

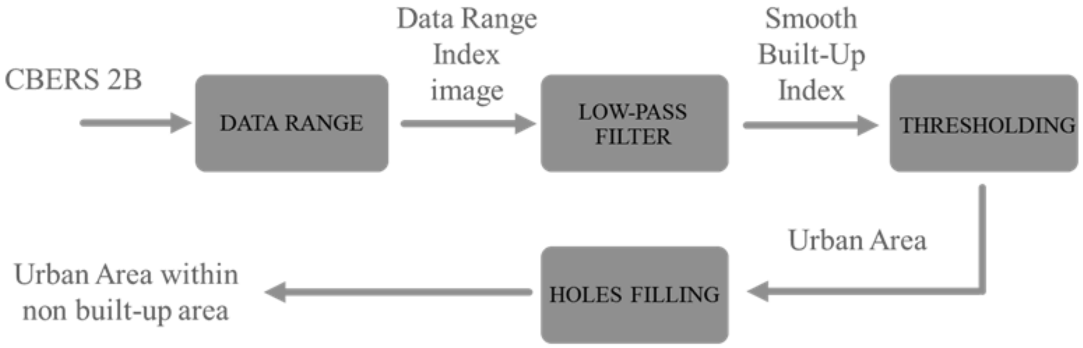

Figure 2 the second processing chain presented and compared in this paper, based on the range textural feature.

3. Experimental results

The data used in this research were acquired by CBERS 2B. The CBERS (Chino-Brazilian Earth Resource Satellite) Program, Lino et al. (2000), born from a partnership between Brazil and China to jointly develop earth observation missions, lead so far to the development of three satellites, called CBERS-1/2 and 2B, all of the operated from the Brazilian INPE ground segment. Specifically, the CBERS-2B satellite carried three sensors, namely the High Resolution CCD Camera (HRCC), the High Resolution Camera (HRC) and Wide Field Imager (WFI). The HRCC is an instrument collecting 5 bands (a panchromatic image, plus RGB and NIR bands), all of them with a spatial resolution of 20 m and a swath of 113 km.

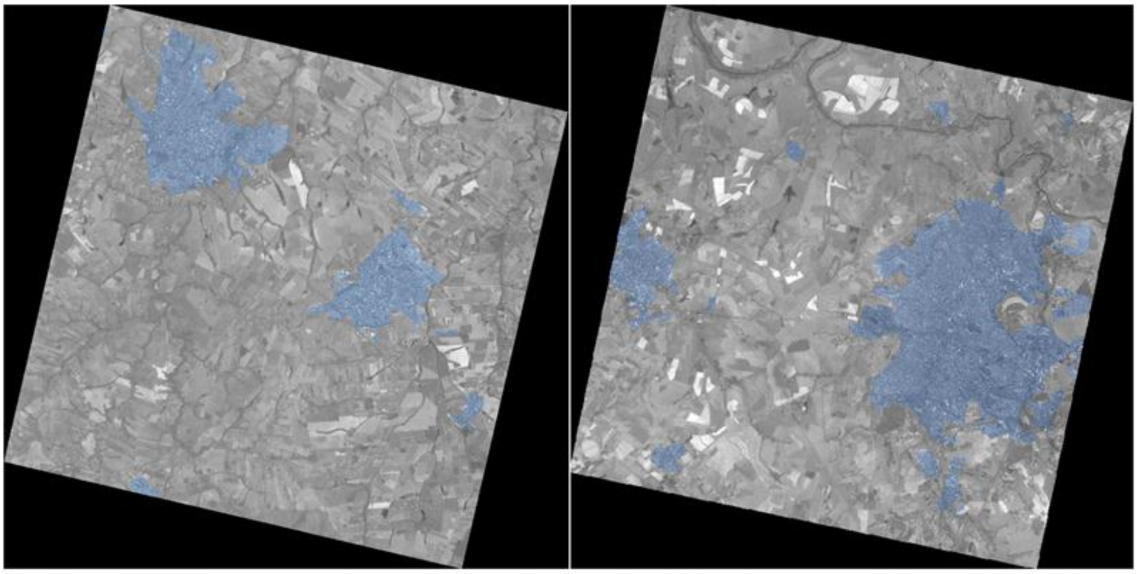

Figure 3. Available ground truth at coarse resolution for the two CBERS-2B scenes analyzed in this work: on the left, the Araçatuba scene, on the right the Ribeirão Preto scene.

The HRC is instead composed by a panchromatic sensor with a finer spatial resolution (2.36 m) and a smaller swath width (27 km). Finally, the WFI acquired only two bands, in the red and near infrared wavelength ranges, with a spatial resolution of 260 m and a swath width of 885 km. For our study, only the images acquired by the HRC sensor were used, as they match the high spatial resolution we are looking for. The images collected by the satellite and freely available through the INPE on-line catalogue, are affected by an imprecise georeferentiation, and had to be manually geometrically corrected using a set of Ground Control Points, selected in an uniform way over each of the images. To perform the geocorrection, as the terrain effect has been already corrected in the delivered data, only a very simple first order polynomial transformation method was adopted to achieve a sub-pixel precision level. More specifically, the images used in this work correspond to two scenes in the vicinity of the Araçatuba (SP, Brazil) and Ribeirão Preto (SP, Brazil) urban areas, respectively. They were collected on 10-10-2008 and 30-07-2008. A first set of ground truth data used for comparison (see Figure 3) are based on photo interpretation of Landsat data acquired a few years before, in 2004, updated using additional data by the same sensor in 2009, Pereira et al. (2005), Rudorff et al. (2010). Accordingly, these reference data have a resolution which is lower than the one of the CBERS data used in this work, and thus they lack by definition some of the smaller settlements that are instead clearly visible in the considered scenes. A second set of ground truth data, available only for a smaller subset of each entire image, were extracted manually by photo interpretation of the same data, and have, obviously the same spatial resolution.

Results of the analysis of the first scene are provided in Figure 4. They show that the “True Positive Rate” decreases for increasing threshold values for all the cases. The so called “False Positive Rate” (i.e., the commission error) decreases as well, because, as the threshold increases small agglomerates of buildings, which are not considered as human settlements in the available ground truth, are discarded. It’s interesting to note that the best results are obtained when the “holes” are filled, because the human settlement definition underlying the Landsat-based ground truth data is that they are compact areas. Therefore, the improvement related to the “hole filling” procedure is due to the inclusion of small non built-up areas inside the human settlements into the final boundary sets.

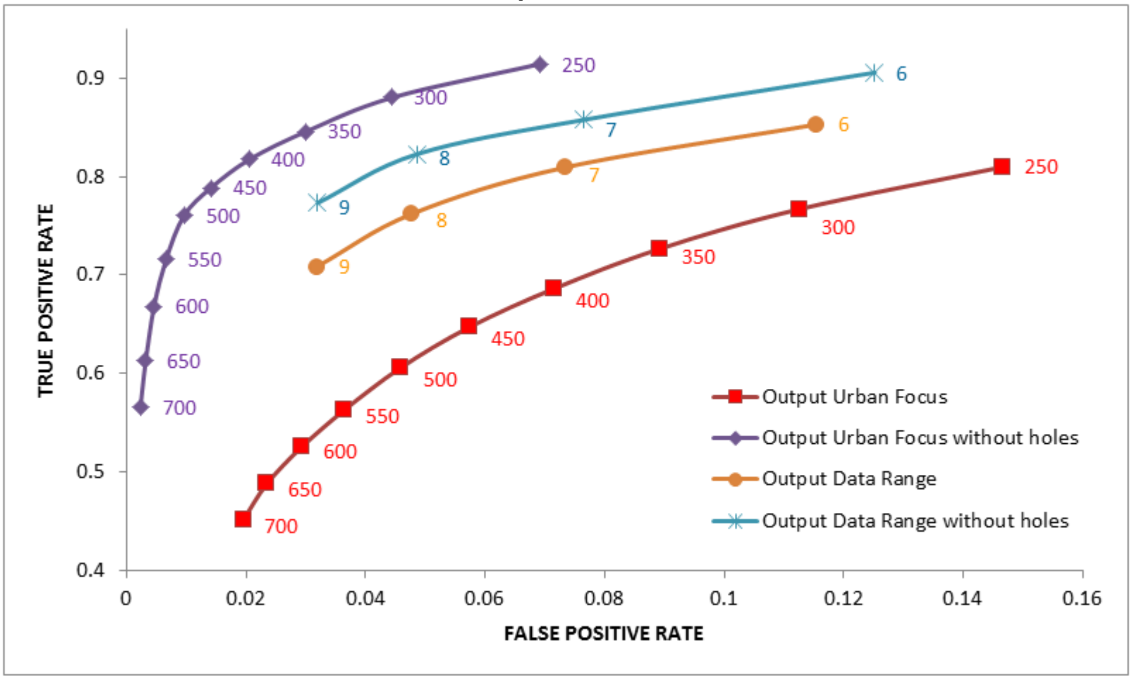

Figure 4. Classification accuracies for the Araçatuba scene, considering the coarse ground truth in Figure 3, and comparing the contrast and the range extraction, with and without the additional processing steps described in Sections 2.1 and 2.2.

Similarly, results of the analysis of the second scene are provided in Figure 5. They show the same behavior of the previous case. In this case, filling the holes improve the results but not as much as in the previous case. As mentioned above, this is due to the limited quantity of non built-up area inside the human settlements in this scene.

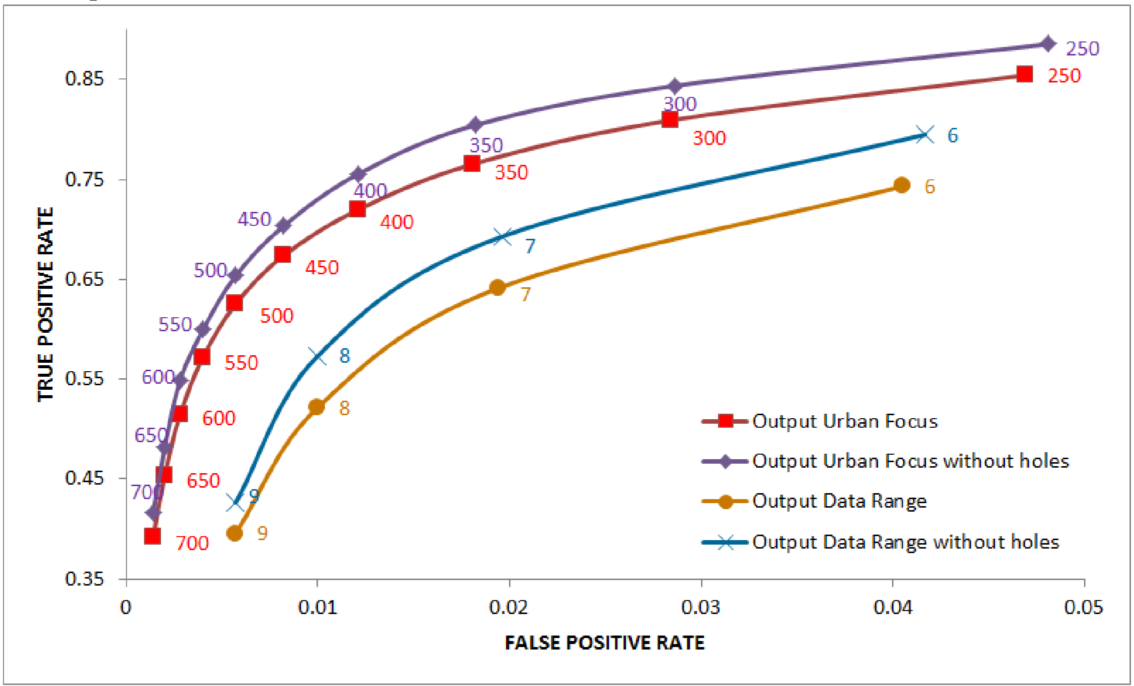

Figure 5. Classification accuracies for the Ribeirão Preto scene, once again considering the coarse ground truth in Figure 3, and comparing the contrast and the range extraction, with and without the additional processing steps described in Sections 2.1 and 2.2.

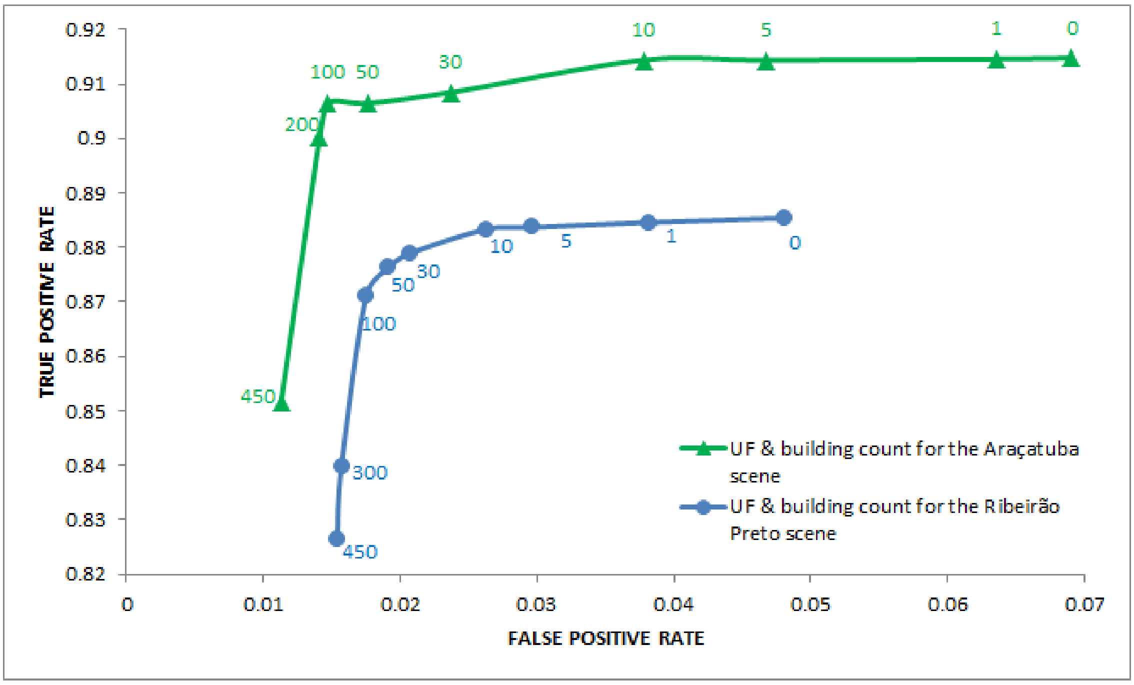

Figure 6. Classification accuracy for both the Araçatuba and the Ribeirão Preto scenes, considering building counts as an additional information inside each of the extracted human settlement, and discarding those with a building count smaller than the mentioned parameter.

Finally, we would like to add that, as mentioned in Iannelli et al. (submitted), it is possible to improve the results by including information about a very rough estimate of the building counts inside each of the human settlements extracted by means of the two previously discussed chains. Of course, as the range results are always worse than the ones obtained by means of the UF processing chain, we will concentrate on this, and the results for the two scenes are proposed in Figure 6.

Figure 6 shows that the selection of urban areas according even to a rough estimate of the number of buildings improves the detection rate by, once again, reducing the "False Positive Rate". The "True Positive Rate'" value does not decrease as in the previous cases because the instances are selected using a model that permits to discriminate the false positives from the human settlements. In the Figure 6 it could be also seen that instances with at least 100 buildings better represents the available Landsat-based ground truth.

The same set of analyses were performed by using the second set of ground truth data, and lead to the results showed in the Figure 7. The curves have the same behavior of the previous cases, after a linear decrease of the "False Positive Rate" keeping the "True Positive Rate" almost equal, the curves start decrease significantly the values of "True Positive Rate". This happens because raising the threshold, it encounter a value that match the model of the ground truth and, after that, start to mismatch again decreasing the "True Positive Rate".

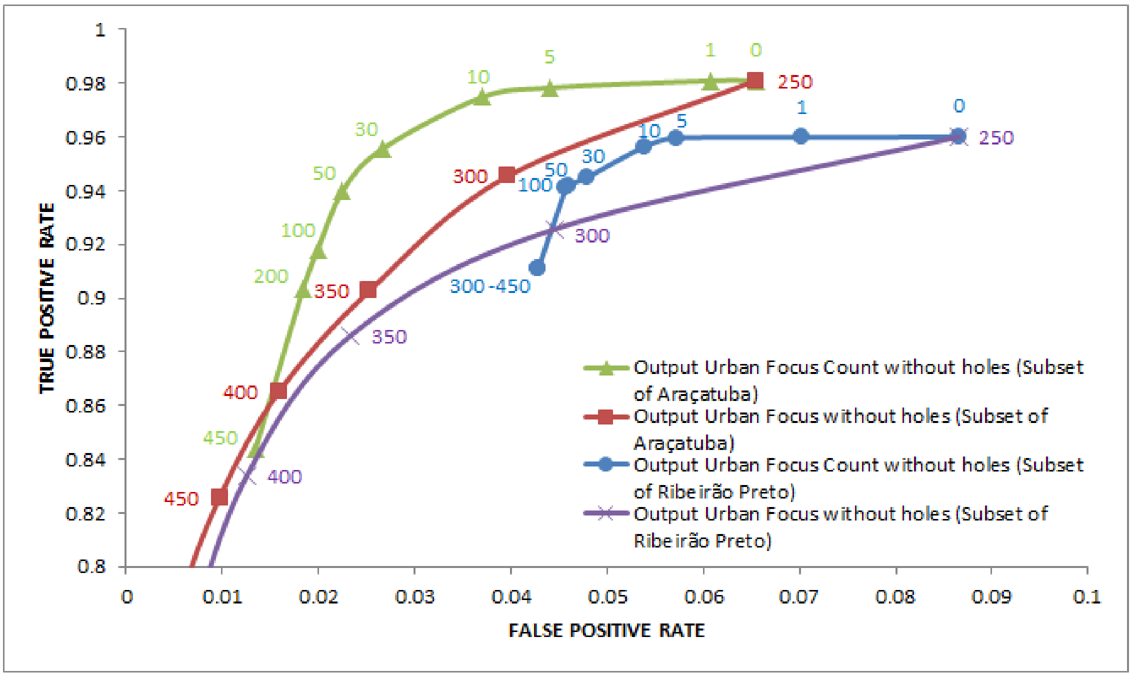

Figure 7. Classification accuracy for both the Araçatuba and the Ribeirão Preto scenes, considering the building counts, compared with the second ground truth.

Looking at the Figure 7, it could be seen that the instances with at least 10 buildings better represent this second ground truth, which is consistent with its finer spatial resolution. In this case its more difficult understands the model because the ground truth is available just on a subset and so, with less image objects representing the human settlements.

4. Conclusion

In this paper we presented a new framework suitable for human settlement extraction from VHR data using textural information, and easy to be tuned to ground truths with different spatial scales, and possibly corresponding to different definitions of "human settlements". The framework proved to be robust and dependent on parameters clearly understandable even from non-technical users, and thus suitable for open-source tools available to the public.

5. Acknowledgments

This work was supported by the TOLOMEO FP7 project.

6. Reference

Elvidge, C. D.; Tuttle, B. T.; Sutton, P. C.; Baugh, K. E.; Howard, A. T.; Milesi, C.; Bhaduri, B.; Nemani, R.; Global distribution and density of constructed impervious surfaces. **Sensors**, 7, 1962-1979; 2007.

Phinn, S.; Stanford, M.; Scarth, P.; Murray, A. T.; Shyy, P. T.; Monitoring the composition of urban environments based on the vegetation-impervious surface-soil (VIS) model by subpixel analysis techniques. **International Journal of Remote Sensing**, 23, 4131-4153; 2002.

Forster, B.; Some urban measurements from Landsat data. **Photogrammetric Engineering and Remote Sensing**, 49. 1693-1707; 1983.

Gluch, R.; Urban growth detection using texture analysis on merged Landsat TM and SPOT-P data. **Photogrammetric engineering and remote sensing**, 68, 1283-1288; 2002.

Pesaresi, M.; Gerhardinger, A.; Kayitakire, F. Monitoring settlement dynamics by anisotropic textural analysis by panchromatic VHR data. Proc. of JURSE 2007, 11-13 April, Paris, France, 2007.

Haralick, R.M.; Shanmugam, K.; Dinstein, I.; Textural features for image classification. **IEEE Trans. Systems, Manufact., Cybernet.**, Vol. 3, no. 6, pp. 610-621, 1988.

Friedl, M. A.; Sulla-Menashe, D.; Tan, B.; Schneider, A.; Ramankutty, N.; Sibley, A.; MODIS Collection 5 Global Land Cover: algorithm refinements and characterization of new datasets. **Remote Sensing of Environment**, vol. 114, pp. 168-182, 2010.

Gamba, P.; Pesaresi, M.; Molch, K.; Gerhardinger, A.; Lisini, G.; Anisotropic rotation invariant built-up presence index, applications to SAR data. Proc. of IGARSS'08, Boston (USA), 6-11 July 2008, vol. V, pp. 338-341.

Lino, C. D. O.; Lima, M. G. R.; Hubscher, G. L.; CBERS – An International Space Cooperation Program. **Acta Astronautica**, Vol. 47, No2-9, pp559-564, 2000.

Pereira, M. N.; Goncalves, C. D. A. B.; Souza, I. M. E.; Garcia, S.; Portela, A. G.; Almeida, C. M.; Florenzano, T. G.; Uso de imagens de satélite como subsídio ao estudo do processo de urbanização (Use of satellite images to support studies on the urbanization process). **Revista de Estudos sobre Urbanização, Arquitetura e Preservação**, São Paulo, SP, v. 46, p. 3-33, 2005.

Rudorff, B. F. T.; Aguiar, D. A.; Silva, W. F.; Sugawara, L. M.; Adami, M.; Moreira, M. A.; Studies on the rapid expansion of sugarcane for ethanol production in São Paulo State (Brazil) using Landsat data. **Remote Sensing** 2(4):1057-1076, 2010.

Iannelli, G. C.; Gamba, P.; Dell'Acqua, F.; Feitosa, R.Q.; Costa, G. A. O. P.; Towards a Free Open Source Software Tool for Human Settlement Detection on VHR Remote Sensing Images.Proc. JURSE 2013, 21-23 April, Sao Paulo, Brazil, 2013. (Submitted; under review).