FOCUS PRE-PROCESSING CHAIN FOR OBJECT DETECTION IN HIGH RESOLUTION REMOTE SENSING IMAGES

C. Valladeau, J. Michel

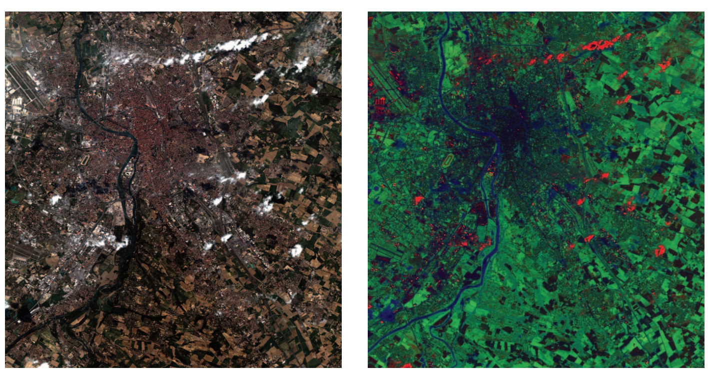
C-S - IGI
Parc de la Plaine
Rue Brindejonc des moulinais, BP 5872
31506 TOULOUSE CEDEX 5 - France

J. Inglada

CNES - DCT/SI/AP - BPI 1219
18 avenue Edouard Belin
31401 TOULOUSE CEDEX 9 - France

1. INTRODUCTION

Efficient object recognition techniques are often time consuming and as such only useable in restricted areas. On the other hand, in operational applications, one starts from full scenes, which are very large in the case of high resolution imagery such as Quickbird products or the upcoming Pleiades ones. Therefore there is still one step missing toward the design of a complete object detection chain. This step aims at quickly pruning regions of the full scene that for sure do not contain any desired object. One can then use more accurate and intensive object detection techniques on the remaining regions. This paper proposes an image pre-processing chain adapted to 4 bands (blue, green, red, near infra-red) high resolution imagery for a given set of template images representing the object of interest. First, we define a coarse description of the images using well known features. This coarse description is then used to derive a likelihood map at a given lower resolution. Finally, this similarity is used to segment regions of interest, on which the algorithm iterates at a higher resolution. With respect to other approaches present in the literature, as for instance [1], our approach has the benefit of being independent of a prior model for the object description. Also, the notion of saliency inspired from the human visual system [2] is not always able to grasp the diversity of *interest objects* in remote sensing imagery. We present here a more physically inspired approach for the problem.



(a) Input image

(b) RGB composition of the coarse description

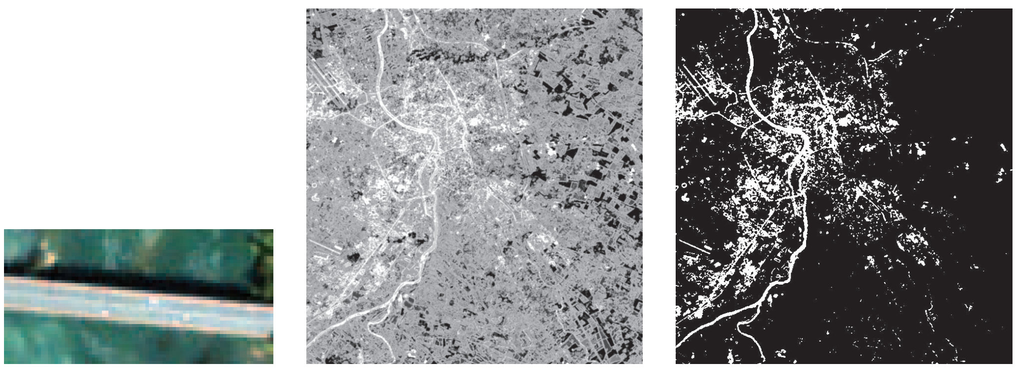
Fig. 1. Coarse image description of a Quickbird scene of the city of Toulouse, France

2. COARSE IMAGE DESCRIPTION

In order to cast images in a common feature space where they can be compared, we developed a coarse description based on the amount of different land cover classes as for instance vegetation, water or clouds. We used the classical normalized difference vegetation index, NDVI, to characterize vegetation. Also, a feature to characterize water was implemented. This water index, is inspired from the NDVI, where the red band is replaced by the green one. It therefore similar to the NDWI [3], but adapted to the available spectral bands. The cloud index uses a simple normalized spectral angle with respect to a reference cloud spectrum. Figure 1 shows the coarse description on a Quickbird scene over the city of Toulouse. We can see that water and vegetation are quite well characterized. A color composition (R = clouds, G = vegetation, B = water) computed on a low resolution image is presented. This result allows to have an interesting synoptic view of the image content. Clouds can be confused with white buildings, but this will be sorted out later in the processing using multi-scale analysis. The approach is not limited to 3 indices, and other indicators, for example for built up areas as for instance modified versions of the NDBI [4] can be used.

3. LOWER RESOLUTION LIKELIHOOD MAP

In this section, we will use the coarse description derived in section 2 to build a lower resolution map of the full scene denoting the likelihood of finding occurrences of the searched object (template image) at the given location. To do so, we first down-sample both the full scene and the template images with a decimation ratio of 8. We then build the histograms corresponding to the indices of our coarse description on the down-sampled template images. Values are weighted according to a Gaussian centered on the template image, so as to give more weight to center pixels. We then walk across the down-sampled full scene with a neighborhood corresponding to the mean size of the down-sampled template images. We use this neighborhood to build and compare the local histograms to the ones of the template images using a distance. The distance can be chosen among the classical set of distances, as for instance the Euclidean one or a spectral angle distance. The inverse of the distance is used as a likelihood measure of finding occurrences. An example is shown in figure 2. One can see that analyzing the image at this resolution allows us to discard around 60 % of the image.



(a) Template image of a bridge (b) Likelihood map (c) Thresholding of the likelihood

Fig. 2. Lower resolution likelihood map to find occurrences of bridges

This procedure can be further refined by iterating, at a higher resolution, for those regions that have been kept after the thresholding. The likelihood will be accurate in discarding regions of no interest up to a certain resolution (about 10 m.). For higher resolutions, other approaches should be considered.

The full paper will show a detailed analysis of the performances of the proposed approach for different data sets and parameters. The relevance of the different indices used for the template matching will be assessed.

4. REFERENCES

[1] Bernard Chalmond, Benjamin Francesconi, and Stéphane Herbin, "Using hidden scale for salient object detection," *IEEE Transactions on Image Processing*, vol. 15, no. 9, pp. 2644–2656, 2006.

[2] Laurent Itti, Christof Koch, and Ernst Niebur, "A model of saliency-based visual attention for rapid scene analysis," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 20, no. 11, pp. 1254–1259, 1998.

[3] Bo-Cai Gao, "NDWI – A Normalized Difference Water Index for Remote Sensing of Vegetation Liquid Water From Space," *Remote Sensing of Environment*, vol. 58, no. 3, pp. 257–266, Dec. 1996.

[4] Y. Zha, J. Gao, and S. Ni, "Use of normalized difference built-up index in automatically mapping urban areas from TM imagery," *International Journal of Remote Sensing*, vol. 24, no. 3, pp. 583–594, Feb. 2003.