OBJECT COUNTING IN HIGH RESOLUTION REMOTE SENSING IMAGES WITH OTB

E. Christophe

CRISP
Block SOC-1, Level 2,
Lower Kent Ridge Road
Singapore 119260

J. Inglada

CNES
DCT/SI/AP
18, Av. E. Belin,
31401 Toulouse Cedex 9, France

ABSTRACT

Satellite observation is particularly enticing due to its large acquisition capabilities. However these large capabilities kindle new challenges for information analysis. Object counting is one of those. To help releasing constraints on the human operator, it is important to free him from repetitive tasks and focus his attention on the high level tasks for which algorithms are not suitable yet. This abstract focuses on building counting in dense areas. The processing is done using the Orfeo Toolbox, an open-source image processing library. This paper proposes several methods with different trade-offs in terms of performance and user involvement. The method has been adapted and successfully used in other situations, as for instance counting tree stands or tents in a refugee camp.

Index Terms— Object Recognition, Object Counting, High Resolution Remote Sensing

1. INTRODUCTION

Object counting is a recurrent challenge in remote sensing applications. Satellite observation is particularly enticing due to its large acquisition capabilities. One bottleneck in the information production chain is the ability to automatically perform image analysis. In many application fields, automatic methods do not provide the quality required by end users. To help releasing constraints on the human operator, it is important to free him from repetitive tasks and focus his attention on the high level tasks for which algorithms are not suitable yet.

One of these repetitive tasks is object counting. Object counting appears in a wide range of domains: agriculture (tree stand counting), economic and development (population evaluation by building counting), biodiversity (animal population study). The problem relies mainly on identifying and locating the objects to count, once these objects have been identified, the counting itself is the trivial part.

In order to detect the objects of interest, many different approaches can be used. One first approach could be using template matching, that is using an example of the class of objects to count – a small image patch, for instance – and try to match it at different image positions. The matching can be done using similarity measures as for instance the correlation coefficient. This approach, although sound and robust for some applications, has the main disadvantage of not allowing variability between objects of the same class, mainly in terms of shape. Generalizations of this approach have been developed [1] by extracting invariant descriptions and using a supervised classification instead of a pixel-based similarity measure. This approach shows interesting results for the case of many examples and counter-examples of the class of interest are given to the classifier. A large amount of examples implies a deep involvement of an operator, which reduces the interest of the approach for an interactive user-in-the-loop approach.

The main idea of the approach presented in this paper is to propose a generic approach which reduces operator intervention.

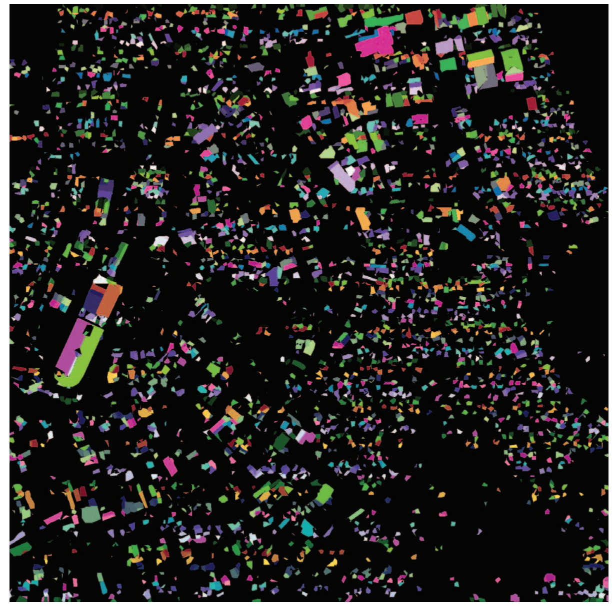
This paper focuses on building counting in dense urban areas. This work has been done in the context of the *PRRS 2008 Algorithm Performance Contest* focusing on building extraction [2]. Data sets are Quickbird images as illustrated on figure 1(a). The processing is done using the Orfeo Toolbox[3], an open-source image processing library. The original algorithm used in the PRRS contest is described in section 2. Then in section 3, simplified versions of this algorithm requiring less involvement of the operator are presented. Finally, results are detailed in section 4.

2. ALGORITHM DESCRIPTION

The algorithm is designed for high resolution optical images. Usually, these images are provided with a high resolution panchromatic image and a lower resolution (typically by a factor of 4) multi-spectral image. To get the best possible results, the required preprocessing is applied to the image: panchromatic and multi-spectral data are combined to get a high resolution 4-bands data set (pan-sharpening).

Contextual information is important to use to avoid obvious mistakes. For example, it is unlikely to find a house in

(a) Pan-sharpened data set originally at 0.6 m (resolution is degraded on the illustration due to copyright restriction). Courtesy of European Commission, Joint Research Centre, ISFEREA Research Action; Includes material ©2005, DigitalGlobe Inc., all rights reserved.



(b) Result of the detection algorithm, the number of detected buildings is 3600 which is an over detection: some big buildings are detected as several

Fig. 1. Input data and final results for the building counting experiment.

the middle of the water unless the goal is specifically to count houses flooded during a natural disaster. Similarly, boats are usually surrounded by water. This basic level information can be exploited by first creating a rough land cover classification. Classes such as water, vegetation, roads, shadows, bare soil and few ad-hoc classes provide a good starting point. To obtain this classification, we use a Support Vector Machine (SVM) classifier [4, 5] on a specific set of features such as the four spectral bands, the NDVI index, a local variance, and morphological profiles. This classification will be used as a mask to remove some obvious false alarms. Figure 2(a) shows the classification.

It is interesting to note that this classification can either be obtained training the system online – that is on the image we want to process – or using an off-line approach. Indeed, since the nomenclature of classes we are using (water, vegetation, roads, buildings) is rather generic and simple, a classifier can be trained using archive images.

The next step is to segment the pan-sharpened image in order to lower the complexity of the input data. The level of details available in high resolution images can have a strong negative effect at some stages of the processing: roof superstructures are irrelevant when trying to extract the whole building for example. The mean shift algorithm [6] provides an efficient way to simplify such images.

The segmented image is combined with the classification to remove irrelevant segments. This is the main step where some simple high level information concerning the object is introduced.

A very useful piece of information which allows to improve the results of the counting is the precise location of the object boundaries. Indeed, adjacent objects which are connected may induce an erroneous count. For this matter, we use the boundaries of the regions obtained in the segmented image in order to disconnect those objects. The boundaries are shown in figure 2(c).

Segments are vectorized to enable higher level post-processing whose goal is to adjust the detected object to the original pan-sharpened data (precise edge adjustment). This step fits the obtained polygons to the input data by introducing shifts to the position of the vertices in order to maximize the overlap with respect to the edges of the original image.

This post-processing step is not needed for object counting and it does not affect the counting result. However, sometimes it may be interesting to filter the detected objects using some of their attributes (shape, radiometry, etc.).

3. SIMPLIFIED VERSIONS

The algorithm described in the preceding section needs a supervised classification of the image to process. This step may be a critical point of the processing chain if an interactive approach is forecasted. It would be interesting to have an algorithm where an operator just clicks on several (2 to 5) exam-

(a) Rough classification of the area to provide contextual information. Three classes which do not contains buildings are clearly extracted: water, vegetation and roads. Several other classes were defined for the different types of buildings. This classification was obtained with a linear SVM classifier.

(b) Mean shift clustering of the Quickbird image.

(c) Object boundaries from the mean shift clustering

Fig. 2. Different steps of the processing for the building detection experiment.

ples of objects of interest in order to initialize the counting algorithm.

Two simplified versions of the object counting algorithm have been developed and are described below. These versions aim to provide a likelihood map of the regions containing objects of interest. They will be followed by the same steps than in the original algorithm, i.e., segmentation, vectorization, etc.

These simplified versions of the original algorithms give good results for the cases where the objects of interest are compact and have an homogeneous radiometry.

3.1. Spectral angle mapper

The idea here is to use the samples selected by the operator in order to build a spectral reference for the object of interest. For this matter, the set of pixels selected by the user will be averaged in order to have a reference spectrum which will be used to compute a spectral angle for all the pixels of the image to process.

3.2. One class SVM

Similarly to the spectral angle, the samples selected by the operator will be used in order to train a supervised classifier. The difficulty here is that the class of interest is easy to define, however, it is very difficult to be exhaustive in giving samples of all the other classes in the images. The solution to this problem here is using a one-class Support Vector Machine [7], where only the class of interest is populated. Therefore, the operator will only select samples of objects of interest as for the spectral angle approach.

3.3. Quality of the results

These simplified versions will achieve poorer performances than the complete algorithm described in section 2. Therefore, before choosing which approach to use it is necessary to understand in which cases they are likely to perform correctly.

As said above, the simplified versions are useful for compact, homogeneous objects. The spectral angle mapper is indicated where the spectral characteristics across different objects are stable. The SVM-based approach is more indicated when the objects to detect can have slightly radiometric differences. In this case, more samples are needed for the algorithm.

4. RESULTS

An example of results for the algorithm with classification is presented in figure 1(b). The input data is a pan-sharpened Quickbird image (60 cm. resolution and 4 spectral bands). The results have been obtained with a specific land use classification and a parameter setting which favors detection over false alarms. A main drawback of this setting is that large buildings where different parts of the roof have different illuminations, can be detected as different buildings.

This algorithm was evaluated in the frame of the *PRRS 2008 Algorithm Performance Contest* [2], for which only the input data was available. Therefore, the results were submitted as a blind test and no optimization of the parameters could be done with respect to a reference data. The evaluation was done by the contest organizers using a ground truth made up of 3065 buildings.

Several evaluation criteria where defined and not only the counting result was used. These criteria allow to analyze the degree of matching between the detected objects and the reference map in terms of geometrical overlap, over-segmentation, distances, etc. (see [2] for more details).

Among the 8 algorithms benchmarked in the contest, the algorithm presented in this paper showed a very good trade-off in terms of detection versus false alarms, but also in terms of matching the shape of the buildings in the ground truth.

As a final illustration for this paper, figure 3 shows the results obtained with one of the simplified versions of the algorithm. Here the spectral angle based approach was used to count tents in a refugee camp. The algorithm shows good results because these objects are compact, radiometrically homogeneous and similar between them. The operator only needed to select 4 examples to obtain these results.

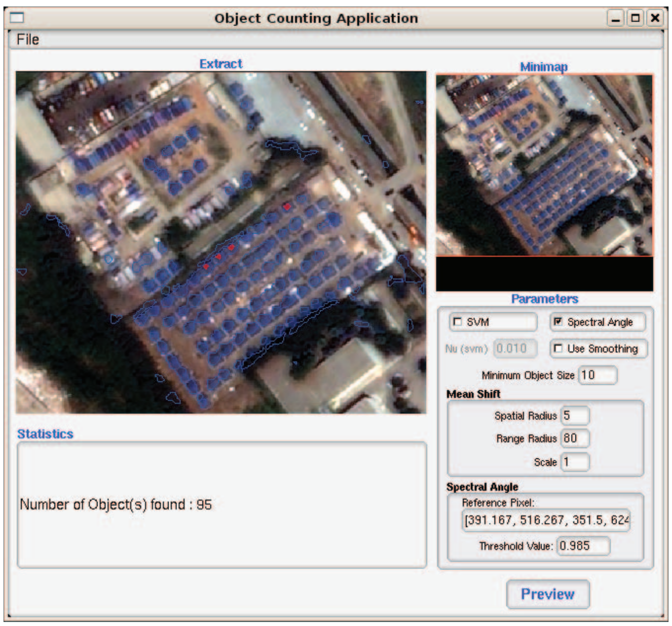


Fig. 3. Example of result using the spectral angle simplified approach.

The application shown in figure 3 is available in the Orfeo-Applications package, so that users can experiment with their own images. It provides the 2 simplified versions of the algorithm and offers an interesting interactivity, since the computations are done in real time in a small image ROI. When the operator is satisfied with the results obtained in the ROI, the processing can be applied to the whole image.

5. CONCLUSIONS

This paper presented an approach to object counting based on a supervised classification followed by a segmentation and some post-processing. All the developments were done using the Orfeo Toolbox library. Some new processing classes were added to the library in during the design of the algorithm.

The approach gives satisfactory results as showed in the *PRRS 2008 Algorithm Performance Contest*. The proposed procedure was further simplified in order to yield a fast and efficient user-in-the-loop approach. The 2 simplified versions of the algorithm were implemented as an Orfeo Application and is available for download at `http://www.orfeo-toolbox.org`.

6. REFERENCES

[1] Jordi Inglada, "Automatic recognition of man-made objects in high resolution optical remote sensing images by SVM classification of geometric image features," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 62, no. 3, pp. 236–248, Aug. 2007.

[2] Selim Aksoy, Bahadir Ozdemir, Sandra Eckert, Francois Kayitakire, Martino Pesaresi, Orsan Aytekin, Christoph C. Borel, Jan Cech, Emmanuel Christophe, Sebnem Duzgun, Arzu Erener, Kivanc Ertugay, Ejaz Hussain, Jordi Inglada, Sebastien Lefevre, Ozgun Ok, Dilek Koc San, Radim Sara, Jie Shan, Jyothish Soman, Ilkay Ulusoy, and Regis Witz, "Performance evaluation of building detection and digital surface model extraction algorithms: Outcomes of the PRRS 2008 algorithm performance contest," in *5th IAPR Workshop on Pattern Recognition in Remote Sensing, Tampa, Florida*, Dec. 2008.

[3] "The ORFEO toolbox software guide," http://www.orfeo-toolbox.org, 2009.

[4] F. Melgani and L. Bruzzone, "Classification of hyperspectral remote sensing images with support vector machines," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 42, no. 8, pp. 1778–1790, Aug. 2004.

[5] G. Camps-Valls and L Bruzzone, "Kernel-based methods for hyperspectral image classification," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 43, no. 6, pp. 1351–1362, June 2005.

[6] D. Commaniciu, "Mean shift: A robust approach toward feature space analysis," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 24, no. 5, pp. 603–619, May 2002.

[7] Yunqiang Chen, Xiang Sean Zhou, and T.S. Huang, "One-class SVM for learning in image retrieval," in *International Conference on Image Processing*, 2001, vol. 1, pp. 34–37.