

# SUPERVISED RE-SEGMENTATION FOR VERY HIGH-RESOLUTION SATELLITE IMAGES

*J. Michel, M. Grizonnet*

CNES, DCT/SI/AP  
18, avenue Edouard Belin, 31401  
Toulouse Cedex 09 - France

*O. Canévet\**

Telecom Bretagne  
Technopôle Brest-Iroise - CS 83818  
29238 Brest Cedex 3

## 1. INTRODUCTION

With the increase of spatial resolution in satellite images, segmentation has become a widely used pre-processing step in a lot of image analysis methods, such as object based image analysis for instance. For this purpose, a wide range of segmentation algorithms, either coming from other domains or dedicated to remote sensing [1, 2], have been proposed in the literature. Yet the segmentation of very high resolution remote sensing images remains a challenging task for mainly two reasons. Firstly, the variability of acquisition parameters such as lightning or sharpness across these wide images, as well as the complexity of the image content itself, makes it impossible to derive appropriate algorithms and parameters with stable segmentation performances for all objects and locations in the image. Secondly, the quality of the segmentation itself is a subjective notion, because it is highly related to the end-user definition of his objects of interest and their granularity. Segmented images are therefore suffering from two common defects called over-segmentation when the object of interest is divided into several pieces and under-segmentation when the object of interest is merged with its surroundings.

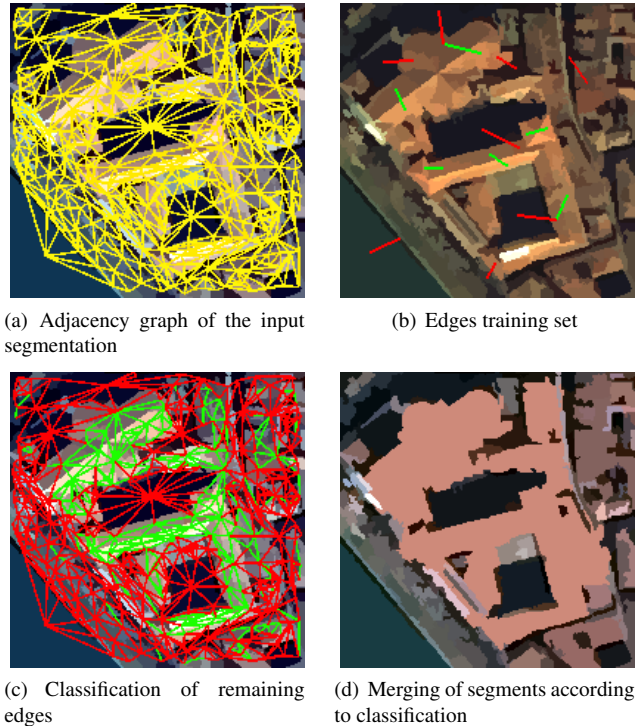
In this paper, we propose a methodology to enhance an existing segmentation for categories of objects of interest by lowering their over-segmentation in a supervised way. A classifier is trained to recognise pairs of adjacent segments belonging to the same object of interest, which are merged according to the classifier decision. Supervision of the classifier intends to capture the notion of object of interest from the end-user perspective, and is therefore a key aspect of the method, for which we propose three alternatives. This methodology as well as the performance evaluation process are detailed in section 2. Results of applying this methodology for different kind of scenes and parameters, with a qualitative and quantitative performance analysis, are shown in section 3.

## 2. PROPOSED METHODOLOGY

### 2.1. Overview

This section presents an overview of the proposed method, its four main steps being illustrated by figure 1. The input of our methodology is a segmentation of the image on which the only assumption we make is that it is broadly over-segmented, with little under-segmentation. This segmentation can therefore be produced by any segmentation algorithm with strict parameters, though we used the Mean-Shift [2] algorithm in our experiments.

This segmentation is represented by the classical OBIA paradigm [3], in which each segment is represented by a unique label and a



**Fig. 1:** The 4 steps of the methodology illustrated on an extract from a Mean-Shift segmentation of a pan-sharpened Quickbird scene over Toulouse, France.

set of features describing it. These segments are gathered in an adjacency graph where nodes are segments and edges represent adjacency between segments, meaning we can find a pair of pixels from the border of two distinct segments that are 8-connexity neighbours. Figure 1(a) shows an example of adjacency graph. The aim of our method is to remove edges between segments belonging to the same object of interest from this graph and to merge these segments. For this purpose, we will build a two-class or one-class training set of edges representing respectively edges to be removed (in one and two class mode) and edges to be kept (in two-class mode), as illustrated in figure 1(b). We designed three different supervision processes which are detailed in section 2.2.

Once the two-class training set has been collected, we describe each edge with a set of features and feed the samples into a SVM classifier [4, 5] for training (one-class training set is used with one-class SVM). The trained classifier is then used to classify the re-

\*Olivier Canévet performed the work during his 6 month internship at CNES

maining edges, as shown in figure 1(c). Section 2.3 explains how the samples are built and how the classifier parameters are chosen.

The last step is to merge adjacent segments whose edges have been marked as to be removed by the classifier, as presented in figure 1(d). Please note that this is a simple pair-wise sequential operation, and the result might not be unique in terms of segment labelling. In our case we decided to assign the lowest label to the resulting merged segment. Also note that this merging step can be in contradiction with the classifier decision, if a path of edges to be merged exist between segments that are directly linked by an edge to be kept. The amount of contradiction encountered during the merging step might be a good measure of the quality of the results.

## 2.2. Supervision

Since this method intends to capture the knowledge and definition of objects of interest from the end-user point of view, it is very important for the supervision process to be as efficient and ergonomic as possible.

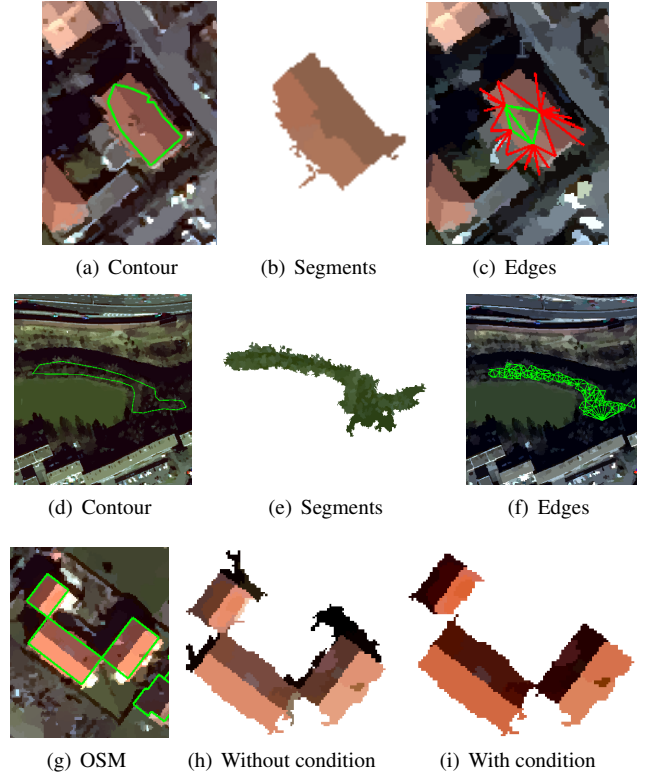
The naive solution would be to ask the user to select pairs of adjacent segments and label each of them manually. Unfortunately, this process is very demanding in terms of user actions, and has only a poor efficiency, which will lead to small training sets with poor representativity.

A second more user-friendly option is to perform group selection: the user selects all the segments composing a given object, and inner edges are labeled as to be kept, while outer edges are labeled as to be removed. This can be done either by segments selection or by drawing the contour of the object, as shown in figure 2. In case of objects with large spatial extent, like roads or rivers, this method is more suited to produce one-class training sets.

The last and most automatic method is to use a GIS database layer from an online database such as Open Street Map [6], if such a layer is available for the object of interest. In this case we can substitute polygons from the GIS layer to the manual contour of the previous method, which allows to get a large amount of training samples with minimal interaction. These training samples are not biased by interpretation of the image, so they also tends to enforce diversity. However, this method suffers from two drawbacks. First, it is likely that the GIS database and the image will correspond to different frames of time, and therefore this method may include in the training set edges corresponding to structures that did not exist at the acquisition time, or miss newly built structures. Second, the GIS database objects have perfect contours that may not fit the actual object in the image because of small mis-registration or imperfection of the actual object. To handle the latter, we added an extra condition such that a region is considered inside the object if its overlap ratio with the contour is greater than a given threshold. An example of supervision by Open Street Map layer is presented figure 2.

## 2.3. Edges samples and learning

Each edge of the adjacency graph is described with a set of features to form samples for the purpose of classification. These features describe each of the segment composing the edge with classical OBIA attributes such as bands statistics and shape features like surface, perimeter and compacity, and we also added features to characterise the relationship between these two segments : the spectral angle between the mean spectral signatures, an euclidean radiometric distance and a measure of the interlocking of the two segments. From one edge, two different samples are generated by mirroring the nodes attributes, in order to enforce that the adjacency graph is undirected.



**Fig. 2:** Examples of two-class manual group selection on a roof (figures 2(a) to 2(c)), one-class manual group selection on a forest (figures 2(d) to 2(f)), and automatic Open Street Map [6] selection on a roof with or without the additional condition (figures 2(g) to 2(i))

The learning and classification is performed with a SVM algorithm [4, 5]. For this purpose, samples are centered and reduced with respect to their statistics on the whole set of edges. With a two-class training set, learning is SVM type is CSVC with a RBF kernel, and the  $C$  and  $\gamma$  parameters are optimised with cross-validation. The one-class training set is an interesting option because it is less demanding in terms of supervision. In this case, the trained SVM algorithm is of type one-class with a RBF kernel, and the  $\nu$  and  $\gamma$  parameters are manually set.

## 2.4. Performances evaluation

Evaluating segmentation performances is a complex task for which a lot of methods have been proposed in the literature. In our case, we are interested in quantifying the possible improvement or deterioration induced by the merging of segments between the input and output images, which can be done by comparing the machine segmentation to a ground truth.

The metrics introduced by Hoover in [7] are based on an overlapping matrix  $O$ , which coefficient  $O_{i,j}$  corresponds to the number of common pixels between region  $i$  of the ground truth and region  $j$  of the machine segmentation. Then, given an overlapping threshold ( $0.5 < t \leq 1$ ), Hoover method builds correspondances (called instances) between the regions of the two segmentations depending on whether they overlap with each other, leading to correct detection, over- and under-segmentation, missed or noise instances.

In the context of satellite images, it is difficult to obtain an ex-



haustive ground truth, and building one by hand is a time consuming task. Moreover, the user may only be interested in some objects of the scene, in which case a full ground truth is useless. Therefore, we decided to evaluate the segmentation with a partial ground truth, in which some objects are segmented, the rest being considered background.

To get a global score on the image, we used the method proposed in [8], which consists in computing scores for all Hoover instances built between the machine segmentation and a partial ground truth. Overall scores are then computed by averaging the partial ones weighted by their size. For a given segmentation, four scores are available showing how much the ground truth is well detected, fragmented, merged or missed.

### 3. RESULTS

This section presents results of the proposed methodology applied to the segmentation of Quickbird and WorldView-2 pan-sharpened scenes. We focused on buildings, roads and vegetation areas and choosed to apply the methodology for each kind of objects independently because it significantly improves the performances and allows to select features and SVM modes adapted to each case. Buildings were handled with a two-class SVM model and radiometric features only, whereas roads and vegetation areas were handled with radiometric features as well as size ratio and interlocking and a one-class SVM. The merging process is then applied sequentially for each object category.

Figure 3 shows the Hoover scores on both building ground truth from OpenStreetMap and segmented image before and after the merging methodology is applied. Please note that the training of the SVM classifiers has been done on another part of the same scene. We can see that the initial segmentation mostly fragments buildings and misses a few of them. After the merging methodology is applied, the amount of well detected buildings increases greatly, even if more buildings are marked as not detected. The overall good detection Hoover index for the segmentation raises from 0.03 to 0.25, while the fragmentation score lowers from 0.7 to 0.3. The missed index increases from the 0.15 to 0.32, which is consistent with the qualitative evaluation.

Figure 4 shows other results from the same scene, focused on the center of the city of Toulouse, France. In such a dense urban area, using OpenStreetMap as a ground truth to evaluate buildings segmentation seems to underestimate the performances, mainly because contiguous building with administrative boundaries are segmented in a single piece, while clearly separated on the ground truth, and because the ground truth does not account for courtyards which are considered as a part of the building itself. We therefore rely on qualitative evaluation, and we can see by comparing figures 4(a) and 4(b) that most of the buildings blocks as well as the bridges are well delineated by the proposed method.

### 4. CONCLUSION

In this paper, we proposed a supervised methodology to enhance an existing segmentation in which we assume that objects of interest are mainly fragmented. We used a SVM classifier to classify edges from the adjacency graph of the initial segmentation, described with features on the pair of segments and their relationship. Pairs of segments are then merged sequentially according to the classifier decision. We also proposed three methods for efficient supervision by the end user.



(a) Initial segmentation contours

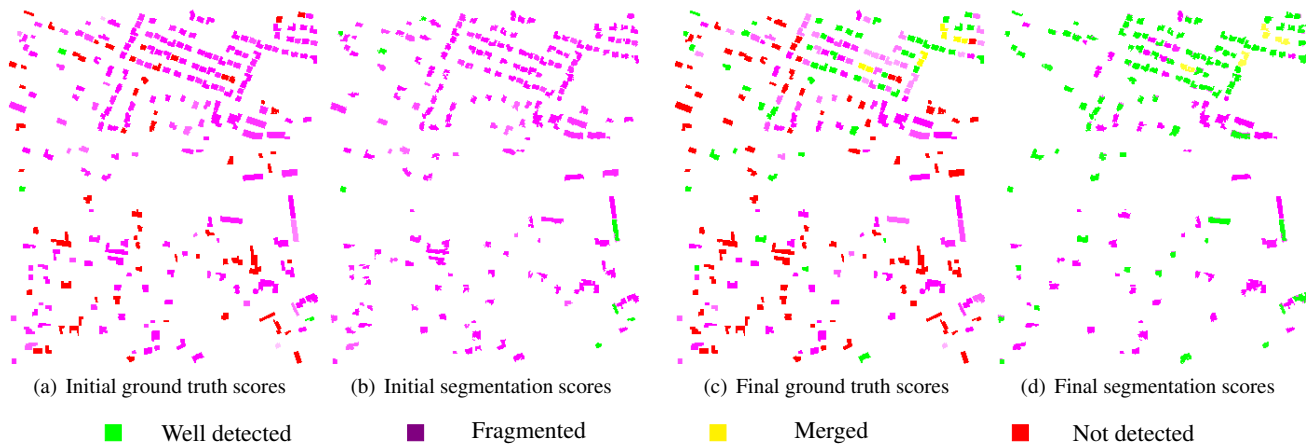


(b) Final segmentation contours



(c) Final segmentation clusters

**Fig. 4:** Segmentation of a pan-sharpened Quickbird scene over the center of Toulouse, France, before and after applying our methodology



**Fig. 3:** Hoover scores on building ground truth from OpenStreetMap and segmentation images before and after applying the proposed methodology to process the initial MeanShift segmentation of a pan-sharpened Quickbird scene of the city of Toulouse, France

The improvement of the segmentation quality is quantified with respect to a GIS ground truth from OpenStreetMap for buildings. Evaluation was conducted both quantitatively by using a derivation of Hoover metrics and qualitatively with a color-mapping of the Hoover instances and visual inspection. Our results on Quickbird and Worldview2 pan-sharpened scenes show that this method allows to significantly improve the amount of well-delineated objects, by reducing fragmentation. Hoover scores also show that the amount of missed objects slightly increases.

This method is of great interest as a pre-processing step for Object Based Image Analysis methods [9] and high-level spatial reasoning [10], for which segmentation quality is a major limitation. As we are able to evaluate quantitatively the performances of our method, a future work would be to optimise the whole process against the detection score for a given class of objects of interest.

This work has been done using the Orfeo ToolBox ([www.orfeo-toolbox.org](http://www.orfeo-toolbox.org)), and will be available for testing in a future release.

## 5. REFERENCES

- [1] J. Shi and J. Malik, “Normalized cuts and image segmentation,” *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 22, no. 8, pp. 888–905, 2000.
- [2] D. Comaniciu and P. Meer, “Mean shift: A robust approach toward feature space analysis,” *IEEE Transactions on pattern analysis and machine intelligence*, pp. 603–619, 2002.
- [3] G. Lehmann, “Label object representation and manipulation with itk,” *Insight J*, pp. 1–34, 2008.
- [4] B.E. Boser, I.M. Guyon, and V.N. Vapnik, “A training algorithm for optimal margin classifiers,” in *Proceedings of the fifth annual workshop on Computational learning theory*. ACM, 1992, pp. 144–152.
- [5] Chih-Chung Chang and Chih-Jen Lin, “LIBSVM: A library for support vector machines,” *ACM Transactions on Intelligent Systems and Technology*, vol. 2, pp. 27:1–27:27, 2011, Software available at <http://www.csie.ntu.edu.tw/~cjlin/libsvm>.
- [6] Openstreetmap France, “Openstreetmap,” <http://www.openstreetmap.fr/> [online], 2011.
- [7] A. Hoover, G. Jean-Baptiste, X. Jiang, P.J. Flynn, H. Bunke, D.B. Goldgof, K. Bowyer, D.W. Eggert, A. Fitzgibbon, and R.B. Fisher, “An experimental comparison of range image segmentation algorithms,” *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 18, no. 7, pp. 673–689, 1996.
- [8] P. Daum, J.L. Buessler, J.P. Urban, et al., “évaluation de segmentations d’images: mesures de similarité avec une référence partielle,” 2009.
- [9] J. Michel, J. Malik, and J. Inglada, “Lazy yet efficient land-cover map generation for hr optical images,” in *Geoscience and Remote Sensing Symposium (IGARSS), 2010 IEEE International*. IEEE, 2010, pp. 1863–1866.
- [10] J. Inglada and J. Michel, “Qualitative spatial reasoning for high-resolution remote sensing image analysis,” *Geoscience and Remote Sensing, IEEE Transactions on*, vol. 47, no. 2, pp. 599–612, 2009.