Semantic Rich ICM Algorithm for VHR Satellite Images Segmentation

Jérémie Sublime[*,***], Andrés Troya-Galvis[**], Younès Bennani[***], Antoine Cornuéjols[*], and Pierre Gançarski[**]

[*]AgroParisTech, UMR INRA MIA 518, 16 rue Claude Bernard, 75231 Paris Cedex 5, France,
{*jeremie.sublime,antoine.cornuejols*}*@agroparistech.fr*

[**]Université de Strasbourg, ICube UMR CNRS 7357 , 300 bd Sébastien Brant - CS 10413 - F-67412 Illkirch Cedex, France,
{*troyagalvis,gancarski*}*@unistra.fr*

[***]Université Paris 13, LIPN UMR CNRS 7030, 99 av. JB Clément, 93430 Villetaneuse, France,
younes.bennani@lipn.univ-paris13.fr

Abstract

In this article we show some applications of a MRF-based segmentation algorithm applied to real data extracted from a very high resolution image. This algorithm has specific features that enable the extraction of semantic information on the clusters in the form of affinity and geographic position properties. The results of the experiments conducted on this data set are interesting both in terms of clustering quality when using common unsupervised learning quality indexes, but also when compared to a ground-truth based on expert maps.

1 Introduction

With the booming number of available satellite images and data, the automatic interpretation of remotely sensed images has become an increasingly active domain. With sensors now capable of getting images with a very high resolution (VHR) on a large spectral resolution, it has become increasingly difficult to design algorithms and methods capable of efficiently processing such data in a reasonable amount of time.

The segmentation of such images can be achieved by using a Markov Random Fields based representation of the images' data [1]. Markov Random Fields (MRF) rely on the notion of neighborhood to represent the dependencies between two neighbor pixels or superpixels. In the pixel model, the goal of MRF-based segmentation is to determine the cluster to which each pixel belongs. In the super-pixel model, the image is preprocessed and pixels are regrouped by patches based on nearly flat gradient areas. It is then these groups of pixels, called superpixels, that are to be assigned to a cluster.

Nowadays, the pixel model is less and less common as single pixels have no signification in VHR images. Thus, superpixel approaches, also known as Object Based Image Analysis (OBIA) [2] methods are preferred.

There are several algorithms capable of achieving the segmentation of images represented as a MRF network : The Graph-Cut Algorithm [3], the Integer Projected Fixed Point method [4], the Graduated Non-Convexity and Concavity Procedure [5], and the Iterated Conditional Modes (ICM) [6].

While it is true that the Iterated Conditional Modes is often considered to be the less effective of these MRF-based segmentation algorithms, it is also the fastest one. In addition to that, in the case of very high resolution satellite pictures using the superpixel model, the data have already been pre-processed and are much less noisy than in the pixel-based model, thus making the ICM algorithm fine enough. Furthermore, the speed of this algorithm is a considerable asset when dealing with huge images with a lot of superpixels to be segmented, especially when these superpixels may have more than the 3 traditional Red, Green and Blue attributes and might have potentially more than four neighbors if their shapes are irregular. Algorithms such as the Integer Projected Fixed Point method, the performances of which are far superior to those of the ICM, are for example way too slow to tackle such data.

It is also worth mentioning that the ICM algorithms has had several improvements since its original version by J. Besag in 1986 :

- It has been adapted to the Gaussian Mixture Models and is now widely used in combination with the EM Algorithm [7].
- Its energy model has been modified to enhance its performances [8]. As a positive side effect of this work, previously unavailable semantic information can now be retrieved during the clustering process.

In this article, we use the latest version of the ICM algorithm and apply it to the segmentation of a data set extracted from a very high resolution image, and we explain how we used the specificity of this algorithm to extract information about the clusters affinities and geographical relations that are not available with other algorithms.

The rest of this article is structured as follows: In section 2, we present our version of the ICM algorithm and its application for semantic information extraction. In section 3, we introduce our data set and the expert ground-truth that we used to validate our

results. In section 4, we show the results of our experiments. Finally, in section 5 we will give a brief conclusion and some future perspectives to this work.

2 Semantic Rich ICM algorithm

We now consider a set of random variables $S = \{s_1, ..., s_N\}, s_i \in 1..K$ that are linked to each other by neighborhood dependencies. These variables represent the unknown states of the superpixels. And we suppose that these superpixels are described by observables data $X = \{x_1, ..., x_N\}$ where the $x_i \in \mathbb{R}^d$ are the vectors containing their attributes. The goal is then to determine the optimal configuration for S. Doing so requires to infer the $s_i \in 1..K$, which is often achieved by using the maximum a posteriori criterion (MAP) to determine S such as:

$$S = \arg\max_S (P(X|S,\Theta)P(S)) \quad (1)$$

$$P(X|S,\Theta)P(S) = \prod_t P(x_t|s_t,\theta_{s_t})P(s_t) \quad (2)$$

We consider that $P(x_t|s_t,\theta_{s_t})$ follows a Gaussian distribution of parameters $\theta_s = (\mu_s, \Sigma_s)$, where μ_s is the mean vector of cluster s, and Σ_s its covariance matrix.

Approaching a solution for this equation requires to locally optimize an energy function deriving from the logarithm of $P(x|s,\theta)P(s)$. This energy function features a local energy term deriving from $P(x|s,\theta)$, and a neighborhood energy term deriving from $P(s)$ that evaluates the probability of a state considering the neighboring superpixels. In the semantic rich version of the ICM algorithm (SR-ICM), this energy function is as follows:

$$U(s,x) = \frac{1}{2}(x-\mu_s)^T\Sigma_s^{-1}(x-\mu_s) + log(\sqrt{|\Sigma_s|(2\pi)^d}) - \sum_{v\in V_x} \tau_{x,v} \times log(a_{s_v,s}) \quad (3)$$

In the previous Equation, V_x is a vector with all the neighbors of the currently observed superpixel. $\tau_{x,v}$ is the percentage of border occupied by neighbor v regarding to the observed superpixel x. $A = \{a_{i,j}\}_{K\times K}$ is a neighborhood compatibility matrix, where each $a_{i,j}$ is the probability of having a neighbor in state j for a superpixel labeled in state i. Here, we note s_v the current state of neighbor v.

In this algorithm A is computed a posteriori after each iteration of the SR-ICM algorithm.

Algorithm 1: Semantic Rich ICM Algorithm

```
Initialize S and Θ with the EM algorithm
Initialize A
while Tr(A) is increasing do
    for each x ∈ X do
        Minimize U(s, x)
    end
    Update A from the new distribution S
end
return S
```

This version of the ICM algorithm is said to be "semantic rich" because of the information that can be extracted from the neighborhood matrix A. The diagonal terms give information on which clusters are compacts or not. Non diagonal terms can be extracted and translated into rules such as: "Elements of cluster B are almost never in contact with elements of cluster C", or "Elements of cluster B are often surrounded by elements of cluster D".

3 VHR Strasbourg Data Set

3.1 Data

In this article we use a data set made from a very high resolution image of the French city of Strasbourg, $1px = (50cm)^2$, an extract of which is shown in Figure 1.

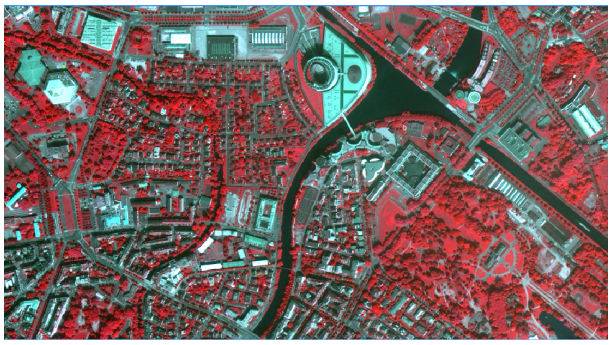


Figure 1. Extract of the original source image (approximately 1/20 of the full image), Pleiades ©CNES2012, Distribution Astrium Services / Spot Image S.A., France, All rights reserved.

This image has been preprocessed into a data set made of 187058 superpixels, each of them described by 27 attributes either geometrical or radiometrical [9]. These attributes include the geographic position of the superpixel, the surface of the area covered by the superpixel, the mean RGB values, the constrast compared to neighbor pixels and superpixels, the brightness, and the standard deviations, among others.

In addition to this information, this data set provides the neighborhood dependencies between the superpixels: number of neighbors, id number of neighbor superpixels, and relative percentage of shared border.

The superpixels in this data set have highly irregular shapes (cf. Figure 2), and consequently the neighborhoods themselves are also irregular. Unlike in the pixel model where 1st order neighborhood usually include 4 or 8 neighbors, in this data set each superpixel can have 1 to 15 neighbors depending on its shape and position.

3.2 Ground truth

In order to validate our results, we had to find a ground truth. It was however too tedious a task to manually label the 187058 superpixels. Therefore we decided to rely on maps of the area made by expert geographers (cf. Figure 3(a)). These maps were produced by a hybrid methodology, mixing data from topographic databases for roads and buildings, a supervised classification for different types of water and veg-

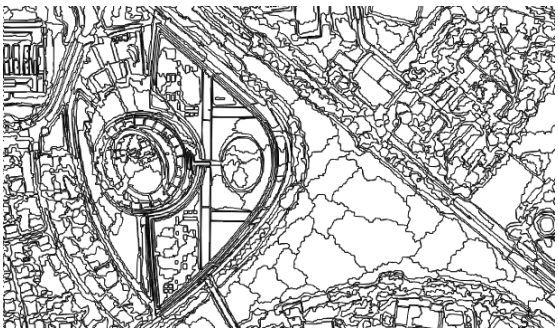
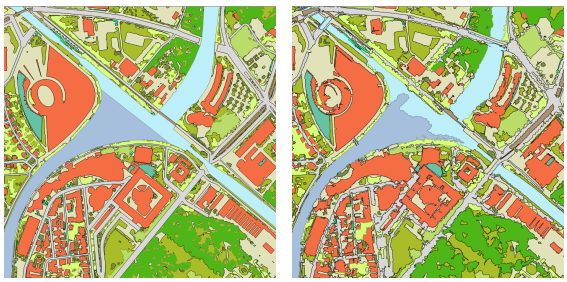
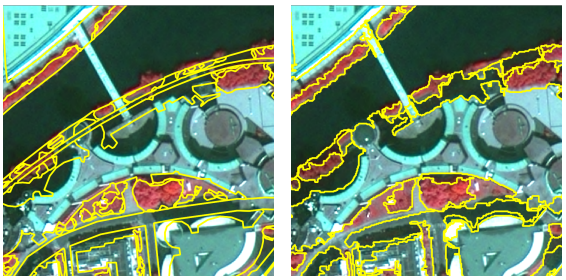


Figure 2. Exemple of superpixel segmentation in the central area of Strasbourg.

etation, as well as further manual refinement in order to reduce classification errors.



(a) Original ground truth. (b) Modified ground truth.



(c) Zoomed boundary representation of the original ground truth maps. (d) Zoomed boundary representation of the modified ground truth.

Figure 3. Extract of the original and improved ground-truth images

A closer inspection of the provided ground truth shows that the segments boundaries do not align well with the superpixels boundaries in the image data (cf. Figure 3(c)). Thus, the original ground truth data set was not suitable to assess the results of our method which considerably relies on radiometrical attributes. Consequently, we decided to improve the ground truth by projecting the superpixel segmentation onto the original maps and using a majority vote approach to determine which expert label should be chosen depending on the covering surface of the superpixel.

It is important to mention that some of the labels provided by the expert geographers are very unlikely to be mapped to a single cluster by an unsupervised algorithm. Such labels include various types of vegetation and forest areas based on their density and the total size they cover, 2 types of water bodies, and different types of buildings, some of them being difficult to distinguish from the sky.

As one can see in figures 3(b) and 3(d) the modified ground truth data corresponds more accurately to the image data. Although, there are still a few errors due to the superpixel segmentation which is not optimal. We are working towards ways of evaluating the ground truth quality in order to better assess the quality of our own results.

The expert geographers provided 15 different classes, this number could be reduced to 7 to 9 classes by regrouping those that are very similar or technically impossible to distinguish for an unsupervised algorithm.

4 Experimental results

In this section, we show our MRF-based segmentation results on the VHR Strasbourg data set. To evaluate the quality of our results, we have been using two different quality measures :

- The Davies-Bouldin Index [10] : An unsupervised quality measure that evaluates the quality of the clusters based on their internal variance and the distance between the prototypes of the different clusters. For this index a lower value is better.
- The Rand Index [11] : This index compares how close two segmentations are. In the context of these experiments we compared our results with our ground-truth made from from expert maps. For the Rand Index, a value close to 1 means a 100% similarity.

In Table (1), we give the results of 3 different segmentations searching for 7, 8 and 9 clusters. As one can see, the Rand Index results compared with the ground truth are quite good (around 80% of similarity), with the 7-cluster segmentation being the closest one to the ground-truth according to the Rand Index, quickly followed by the 9-cluster segmentation which is the best one from a clustering point of view (lowest Davies-Bouldin Index).

These results are surprisingly good given the fact that the ICM is unsupervised, and thus it tends to prove the effectiveness of the semantic rich version of this algorithm to process this type of satellite images. Four extracts of the 9-cluster segmentation are shown in Figure 4.

Table 1. DB Index and Rand Index results for different number of clusters (100 simulations)

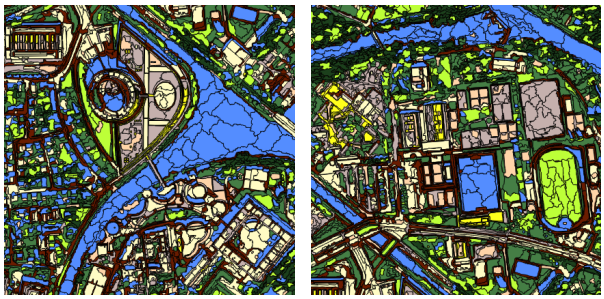
SR-ICM	DB Index	Rand Index
7 clusters	2.756 ± 0.31	0.8264 ± 0.040
8 clusters	3.329 ± 0.24	0.7959 ± 0.017
9 clusters	3.112 ± 0.27	0.8168 ± 0.025

It is however important to emphasize that all 3 segmentations suffer from problems that are very common in image segmentation. Regardless of the 27 attributes, it seems that the original colors still are the most important factor. For instance water bodies, shadow areas and dark roofs are often grouped in a single cluster. With our current knowledge of this data set, it is difficult to say whether this problem comes from the segmentation algorithm, from the data preprocess that created superpixels from these shadow areas, or a bit of both. These segmentation defaults are quite easy to spot on full scale color versions of the image extracts shown in Figures 4(a) and 4(c).

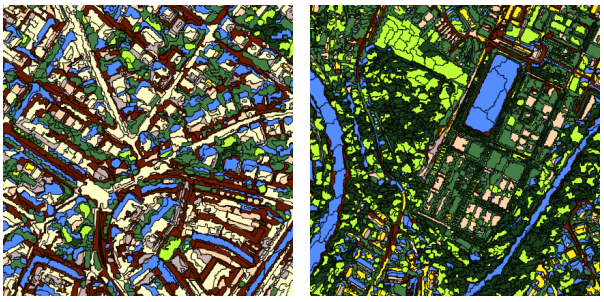
On the semantic side, after linking each cluster to the corresponding expert label from the visual results, the following properties were found :

- There is a strong neighborhood connection between the tree areas and grass areas (dark green and light green respectively), with a transition probability of ≈ 0.65 from a tree superpixel to a grass superpixel. This interesting property from the neighborhood matrix A could be translated as "Trees are often surrounded by grass".
- There is a double side low transition probability (< 0.02) from modern urban buildings superpixels (in yellow) to water superpixels. It matches with the fact that modern buildings are rarely built directly adjacent to a river, nor in the middle of it.
- The water cluster has an average transition probability to itself (≈ 0.47) . This is consistent with the river and water bodies of the city of Strasbourg being rather linear and small.

While these results may seem obvious, considering that they are coming from an unsupervised algorithm that has no external knowledge on the true nature of each cluster, they are still quite impressive.



(a) European Parliament (b) A Sport complex



(c) Strasbourg residential area (d) A graveyard

Figure 4. Extracts of the 9-cluster segmentation

Table 2. Computation times for different number of clusters

Cluster Number			Computation time
SR-ICM	7	clusters	27630ms
SR-ICM	8	clusters	31495ms
SR-ICM	9	clusters	34438ms

Finally, in Table (2), we give the computation times associated with the previous experiments for 10 iterations of the EM algorithm followed by the SR-ICM algorithm iterating until convergence. These computation times were acquired using an i5-3210M 2.5GHz processor. The data set processed by the algorithms weighted around 60MB and included : the 187058 superpixels' 27 attributes and their neighborhood graph.

As one can see, the computation times are quite low given the size and complexity of the data set as well as the quality of the results.

5 Conclusion

In this article, we have shown an application of a semantic rich Iterated Conditional Modes algorithm tuned to perform the segmentation of a very high resolution satellite image. The results of our experiments have shown this algorithm to be suited and fast enough to process such images. Furthermore the quality of the semantic approach in the context of unsupervised learning has also been validated.

In our future works we plan to improve and extend this algorithm to apply it on other images.

Acknowledgments

This work has been supported by the ANR Project COCLICO, ANR-12-MONU-0001.

References

[1] S. Roth, and M. J. Black.: "Fields of experts" *Markov Random Fields for Vision and Image Processing*, MIT Press, pp.297–310, 2011.

[2] T. Blaschke.: "Object based image analysis for remote sensing" {*ISPRS*} *Journal of Photogrammetry and Remote Sensing*, pp.2–16, 2010.

[3] Y. Boykov, et al.: "Fast approximate energy minimization via graph cuts" *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 23(11), pp.1222–1239, 2001.

[4] M. Leordeanu, et al.: "An integer projected fixed point method for graph matching and map inference" *NIPS*, 2009.

[5] Z. Y. Liu, et al.: "MAP Inference with MRF by Graduated Non-Convexity and Concavity Procedure" *ICONIP 2014, Part II, Lecture Notes in Computer Science*, volume 8835, pp.404–412, 2014.

[6] J. Besag: "On the Statistical Analysis of Dirty Pictures" *Journal of the Royal Statistical Society*, Series B (Methodological) 48(3), pp.259-302, 1986.

[7] Y. Zhang, et al. : "Segmentation of brain mr images through a hidden markov random Field model and the expectation maximization algorithm" *IEEE Transaction on Medical Imaging*, volume 20, pp.45–57, 2001.

[8] J. Sublime, et al.: "A new energy model for the hidden markov random Fields" *ICONIP 2014, Part II, Lecture Notes in Computer Science*, volume 8835, pp.60–67, 2014.

[9] S. Rougier, and A. Puissant: "Improvements of urban vegetation segmentation and classification using multitemporal Pleiades images" *5th International Conference on Geographic Object-Based Image Analysis*, p.6, 2014.

[10] D. L. Davies and D. W. Bouldin: "A cluster separation measure" *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Volume 1(2), pp.224–227, 1979.

[11] W. Rand.: "Objective criteria for the evaluation of clustering methods" *Journal of the American Statistical Association*, pp.846–850, 1971.