A strategy to improve quality and speed of multi-scale high-resolution image analysis through superpixels

Angel García-Pedrero[1], Consuelo Gonzalo-Martín[2], Mario Lillo-Saavedra[3], and Dionisio Rodríguez-Esparragón[4]

[1]*Universidad Politécnica de Madrid, Facultad de Informática, DATSI, Madrid, Spain; am.garcia@alumnos.upm.es*
[2]*Universidad Politécnica de Madrid, Centro de Tecnología Biomédica, Madrid, Spain; consuelo.gonzalo@upm.es*
[3]*Universidad de Concepción, Facultad de Ingeniería Agrícola, Departamento de Mecanización y Energía. Chillán, Chile; malillo@udec.cl*
[4]*Universidad de Las Palmas de Gran Canaria, Instituto de Oceanografía y Cambio Global, Las Palmas de Gran Canaria, Spain; drodriguez@dsc.ulpgc.es*

Abstract. With the advent of higher-resolution remote sensing imagery the need for more efficient multi-scale image analysis methods has grown more than ever. This paper presents a superpixel-based methodology for hierarchical image analysis. The use of superpixels as pre-processing block reduces the execution time of the analysis and also provides a good starting point to build meaningful objects at different scales through a merging process. At each scale a set of most representative features are extracted, and then regions of interest are identified through a classification method. Only classes of interest are analyzed in a later segmentation step.

Keywords. Superpixel, object-based image analysis, multi-scale analysis.

1. Introduction

Object-based image analysis (OBIA), in a geographic context, is a newly developed technique that outperforms very high-spatial resolution (VHSR) image analysis when it is compared with pixel-based approaches [1]. Broadly speaking, the OBIA approach starts with a segmentation process, followed by successive analysis, usually at different hierarchical levels (scale) in order to create relationships between objects and super-objects [2]. Despite the advances in OBIA methodologies, there are still many issues to resolve before achieving an unsupervised image analysis, due mainly to the fact that these types of methods depend on image segmentation.

Even though segmentation methods are commonly used in image-processing tasks, they have some drawbacks. On the one hand, since segmentation is an ill-posed problem, it has no unique solution [3]. Therefore, in real world problems erroneous segments are inevitably detected while genuine segments are omitted. On the other hand, the computational cost of segmentation operation should be considered when a VHSR image is analyzed. For instance, depending on the computational complexity of the algorithm and the tuning of its parameters, obtaining a good high-resolution image segmentation could take several hours with a standard computer. In practice, when choosing a method of segmentation, it is necessary to establish a trade-off between efficacy and efficiency.

In recent years, new types of image segmentation methods, known as superpixel methods [4], have been developed in the area of computer vision. A superpixel is a small, local, and coherent cluster that contains a statistically homogeneous image region according to certain criteria [5] such as color, texture, among others. Superpixels reduce the influence of noise, preserve most edges of images, are approximately uniform in size and shape, and improve the computational speed of later steps. Superpixels are a form of image over-segmentation that aims to represent the image in a more useful manner, since this representation reduces the intra-class spectral variability of the pixels. In this context, it is assumed that every superpixel belongs completely to a single meaningful object; but an object can be composed by several superpixels.

Despite its benefits in areas such as computer vision, superpixel methods still have not been completely explored in remote sensing applications. Only a few papers can be found in literature. In [6], superpixels were employed as basic pre-processing blocks for change detection for QuickBird images. Superpixel-based classifications were also performed over aerial images [7] and Polarimetric SAR (PolSAR) imagery [8]. In [9], superpixels are used as operation units in order to segment PolSAR images in a hierarchical manner. Segmentation at each level is made through a merging process where adjacent superpixels are joined according to similar criteria based on contextual information.

In this paper, a novel superpixel/object-based approach is proposed for multi-scale image analysis. The approach consists of two main stages: 1) Superpixel segmentation: A multi-spectral image is firstly segmented into a uniformly distributed compact and homogeneous map of superpixels. These superpixels are considered the basic processing unit of the analysis; and 2) Hierarchical object analysis: A region adjacency graph representation (RAG) is made using the map of superpixels. This representation allows the use of neighborhood relationships and also facilitates the merging/splitting adjacent regions. This two-step process represents the zero level of the complete hierarchical structure. A merging process based on specific features for each scale creates later structure levels.

The proposed approach is expected to take advantage of superpixels and OBIA characteristics in order to improve speed and quality of VHSR image analysis. First, the superpixels preserve the statistical characteristics of homogeneous pixels. Second, it considers the superpixel as the smallest processing block, which reduces the number of instances to analyze and allows the use of more complex analysis methods. And finally, the OBIA paradigm is applied in order to generate a hierarchical structure for image analysis.

2. Methods and Materials

2.1. Dataset

The data used in this work consist of an image registered by the WorldView-2 sensor covering a cropland region of Valparaiso Region, Chile. Coordinates of upper left corner of the image are (70º33’57’’W, 32º50’42’’S). The scene has an area of 104.86 ha and corresponds to 2048 x 2048 pixels for the panchromatic image. In order to improve the quality of the analysis the multispectral image was pansharpened using the robust WATxFRAC algorithm [10]. Fig. 1 shows a real color composition (bands 5, 4, and 2) of the pansharpened image.

2.2. Methodology

As it has been previously mentioned, the proposed methodology consists of two phases: superpixel preprocessing and, a hierarchical object analysis, where the last step is subdivided into three sub-steps. An external prior knowledge about of the most representative features of an object in each scale is incorporated. For instance, NDVI is useful to identify vegetation in incipient levels, but in later scales features as elongation could be of more help to identify rows of crops. This type of feature knowledge can be obtained with the help of an expert or through a machine learning process.

The workflow is shown in Fig. 2. And it will describe in later sub-sections.

Figure 1. Real color composition (bands 5, 4, and 2) of the pansharpened image.

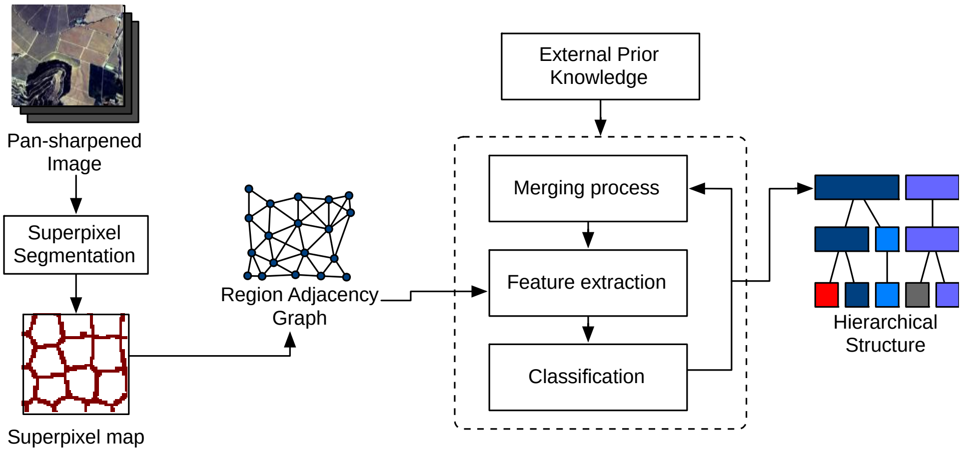


Figure 2. The proposed methodology.

2.2.1. SLIC Superpixel Algorithm

In the first step of the processing, the super linear iterative clustering (SLIC) algorithm [4] is used to segment a multispectral image to generate a superpixel map. SLIC is an efficient clustering technique based on the well-known *K*-means method. The algorithm has a computational complexity that is linearly proportional to the number of image pixels.

SLIC starts sampling *K* regularly spaced cluster centroids, followed by an iterative clustering process. The clustering distance is a weighted relationship between color and spatial measures. The first measure ensures the superpixel homogeneity, and the second one enforces compactness and regularity in superpixels shape. The color distance between pixels m and n is calculated as follows:

$$d_c = \sum_{i=1}^{B} (p_m^i - p_n^i)^2 \tag{1}$$

where B denotes the number of spectral bands, and p^i represents the digital number in the i–th band. The spatial distance is defined as:

$$d_s = (x_m - x_n)^2 + (y_m - y_n)^2 \tag{2}$$

where x and y denote the position of the pixel. And then the clustering distance is defined as:

$$d_w = \sqrt{d_c} + \sqrt{\left(\frac{m}{s}\right)^2 d_s} \tag{3}$$

where s is the sampling interval of the centroids, and m controls the compactness of superpixels. A larger value of m emphasizes the importance of spatial proximity resulting in more compact superpixels. On the other hand, a larger value of s increases the superpixels size.

2.2.2. Hierarchical Object Analysis

In order to create a hierarchical structure, a region adjacency graph (RAG) representation is generated from a superpixel map. This graph representation allows defining ofe merging and splitting operations to make an incremental construction resulting in a hierarchical final representation.

Starting from a superpixel RAG level, a merging process is made to create a next level of analysis. This process intends to be directed through an external source of knowledge that must contain a set of the most important features that conceptually define the object of interest at a particular scale.

As shown in Fig. 2, this phase includes three steps: feature extraction, classification, and splitting process. Each step is described below:

- Feature extraction is one of the most important steps in OBIA process. Several features can be extracted from objects, however depending on the scale of analysis some features could be more relevant than another. For this reason, diverse features are extracted from segments at each scale of the hierarchical structure and they are employed to generate new objects in a posterior level.
- Classification is carried out to detect regions of interest; this step is relevant when large images are analyzed therefore instead of analyzing the whole image only a reduce region marked as relevant is dissected in more detailed in next level. From this, the analysis is considered to be target specific, which significantly reduces the processing time.
- Merging process is done, only over regions of interest, according to the set of features defined for each level, a set a similarity measure between node u and its neighbor v is calculated. If the distance is greater than a specified threshold th, the nodes are merged.
  In next level, the new regions are intended to merging by following the above procedure but with another set of features.
  The details of the hierarchical merging algorithm are described in Algorithm 1.

Algorithm 1 Hierarchical Object Analysis algorithm

1. Construct a RAG R from superpixel map. This is the level 0
2. For $i \leftarrow 1$ to L required levels:
   - 2.1. Calculates the corresponding level i set features F_i
   - 2.2. Classify regions according classes C_i of interest for level i
   - 2.3. For each node u in R do:
     - 2.3.1. $N \leftarrow$ Neighbors of u
     - 2.3.2. For each node u in N do:
       - Calculate a set similarity measures m between pairs of nodes u and v, according features in F_i
       - If $m \geq th$ then:
         - Merge u and v
         - Modify R

Similarity criteria

The criterion used in this paper to determine similarity S_f between neighbor regions u and v is

$$S_f = \exp(-\|F_u - F_v\|) \quad (4)$$

where $\|\cdot\|$ represents the two-norm distance. F is the set of features to evaluate. The value of S_f is set to [0, 1], and the maximum similarity between two regions is 1.

A penalty term P, based on the information of region borders, was added to the similarity function. P evaluates the difference between the shared border between region and its length. P is defined as:

$$P = \exp\left(-\frac{1}{|sh(u,v)|}\right) + \exp\left(-\|Fs_u - Fs_v\|\right) \quad (5)$$

where $|sh(u,v)|$ is the number of pixels sharing the edge of the segments u and v. Thus the longer edge is larger penalty factor. A small edge will be penalized less. Fs_u is the average value of pixels on the edge of segment u.

2.2.3. *Refinement process*

In addition, a refining process is carried out to improve the results of the analysis. It determines which segments were misclassified at the current level and must be re-labeled. In this sense, the segments of size smaller than a threshold th_r are compared with all of its neighboring segments. When one of the larger neighbors has the same type of analyzed segment, at a lower level, it acquires the label of its neighbor. Otherwise, the segment is labeled as unknown.

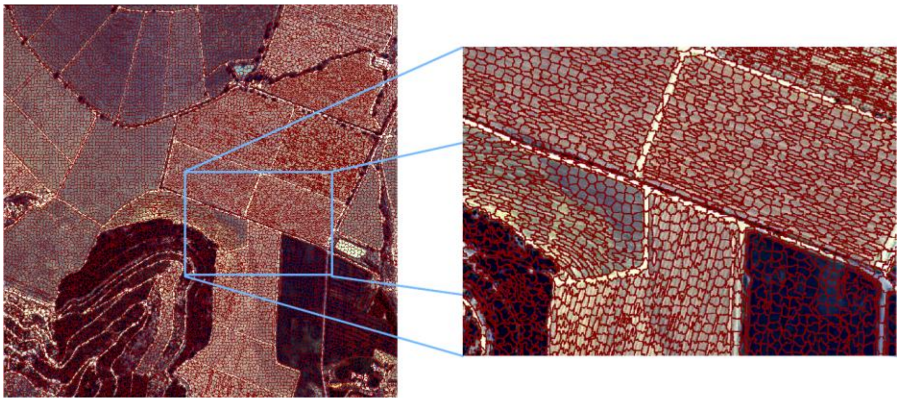
3. Results

The methodology described above and illustrated in Fig. 2, has been applied to image displayed in Fig. 1 in order to analyze two vegetation covers. A superpixel map has been created using a compactness factor m of 100. A total of 17430 superpixels were automatically generated, they represent the 0.41% of observations to analyze respect to the entire number of image pixels. The superpixel map is shown in Fig. 3a; as can be appreciated superpixels adhere well to the edges of objects.

In first level, a feature vector based on remote sensing spectral indices was calculated for each superpixel. The selected indices were NDVI, NDWI, and NDSI [11]. These indices are commonly used in land cover classification with relative success, especially when analyzing natural covers [12]. Then a superpixel-based classification was performed using a multi-label support vector machine (SVM) with a RBF kernel [13]. The target classes were: water, vegetation, and soil. Results of the classification show a smooth chart without inter-class variance (Fig. 4a).

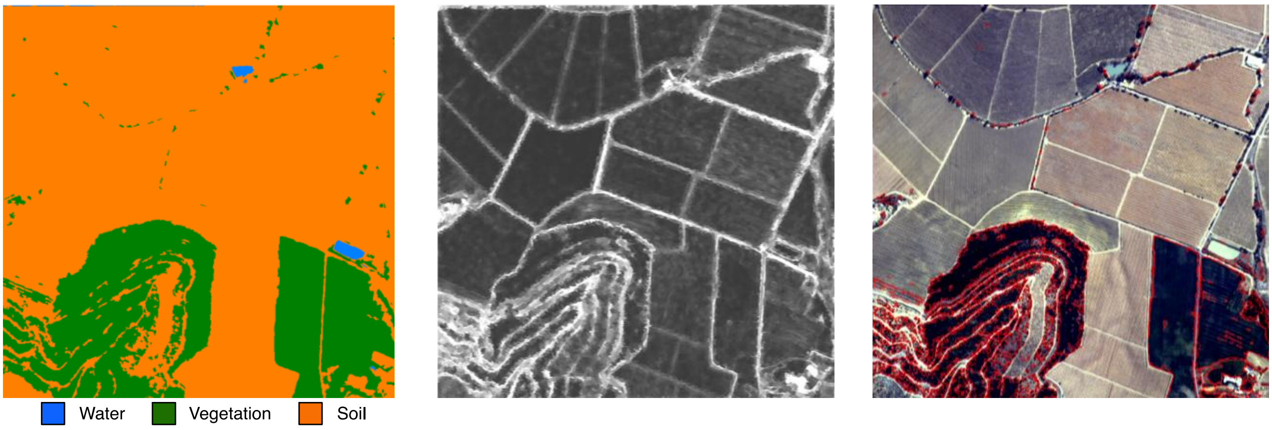
Since the interesting objects belong to a vegetation land cover, the merging process was done using NDVI as principal feature; however in order to improve quality of the segments respect to image edges, a gradient map based on superpixels was used as ancillary information. The gradient map was calculated over a smoothed superpixel map. Smoothing process was carried out using the anisotropic diffusion filtering [14]. The resulting gradient map is shown in Fig. 4b. A threshold th of 0.95 was determined through experimentation. The results of merging process are shown in Fig.

4c. As can be seen the combination of low cues as gradient map and a priori feature knowledge helps to preserve almost all the natural image edges, however some artifacts are also created.



(a) (b)

Figure 3. (a) Superpixel map. (b) Zooming into a small patch of superpixel map.



(a) (b) (c)

Figure 4. (a) Superpixels are classified into three land-cover classes. (b) Gradient superpixel map. (c) Superpixels are agglomerated using NDVI. Segment edges are colored red.

In second level, an analysis has been done to identify two classes of interest: avocado and alfalfa. Since both classes belong to vegetation, the segments that are not members of vegetation are discarded. Aimed analysis dramatically reduces the number of segments to study. For classification, textural features such as local entropy, local standard deviation and local range [15] were used.

Classification results are shown in Fig. 5. A preliminary evaluation of the results based on superpixels indicates an overall accuracy of 80.9%.

After the refining process is realized, some segments acquire the label of larger size, especially segments on the edges of objects. As shown in the Fig. 6 many of the segments are also regarded as "vegetation unknown".

4. Conclusions

This paper has presented a new methodology for the hierarchical analysis of satellite images. From a conceptual point of view, the proposed methodology allows to make a multi-scale analysis from the most relevant characteristics of the objects in each level.

The methodology employs superpixels instead of traditional pixels as minimum-size processing blocks. The use of superpixels decreases the computational cost of subsequent analysis steps and facilitates the analysis of large images. In the case under consideration, a data reduction of 99% has been reached in first level. The proposed method also contemplates the identification of objects of interest by classification process, subsequent discarding irrelevant segment reduces the space for further analysis, decreasing the complexity.

Preliminary results show the ability of the proposed methodology for multi-scale analysis of images, however it must be refined and generalized to a larger number of scales.

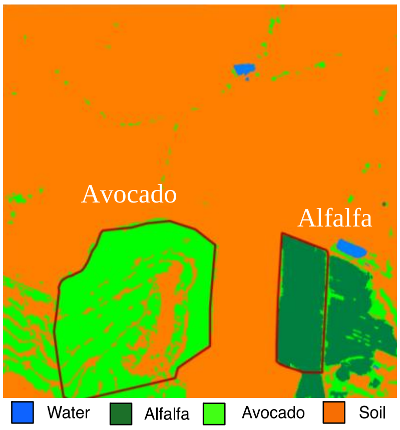


Figure 5. Vegetation regions are classified into alfalfa and avocado. Ground truth borders are colored red.

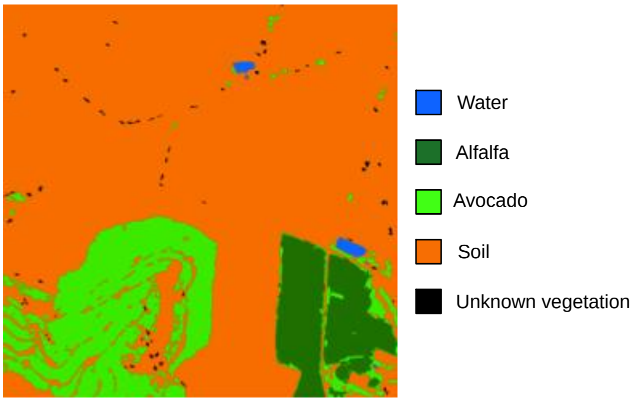


Figure 6. Results of the refinement process are shown. A new class called “unknown vegetation” is generated.

Acknowledgements

The first author acknowledges to CONACYT the support provided through the grant for doctoral studies number 216146. This work has been funded by the Universidad Politécnica de Madrid (AL13-PID-07) and the Fondo de Fomento al Desarrollo Científico y Tecnológico de Chile (FONDEF- D09I1069).

References

[1] Corcoran, P., & Winstanley, A. 2008. Using texture to tackle the problem of scale in land-cover classification. In T. Blaschke, S. Lang, & G. Hay (Eds.), Lecture Notes in Geoinformation and Cartography (pp. 113–132). Springer Berlin Heidelberg.

[2] Kim, M., Warner, T. A., Madden, M., & Atkinson, D. S. 2011. Multi-scale GEOBIA with very high spatial resolution digital aerial imagery: scale, texture and image objects. International Journal of Remote Sensing, 32(10), 2825–2850.

[3] Blaschke, T., Lang, S., & Hay, G. (Eds.). 2008. Object-Based Image Analysis: Spatial Concepts for Knowledge-Driven Remote Sensing Applications (pp. 1–817). Springer.

[4] Achanta, R., Shaji, A., Smith, K., Lucchi, A., Fua, P., & Süsstrunk, S. 2012. SLIC Superpixels Compared to State-of-the-Art Superpixel Methods. IEEE Transactions on Intelligent Transportation Systems, 34(11), 2274–2282.

[5] Ren, X., & Malik, J. 2003. Learning a classification model for segmentation (pp. 10–17 vol.1). Presented at the Computer Vision, 2003. Proceedings. Ninth IEEE International Conference on, IEEE.

[6] Zhaocong, W., Zhongwen, H., & Quian, F. 2012. Superpixel-based unsupervised change detection using multi-dimensional change vector analysis and svm-based classification. Presented at the Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences, XXII ISPRS Congress, Melbourne, Australia.

[7] Kluckner, S., Donoser, M., & Bischof, H. 2010. Super-Pixel Class Segmentation in Large-Scale Aerial Imagery (pp. 1–8). Proceedings Annual Workshop of the Austrian Association for Pattern Recognition.

[8] Liu, B., Hu, H., Wang, H., Wang, K., Liu, X., & Yu, W. 2013. Superpixel-Based Classification With an Adaptive Number of Classes for Polarimetric SAR Images. IEEE Transactions on Geoscience and Remote Sensing, 51(2), 907–924.

[9] Yu, H., Zhang, X., Wang, S., & Hou, B. 2013. Context-Based Hierarchical Unequal Merging for SAR Image Segmentation. IEEE Transactions on Intelligent Transportation Systems, 51(2), 995–1009.

[10] Lillo-Saavedra, M.; Gonzalo, C.; & Lagos, O. 2011. Toward reduction of artifacts in fused images. Int. J. Applied Earth Observation and Geoinformation 13(3): 368–375.

[11] Wolf, A. 2010. Using WorldView 2 Vis-NIR MSI imagery to support land mapping and feature extraction using normalized difference index ratios. DigitalGlobe 8-Band Research Challenge, 1-13.

[12] Kulawardhana, R. W. (2011), Remote sensing of vegetation: principles, techniques and applications. By Hamlyn G. Jones and Robin A Vaughan. Journal of Vegetation Science, 22: 1151–1153.

[13] Chih-Chung & Chih-Jen. 2011. LIBSVM: A library for support vector machines. ACM Transactions on Intelligent Systems and Technology 2(3):1–27. Software available at http://www.csie.ntu.edu.tw/~cjlin/libsvm.

[14] Perona & Malik. 1990. Scale-space and edge detection using anisotropic diffusion. IEEE Transactions on Pattern Analysis and Machine Intelligence, 12(7):629-639.

[15] Gonzalez, R.C.; Woods, R.E. & Eddins, S.L. 2003. Digital Image Processing Using MATLAB, New Jersey, Prentice Hall, Chapter 11.