

# THE SYNTHETIC IMAGE TESTING FRAMEWORK (SITEF) FOR THE EVALUATION OF MULTI-SPECTRAL IMAGE SEGMENTATION ALGORITHMS

André R. S. Marçal, Arlete Rodrigues, Mário Cunha

Centro de Investigação em Ciências Geo-espaciais  
Faculdade de Ciências, Universidade do Porto

DMA, Rua do Campo Alegre, 687, 4169-007 Porto, Portugal – andre.marcal@fc.up.pt

## ABSTRACT

The segmentation stage is a key aspect of an object-based image analysis system. However, the segmentation quality is usually difficult to evaluate for satellite images. The Synthetic Image TEsting Framework (SITEF) is a tool to evaluate and compare image segmentation results. This paper presents the SITEF with an extension to model adjacency effects between neighboring parcels, using the sensor's point spread function and a grid offset. A practical application of SITEF is presented using a SPOT HRG satellite image, with 6 vegetation land cover classes identified on a mountainous area. The segmentation results were evaluated under various perspectives, including the parcel size and shape, the land cover types, the sensor grid offset and one parameter used in the segmentation algorithm.

**Index Terms**— Image segmentation, synthetic images, segmentation evaluation

## 1. INTRODUCTION

One of the most challenging tasks in remote sensing at present is how to handle the huge amounts of image data acquired by the current Earth Observation Satellites (EOS). An alternative approach to the standard per-pixel analysis of multi-spectral EOS images has evolved over the last decade, due to the dramatic increase in commercially available high resolution (5.0 m and finer) remote sensing imagery [1]. Instead of focusing on individual image pixels, the object-based image analysis approach consists of partitioning an image into meaningful image objects. This approach is based on the fact that image pixels are not considered to be “natural” elements of an image scene.

A common element of all object-based image analysis systems is the segmentation stage, where the image is partitioned in a number of objects (or segments). This is a critical stage of the whole process, because if the segmentation fails to identify as an object a given element present in the image, the subsequent stages will generally be unable to recognize it or classify it. An evaluation of the segmentation algorithms performance on the image data used is therefore an important aspect for object-based image analysis. However, there is no established standard procedure for the evaluation of the segmentation results produced for EOS images [2]. The most common methods are based on discrepancy measures between the segmentation result and a reference [3], but it is usually difficult to have abundant reference segmentation data for EOS images.

The purpose of this work is to present the Synthetic Image TEsting Framework (SITEF), a tool to evaluate the performance of segmentation algorithms on multi-spectral EOS images. The method is based on the production of synthetic images with the spectral characteristics of the image pixels extracted from a signature EOS image. The work presented here is an evolution of the method described in [4], by considering adjacency effects between neighboring parcels. A practical experiment was carried out to illustrate the usefulness of the proposed methodology. The SITEF software and several image datasets are freely available at [www.fc.up.pt/sitef](http://www.fc.up.pt/sitef).

## 2. SYNTHETIC IMAGES

A multi-spectral synthetic image is created using a base image, with pre-defined geometric characteristics, and a signature image where a number of land cover types are identified.

### 2.1. Base image

The base image is produced with rectangular areas (parcels) assigned to different class labels. Two neighboring parcels always belong to different classes. The image characteristics depend on the following parameters: number of classes or land cover types ( $t$ ), size of the smallest unit ( $u$ ), range of sizes ( $s$ ) and a repetition factor ( $r$ ).

The process is best understood using an example, such as the images presented in figure 1. For these base images, the size of the smallest unit ( $u$ ) is 4, which means that the smallest parcels on the top left corner are 4x4 pixels. The range of sizes ( $s$ ) is 4, so that the images have parcels between 1x1 and 4x4 units (16x16 pixels). The repetition factor ( $r=2$ ) indicates the number of similar columns and lines. The images are 80x80 pixels, with a total of 64 different parcels, 4 of each size ( $i*j$  units, with  $i,j=1,\dots,4$ ). The only difference between the three base images in figure 1 is the number of classes:  $t=4$  (a),  $t=5$  (b), and  $t=6$  (c).



Figure 1. Base images with 4 (a), 5 (b) and 6 (c) classes, all with  $u=4$ ,  $s=4$  and  $r=2$ .



Figure 2. Example of a signature image (left) with 4 land cover types identified (right).

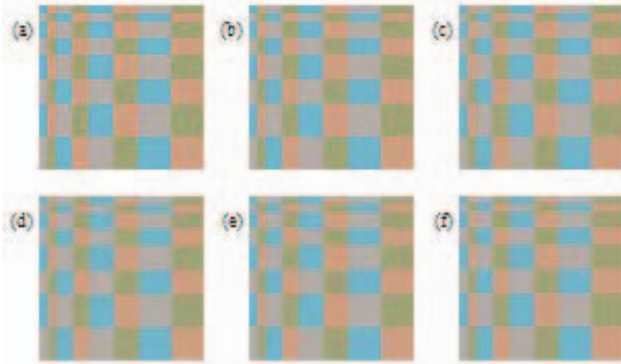


Figure 3. Synthetic images produced with base image of fig.1a and signature image of fig.2: raw (a), PSF only (b) and with PSF and grid offset of (1/4;0) (c), (1/2;0) (d), (1/4;1/4) (e), (1/2;1/2) (f).

## 2.2. Raw multi-spectral image

The base image provides the geometric characteristics of the synthetic image, with the class labels corresponding to land cover types. The pixel values of the multi-spectral synthetic image are obtained using a reference or signature image, where training areas are identified for each land cover type. Figure 2 shows an example of a signature image (a RGB color photo) with 4 land cover types.

The synthetic image produced has the same number of bands as the signature image. Each pixel in the base image is replaced by a pixel vector selected randomly from the corresponding training area. An example of a multi-spectral synthetic image is presented in figure 3a, using the base image of figure 1a and the signature image of figure 2.

## 2.3. Including adjacency effects

The multi-spectral synthetic image produced using the base and signature images assumes that an ideal sensor is used. However, this is not the case with real sensors, as the instantaneous field of view (IFOV) does not correspond to a square pixel. A more realistic approach is to use the Point Spread Function (PSF) of a sensor. Furthermore, even if the land parcels were as the ideal synthetic images, in a real case scenario there might be an offset between the image grid and the ideal grid on the ground. Both the PSF and the grid offset result in a mixture of the pixel values with their neighbors.

The final synthetic image ( $I$ ) is obtained as the convolution of the raw image ( $I_0$ ) with a PSF ( $f$ ) and a grid offset function ( $g$ ),

$$I = I_0 \otimes f \otimes g \quad (1)$$

where  $\otimes$  is the convolution operator. The function  $f$  used was the empirical normalized PSF matrix of  $9 \times 9$  [5]. The grid offset function ( $g$ ) is a  $3 \times 3$  matrix, modeling the offset between two grids. Besides the absence of offset (0;0), the offset values used were (in pixel units): (1/4;0), (1/2;0), (1/4;1/4), (1/2;1/2).

As an illustration of the adjacency effects introduced by the process described, figure 3 shows a raw synthetic image (a), the result after the PSF convolution (b), as well as using the PSF plus various levels of the grid offset functions (c-f). In this figure it is clear that the smoothing increases with the grid offset level.

## 3. SEGMENTATION EVALUATION

There is an exact knowledge of the location of every land parcel (object) in a synthetic image. This is used to provide a reference or expected (ideal) segmentation, allowing for a quantitative evaluation of a segmentation result. The Hammoude metric and three external similarity indices were used for this purpose.

The Hammoude metric evaluates the similarity between two segmentations proposed for an object ( $X$  and  $Y$ ), by comparing the number of common and non-common pixels in the two segmentations [6]. The Hammoude metric is computed by

$$H = \frac{\#(X \cup Y) - \#(X \cap Y)}{\#(X \cup Y)} \times 100 \quad (2)$$

where  $X$  and  $Y$  are two binary representations of the segmented object, and the operator  $\#$  returns the “number of pixels ON” of a binary image. The Hammoude metric, in the form presented in (2), has values between 0 and 100, with a value of 100 occurring when there is no intersection between segmentations (completely dissimilar), and a value of 0 when the two segmentations are equal.

The evaluation of a segmented image using a reference can also be seen as a problem of comparing two classification results (data partitions), where each object (or segment) is assigned a class label. The segmented images can thus be evaluated using external indices of cluster validity, which access the degree to which two classifications of the data agree [7], such as the Rand (R), Corrected Rand (CR) and Jaccard (J) coefficients [4]. The computation of these external similarity indices is based on an inspection of labels assigned to all pixel pairs. Details about the computation of the indices can be found in [4] and in [7]. These indices usually range between 0 and 1, with 1 assigned to a perfect result. However, in order to match the range used with the Hammoude metric, the values of R, CR and J presented here are scaled to 0-100, with 0 corresponding to a perfect result.

## 4. EXPERIMENTAL PROCEDURE

Three base images were created, all with  $u=3$ ,  $s=8$  and  $r=5$ . These images are 540 by 540 pixels, with a total of 1600 parcels. The smallest parcels (1x1 units) are  $3 \times 3$  pixels, and the largest ( $8 \times 8$  units) are  $24 \times 24$  pixels. There are 25 parcels for each square size (between 1x1 and  $8 \times 8$  units) and 50 rectangular parcels of  $i$  by  $j$  units (with  $i=2$  to 8,  $j=1$  to 7, and  $i > j$ ). In total there are 36 different parcel shapes/sizes. The base images were created with 5, 6 and 8 land cover types ( $t=5, 6$  and 8).

A SPOT HRG satellite image covering a mountainous area in Montalegre, NW Portugal, was used as signature image. The SPOT HRG image has 4 spectral bands in the visible and near

infrared, and a spatial resolution of 10 meters. A total of 8 land cover types were identified in the signature image, listed in table I.

Test Images (TI) were created with eight different sets of land cover classes: all 8 classes (TI8), only the 6 vegetation classes (TI6), and with the six sub-sets of 5 vegetation classes (TI5a to TI5f, where the last character indicates the vegetation class missing, e.g. TI5a only uses classes B-F).

For each set of land cover classes, a total of 6 synthetic images were created: 1 raw image, without considering adjacency effects, 1 using the PSF only, and 4 using the PSF and grid offsets. A total of 48 images were created altogether. These images were segmented using the multiresolution algorithm available on Definiens 7 software [8], using the default values for all parameters. For one test image (TI5d), the scale parameter of the multiresolution algorithm was tested with 5 different values.

TABLE I. LAND COVER CLASSES USED IN THE SIGNATURE IMAGE. (\*) NON-VEGETATION CLASS.

Label	Class description
A	Irrigated permanent semi-natural mountain meadows
B	Non-irrigated perman. semi-natural mountain meadows
C	Evergreen forest
D	Deciduous forest
E	Communitarian pastures
F	Annual crops
G (*)	Sand
H (*)	Water

TABLE II. AVERAGE H FOR PARCELS SIZED I BY J UNITS (TI6).

i,j	1	2	3	4	5	6	7	8
1	51.1							
2	34.5	12.7						
3	25.6	4.2	1.0					
4	9.6	3.8	1.8	1.2				
5	3.8	2.9	1.3	1.2	1.0			
6	4.8	1.7	1.7	1.1	0.9	1.1		
7	4.6	1.6	1.5	1.4	0.8	1.0	0.6	
8	5.4	1.5	1.3	0.9	0.8	0.9	0.6	0.7

TABLE III. AVERAGE H FOR CLASS / SET OF CLASSES.

Class	TI5a	TI5b	TI5c	TI5d	TI5e	TI5f
A		0.8	5.6	5.0	5.8	5.6
B	9.1		13.0	4.6	11.9	11.6
C	0.2	0.2		0.0	0.0	0.0
D	8.5	0.9	9.4		8.7	8.8
E	4.4	4.3	4.7	4.5		0.0
F	4.2	4.1	4.5	4.4	0.0	

TABLE IV. AVERAGE H FOR 5 GRID OFFSET VALUES (TI5D).

Class	(0;0)	(1/4;0)	(1/2;0)	(1/4;1/4)	(1/2;1/2)
A	5.0	13.6	16.6	22.1	24.5
B	4.6	5.1	8.2	14.1	16.9
C	0.0	4.0	10.1	13.8	20.4
E	4.5	11.9	17.5	25.0	24.0
F	4.4	5.3	8.6	14.2	14.9
All	3.7	8.0	12.2	17.9	20.1

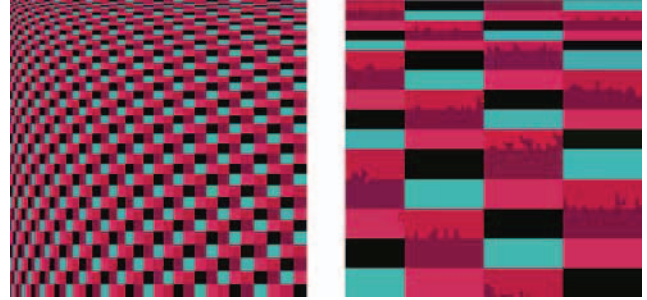


Figure 4. Segmentation results for TI5d (left) and 6x detail of the top right corner (right).

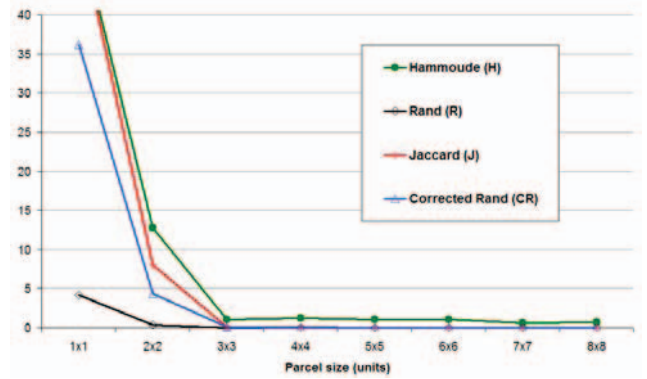


Figure 5. Average H, R, J and CR for the square parcels of TI6.

## 5. RESULTS

The segmentation results provided by the segmentation software were compared with the expected (ideal) segmentation, using the Hammoude metric and the similarity indices, for the 1600 parcels (objects) on each test image. The segmentation evaluation can be done through multiple perspectives, such as: for different parcel sizes and shapes, for the different land cover types and sets of land cover types, for the different parameter settings used in the segmentation algorithm.

The results for TI8 showed that the segmentation of land parcels of sand and water is nearly always perfect, as the spectral signature of these two land cover types is very distinctive. As an illustration of a more typical case, the segmentation of one test image with only vegetation classes (TI5d) is presented in figure 4. The complete segmented image (540 by 540 pixels) is presented together with a detail of the top right corner, where the mis-segmentation of some parcels is clearly visible.

The average values of H, R, CR and J are presented in figure 5, for the square shaped parcels of TI6 (with PSF and without grid offset). The plot clearly shows that the major problems in segmentation occur for small parcels (1x1 and 2x2 units). It also shows the strong correlation between H, J and CR. This was identified in [4], as well as the poor ability of the Rand index to discriminate between good and bad segmentation results. Although the four metrics were computed for all test images, only the Hammoude metric results are presented throughout.

Table II shows the average H for all parcel shapes/sizes of TI6 (with PSF and without grid offset), including the square parcels plotted in figure 5. There is a clear improvement in the



segmentation results as the parcel size increases. For parcels with the same area but different shapes (e.g 1x8 and 2x4), the best results are obtained for parcels with the smaller perimeter.

A factor that influences the segmentation result even more than the parcel size and shape is the land cover type of a parcel and the land cover types of its neighbors. This can be evaluated with the results presented in table III, for images with PSF and without grid offset. For example for TI5d (figure 4) there is one class with a perfect result ( $H=0.0$ ) while the other classes all have non-negligible values of  $H$  (between 4.4 and 5.0). It is also interesting to observe that the performance of a class greatly depends of the set of land cover types present in the image. For example, for class B, the average  $H$  varies between 4.6 and 13.0, for the various combinations of vegetation land cover types tested. This is not surprising, because when two neighboring parcels have similar spectral signatures, is difficult for the segmentation algorithm to establish the expected boundary between the two parcels.

The adjacency effects between neighboring parcels are modeled by the sensor PSF and a grid offset parameter. The larger the offset the more smoothing is introduced, resulting in a mixture of the signals recorded for pixel values, particularly at the edges of parcels. The segmentation quality is obviously affected by this mixture. The average  $H$  for the land cover types of TI5d are presented in table IV, for the 5 values of grid offset tested. The values of  $H$  for the whole image increase steadily with the offset, from 3.7 to 20.1 for the worst case scenario. This is a general behavior that is observed for all land cover types and parcel sizes. Figure 6 shows a plot for the whole image and for parcels with roughly the same area (15-16  $u^2$ ) but different shapes. The square parcels (4x4 units) are less affected by the increasing grid offset, as they have less pixels at the edges, which are the most influenced by the adjacency effects.

A final evaluation involves the selection of an optimum value for the scale parameter ( $S$ ) on the multiresolution algorithm of Definiens 7 software [8]. The image TI5d with PSF and grid offset (1/4;0) was used for this test. The image was segmented with 5 values assigned to  $S$  (6,8,10,12,14). The average  $H$  for the square parcels are presented in figure 7. The optimum value of  $S$  depends of the parcel size, with lower values being favored by the smaller parcels. This procedure can be used to fine tune the segmentation parameters for the requirements (spatial, spectral or other) of a specific application.

## 6. CONCLUSIONS

The evaluation of satellite image segmentation results based on the SITEF provides multiple perspectives that can assist in the selection of the most suitable segmentation algorithm or to fine tune its parameters. The practical experiment presented shows the potential of the method and the importance of various aspects for the segmentation result. The extension of the method to model the adjacency effects between neighboring parcels, presented in this paper, provides synthetic images that reproduce in a more realistic way the behavior of real land parcels.

## 7. AGNOWLEDGMENTS

The authors wish to thank Fundação Calouste Gulbenkian and FCT (LAMSAT XXI project - PTDC/AGR-AAM/67182/2006) for the financial support, the OASIS program (n° 307) for providing the SPOT data and I. Poças for the identification of land cover classes.

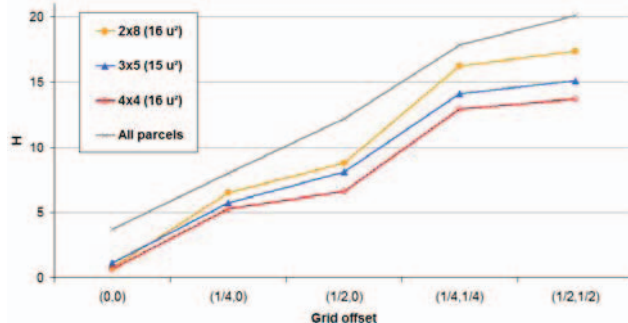


Figure 6. Average H for different grid offset values (TI5d).

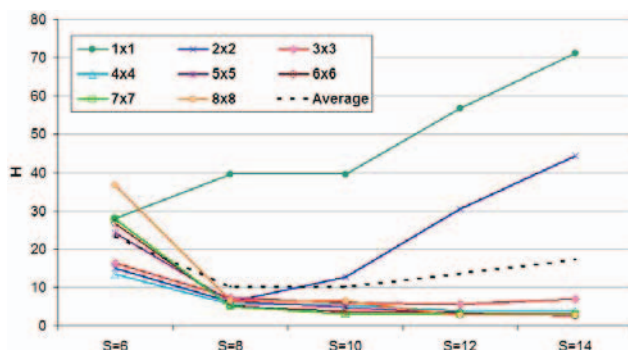


Figure 7. Average H for different values of the scale parameter ( $S$ ) for square parcels (TI5d).

## 8. REFERENCES

- [1] G.J. Hay, G. Castilla, M.A. Wulder, J.R. Ruiz, "An automated object-based approach for the multiscale image segmentation of forest scenes", *International Journal of Applied Earth Observation*, Elsevier, Vol. 7, pp. 339-359, 2005.
- [2] M. Neubert, H. Herold, G. Meinel, "Assessing Image Segmentation Quality Concepts, Methods and Application", In: Blaschke, T.; Hay, G. & Lang, S. (Eds.): *Object-Based Image Analysis - Spatial concepts for knowledge-driven remote sensing applications*, Springer, Berlin, pp. 769-784, 2008
- [3] A.P. Carleer, O. Debeir, E. Wolff, "Assessment of very high spatial resolution satellite image segmentations", *Photogrammetric Engineering and Remote Sensing*, ASPRS, Vol. 71, pp. 1285-1294, 2005.
- [4] A.R.S. Marçal, A. Rodrigues, "A method for multi-spectral image segmentation evaluation based on synthetic images", *Computers & Geosciences*, Elsevier, (in press), 2009, DOI: 10.1016/j.cageo.2008.11.008
- [5] C.P. Ruiz, F.J.A. López, "Restoring SPOT images using PSF-derived deconvolution filters", *International Journal of Remote Sensing*, Taylor & Francis, Vol. 23, pp. 2379-2391, 2002.
- [6] A. Hammoude, "Computer-assited endocardial border identification from a sequence of two-dimensional echocardiographic images", unpublished Ph.D. dissertation, University of Washington, Seattle, WA, 1998.
- [7] R.C. Dubes, "How many clusters are best? - An experiment", *Pattern Recognition*, Vol. 20, pp. 645-663, 1987.
- [8] DEFINIENS, "Definiens Developer 7 User Guide", 2008