

A FRAMEWORK FOR THE EVALUATION OF MULTI-SPECTRAL IMAGE SEGMENTATION

André R. S. Marçal, Arlete S. Rodrigues

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 ; dr.arlete@gmail.com

KEY WORDS: Image segmentation, synthetic images, similarity indices, segmentation metrics, segmentation evaluation

ABSTRACT:

A general framework for testing the quality of the segmentation of a multi-spectral satellite image is proposed. The method is based on the production of synthetic images with the spectral characteristics of the image pixels extracted from a signature multi-spectral image. The knowledge of the exact location of objects in the synthetic image provides a reference segmentation, which allows for a quantitative evaluation of a segmentation algorithm applied to the image. The Hammoude metric and the external similarity indices Rand, Corrected Rand and Jaccard are used.

A practical application was carried out to illustrate the value of the proposed method. Two satellite images, from SPOT HRG and Landsat TM, were used to extract the spectral signature of 8 land cover types. Six test images were produced using all 8 land cover classes and with two different sub-sets with 5 classes. The segmentation results provided by a standard algorithm were compared with the reference or expected segmentation. An evaluation of the parameters used in the eCognition software segmentation algorithm was also carried out, using the proposed indices.

1. INTRODUCTION

Multi-spectral images acquired by Earth Observation Satellite (EOS) are increasingly being used. Over the last decade an alternative approach to the standard per-pixel analysis has evolved to extract meaningful information from EOS images. Instead of focusing on individual image pixels, the object-based image analysis approach consists of partitioning an EOS image into meaningful image-objects, and assessing their characteristics through spatial, spectral and temporal scale (Hay & Castilla, 2006). One of the reasons for the development of object-based methods has been the dramatic increase in commercially available high resolution digital remote sensing imagery, with spatial resolutions of 5.0 m and finer (Hay et al., 2005). Also it has been recognised that the image pixel is not a “natural” element 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 clearly a critical stage of the whole process. If the segmentation fails to identify as an object a given element present in the image, the subsequent stages will generally be unable to recognise or to classify this element. It is generally accepted that only a few of the various image segmentation methods that have been used in image analysis lead to qualitatively convincing results, which are robust and applicable under operational settings (Baatz & Schape, 2000). The quality of the results produced by a segmentation algorithm are strongly dependent on the image data, the scene characteristics and the parameters used. The correct evaluation of the segmentation process is therefore an important aspect of any object based image analysis system. However, it is difficult to evaluate the quality of the segmentation result for EOS images, as there are usually no reference values that can be used for comparison.

The purpose of this work is to present a framework for the evaluation of image segmentation algorithms based on the production of synthetic multi-spectral images. These images are produced with controlled characteristics, both spatially and

spectrally, simulating any satellite sensor and set of land cover types.

2. EVALUATION OF SEGMENTATION RESULTS

Image segmentation is the process of partitioning an image into a set of non-overlapping regions whose union is the entire image. The goal of the image segmentation stage is usually to identify the object of interest in an image, separating it from the background, or alternatively to divide the image in a small number of segments or objects. However, in the case of EOS images, the number of objects present in the image is normally very large. This brings a problem of evaluating the segmentation result, as the commonly used metrics, such as the true detection rate, average distance or Hausdorff distance are mostly suitable for the case of a single interest object (Chalana & Kim, 1997). Another limitation is the lack of a reference segmentation. There are some examples of evaluation of EOS image segmentation results using manually established reference objects (Moller et al, 2007), but usually there are no reference segmentation available for comparison. The proposed methodology based on synthetic images provides a reference (expected) segmentation, and makes use of the Hammoude metric and the external similarity indices Rand, Corrected Rand and Jaccard to evaluate quantitatively the segmentation result.

2.1 Standard methods

The Hammoude metric allows for an evaluation of 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 (Hammoude, 1988). The Hammoude metric is computed by

$$H = \frac{\#(X \cup Y) - \#(X \cap Y)}{\#(X \cup Y)} \quad (1)$$

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 has values between 0 and 1, with a value of 1 occurring when there is no intersection between segmentations (completely dissimilar), and a value of 0 when the two segmentations are equal. In order to have consistency between various metrics / indices, the inverted Hammoude index (H') is also used, where H'=1-H (H'=1 for a perfect match and H'=0 for two non-intersecting objects).

2.2 External similarity indices

The problem of comparing two segmented images can alternatively be seen as a problem of comparing two classifications (data partitions), where each object (or segment) is assigned a class label. The similarity between two segmented images can thus be evaluated using external indices of cluster validity, which access the degree to which two classifications of the data agree (Dubes, 1987).

The external indices used here are the Rand coefficient (Rand, 1971), the Corrected Rand coefficient (Hubert & Arabie, 1985) and the Jaccard coefficient (Dubes, 1987). The computation of these indices is based on 4 variables (a,b,c,d) that are obtained by inspecting the labels assigned to all pairs of patterns (pixels) in the image dataset. Each image pixel (i) has two class labels, X(i) and Y(i), corresponding to the object numbers assigned by segmentations X and Y. The variable a is computed by counting the number of pixel pairs that have the same class in X and the same class in Y (2a). The variable d corresponds to the number of pixel pairs with different labels in both X and Y (2d), while b and c refer to pixel pairs that are assigned the same label in one of the classified images and different labels in the other (2b, 2c). The four values (a,b,c,d) sum is the number of ordered pairs of pixels: a+b+c+d=N(N-1)/2, where N is the total number of image pixels.

$$a = \# \{ (i,j): i>j, X(i)=X(j), Y(i)=Y(j) \} \quad (2a)$$

$$b = \# \{ (i,j): i>j, X(i)=X(j), Y(i) \neq Y(j) \} \quad (2b)$$

$$c = \# \{ (i,j): i>j, X(i) \neq X(j), Y(i)=Y(j) \} \quad (2c)$$

$$d = \# \{ (i,j): i>j, X(i) \neq X(j), Y(i) \neq Y(j) \}. \quad (2d)$$

The Rand coefficient (R) is the relative number of pixels pairs that are treated in the same way under both classifications (Dubes, 1987), and is computed by

$$R = \frac{a+d}{a+b+c+d} \quad (3)$$

The Jaccard coefficient (J) is also based on the relative number of pixels pairs that are treated in the same way under both segmentations, but discounting the situations where both classifications assign different classes to pixel pairs (d).

$$J = \frac{a}{a+b+c+d} \quad (4)$$

Both Rand and Jaccard indices vary between 0 and 1, with 1 corresponding to perfectly matched classifications. The Corrected Rand coefficient (CR) is a modified version of the standard Rand index, proposed by Hubert & Arabie (1985),

where the range of values (0 to 1) is adjusted so that a fully random classification would correspond to a value of 0. The Corrected Rand coefficient is computed by (5), where R_{max} is the maximum possible value for R (R_{max}=1) and R_{exp} is the expected value of R under randomness (Hubert & Arabie, 1985). In order to have a range between 0 and 1 for CR, negative values can be truncated to 0.

$$CR = \frac{R-R_{exp}}{R_{max}-R_{exp}} \quad (5)$$

As the patterns considered are the image pixels (N), which is usually large, the very large number of pixel pairs (~N²/2) could cause problems in the computation of a, b, c, d. However, Hubert & Arabie (1985) presented a combinatorial approach that is computationally very efficient, making the external similarity indices suitable for large images with any number of classes, such as the labelled images produced by segmentation of EOS images.

3. SYNTHETIC IMAGE PRODUCTION

A methodology to produce synthetic test images was developed to simulate a scenario with land parcels of various sizes, belonging to a number of cover types. Initially, a base image is produced with rectangular areas (parcels) assigned to different class labels. A multi-spectral signature image is used to obtain multi-spectral pixels belonging to each land cover type, and together with the base image create the final multi-spectral synthetic image.

3.1 Base image

A base image is produced given the following parameters: the number of land cover types (t), the size of the smallest unit (u), the range of sizes (s) and a repetition parameter (r). The base image has a total of r²s² parcels of rectangular shape, with sizes from 1*1 to s*s units. There are r² single unit parcels (of u by u pixels), and generally r² parcels of i*j units, with i,j=1,...,s. It is worth noting that two neighbouring parcels always belong to different classes.

As an example, Figure 1 shows three base images, with u=4, s=4, r=2, with 4, 5 and 7 classes. The smallest squares on the top left section of these images have 4 by 4 pixels (1 unit), while the largest ones on the lower right part of the images have 16 by 16 pixels (4 by 4 units, s=4). In this case the repetition parameter (r) is 2, which means that there are 4 parcels of each size (r²). The whole images are 80 by 80 pixels, with a total of 64 different parcels. There are 4 single unit parcels (4 by 4 pixels), and 4 parcels of i*j units, with i,j=1,...,4.

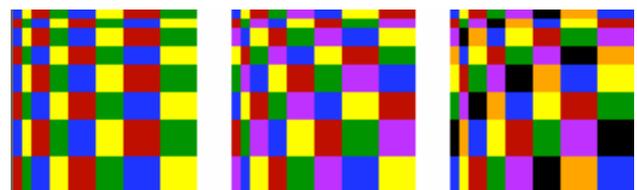


Figure 1. Synthetic base images (u=4, s=4, r=2) with the number of classes t=4 (left), 5 (centre) and 7 (right).

3.2 Multi-spectral synthetic image

The production of a multi-spectral synthetic image from the base image requires a signature image with training areas identified for each land cover type. The multi-spectral synthetic image is produced, with the same number of bands as the signature image, by replacing each pixel in the base image by a pixel vector selected randomly from the corresponding training area. An example is presented in Figure 2, where four land cover types were identified in a 3 band (RGB true colour) signature image. The synthetic multi-spectral image was produced using the first base image of Figure 1, with $u=4$, $s=4$, $r=2$, $t=4$.



Figure 2. Signature image (left) with 4 land cover types (centre) and the resulting synthetic image (right).

3.3 Evaluation of the synthetic image segmentation

The multi-spectral synthetic image is used to evaluate the performance of a given segmentation algorithm. The expected or ideal segmentation consists of identifying each parcel as an object. A reference segmentation image is created by assigning a different object label to each parcel in the base image. The segmentation result (Y) obtained from the application of a segmentation algorithm to the multi-spectral synthetic image is then compared with the reference segmentation (X).

For the Hammoude metric, a value of H_m is computed for each object m of X , by comparing it with the object n of Y that contains the central pixel of m . Average values of H are computed for all parcels of the same size and for all parcels belonging to the same class in the base image. A global value of H is also obtained averaging the H_m of all parcels in the image.

For the external similarity indices, the segmentation images are divided in sub-images, corresponding to the areas where the objects have the same size in X . The Rand, Corrected Rand and Jaccard indices are computed for each of these sub-images. As there are normally no difference between the horizontal and vertical directions, the average values are computed for all parcels of size $i*j$ and $j*i$ (e.g. no distinction is made between parcels of 2 by 1 and 1 by 2 units). The average indices computed are thus based on r^2 objects for square parcels, and $2r^2$ objects for the other parcel sizes.

4. EXPERIMENTAL SETUP

A practical experience was carried out to illustrate the usefulness of the proposed method. Six multi-spectral synthetic images were initially created, and then segmented using the eCognition software (Baatz et al., 2001). The results provided by the segmentation software were compared with the expected (ideal) segmentation using the proposed indices.

4.1 Synthetic image characteristics

Two base images were created, one with 5 classes ($t=5$) and one with 8 classes ($t=8$), both having $u=3$, $s=8$ and $r=5$. These images are 540 by 540 pixels, with a total of 1600 parcels. The

smallest parcels of 1 by 1 units are 3 by 3 pixels, and the largest, of 8 by 8 units, are 24 by 24 pixels. There are 25 parcels for each of eight square sizes (between 1 by 1 and 8 by 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 sizes.

Two satellite images, from SPOT HRG and Landsat TM sensors, were used as signature images. These images cover a mountainous area near Montalegre, Portugal. The SPOT image has 4 spectral bands in the visible and near infrared parts of the electromagnetic spectrum, and a spatial resolution of 10 meters. The Landsat image has 6 spectral bands with 30 m pixel resolution (the thermal band of TM was not used).

A total of 8 land cover types were used, with training areas identified manually in each signature image. The land cover types used were: 1 - Irrigated permanent semi-natural mountain meadows; 2 - Non-irrigated permanent semi-natural mountain meadows; 3 - Evergreen forest; 4 - Deciduous forest; 5 - Communitarian pastures; 6 - Annual crops; 7 - Sand; 8 - Water. Two sub-sets of 5 out of the 8 classes were also considered, one where the 5 most different classes were selected (5d – classes 1, 3, 4, 7, 8), and another with the 5 most similar classes (5s – classes 1, 2, 4, 5, 6). The three sets of classes were used, together with the two signature and two base images, to produce six synthetic test images (named SPOT8, SPOT5d, SPOT5s, TM8, TM5d and TM5s). The synthetic test images have the same number of bands as the signature satellite images (4 bands for SPOT and 6 bands for TM). Figure 3 shows RGB colour composites of the three SPOT test images (bands 321).

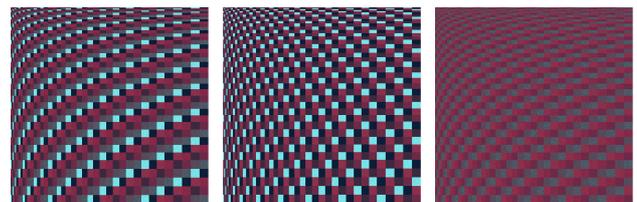


Figure 3. Synthetic test images SPOT8 (left), SPOT5d (centre) and SPOT5s (right), RGB colour composites of bands 321.

4.2 Segmentation

The six test images were segmented using the eCognition software, with the following parameters: scale parameter – 10, color – 0.8, shape – 0.2, smoothness – 0.9 and compactness – 0.1 (Baatz et al., 2001). The test image SPOT5s was also used to evaluate the influence of the eCognition segmentation parameters in the final result. This is the most challenging of the six test images, as the land cover classes used are all spectrally very similar. A set of reference values were established for the eCognition segmentation parameters. Each parameter was then changed within a range of values, with all remaining parameters fixed at their reference values. The values tested (the reference values are in brackets) were: scale parameter – 4, 6, 8, (10), 12, 14, 16, 18, 20; color – 0.50, 0.55, 0.60, 0.65, 0.70, 0.75, (0.80), 0.85, 0.90, 0.95; smoothness – 0.50, 0.55, 0.60, 0.65, 0.70, 0.75, 0.80, 0.85, (0.90), 0.95, 1.00. The parameters shape and compactness are dependent on the values assigned to color and smoothness (shape+color=1, smoothness+compactness=1). A total of 30 segmentation results were thus produced for the test image SPOT5s.

The 35 segmentation results produced by eCognition (30 results for SPOT5s and 5 for the other test images) were converted to raster, and the image pixels labelled with the corresponding object number.

5. RESULTS

The results are presented initially for the six test images segmented with the reference parameters, and in a separate section the analysis of the results produced by changing the eCognition segmentation parameters for SPOT5s.

As an illustration of the type of results produced by the segmentation process, a section of the segmented SPOT images is presented in Figure 4. This section corresponds to the area of the base image with parcels of 6 by 1, 6 by 2 and 6 by 3 units (there are 25 parcels of each size). The results obtained for SPOT5d for this section (figure 4, centre) perfectly match the reference segmentation image, unlike the other two results. For SPOT8 about half of the parcels are segmented exactly, while for the other half there are a few pixel differences between the two segmentations. For SPOT5s there are considerable differences between the reference and eCognition based segmentations. For this image, two parcels were segments into a single object in two instances.

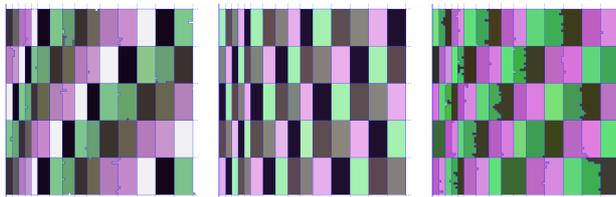


Figure 4. Detail of the images produced by segmentation of SPOT8 (left), SPOT5d (centre) and SPOT5s (right).

5.1 Test images from SPOT

The total number of objects in each segmented image is less than the expected (ideal) number of 1600 objects present in the reference segmentation: 1563 for SPOT8, 1596 for SPOT5d and 1545 for SPOT5s. These values, as well as the overall average Hammoude metric (H), Rand (R), Corrected Rand (CR) and Jaccard (J) indices are presented in Table 1 for the three SPOT test images. The low values for H and very high values of R, CR and J indicate a good segmentation overall, close to the expected segmentation. However, there are great variations in the segmentation quality in terms of scale and land cover type, which will be analysed next.

	SPOT8	SPOT5d	SPOT5s
No. objects	1563	1596	1545
H	0.0430	0.0025	0.0820
H'	0.9570	0.9975	0.9180
R	0.9977	0.9998	0.9941
CR	0.9767	0.9977	0.9549
J	0.9623	0.9962	0.9348

Table 1. Global segmentation results for the SPOT test images.

5.1.1 Land cover classes. An inspection of the results for the section presented in Figure 4 reveals that not only the set of land cover classes used to produce the synthetic image influences the results, but also that within each single image there are classes that are more likely to be correctly segmented than others. The colours assigned to each object in Figure 4 are the RGB average of its pixels, with a fixed RGB colour composite and histogram enhancement assigned to each image. The average value of H computed for each land cover class is presented in Table 2. There are two classes that have a perfect score, water (ID=8) and sand (ID=7), which are spectrally very different than the remaining classes. For the other six classes, the segmentation results are better in the presence of spectrally

distinct classes, which results in generally better scores for all classes in SPOT5d, followed by SPOT8 and by SPOT5s.

Class ID	SPOT8	SPOT5d	SPOT5s
1	0.0279	0.0047	0.0445
2	0.1142	(*)	0.1504
3	0.0080	0.0016	(*)
4	0.0944	0.0063	0.1083
5	0.0526	(*)	0.0522
6	0.0468	(*)	0.0546
7	0.0000	0.0000	(*)
8	0.0000	0.0000	(*)

Table 2. Average Hammoude metric for each land cover class for SPOT test images. (*) Class not used.

5.1.2 Parcel size. Another relevant aspect that influences the segmentation results is the parcel size and shape. The average values of H were computed for all sub-section of the test images with parcels of the same size. The values of the external similarity indices R, CR and J were also computed for each image sub-section.

For SPOT5d the average values of H were null for all sub-sections, except for the smallest, with parcels of 1 by 1 units, that had a value of H=0.1600. For this image sub-section, the values of the external similarity indices were: R=0.9872, CR=0.8555, J=0.7576. The segmentation result for SPOT5d is perfect, except for the smallest parcels. The image was segmented into 1596 objects (only 4 less than a perfect result) and the overall values were H=0.0025 (or 0.25%), R=0.9998, CR=0.9977 and J=0.9962.

The segmentation results for images SPOT8 and SPOT5s were not as good as for SPOT5d, particularly for SPOT5s due to the fact that all classes are spectrally very close. A more detailed analysis of this image is therefore presented, but the same type of behaviour is exhibited by SPOT8 image.

There is a clear relationship between the score of H, R, CR and J with the parcel size, with much better results achieved for larger objects. The average value of H for each sub-section of SPOT5s is presented in Table 3. For large parcels, above 3x3 units, H is always below 0.05 (or 5%), except for 1 case (0.55). On the other extreme, for parcels of 1x1 units H=0.715, which is a very bad result.

Size	1	2	3	4	5	6	7
1	.715						
2	.586	.196					
3	.318	.085	.031				
4	.197	.058	.032	.033			
5	.063	.036	.022	.019	.021		
6	.082	.051	.021	.020	.019	.015	
7	.077	.048	.017	.025	.013	.055	.049
8	.078	.025	.022	.026	.016	.017	.040

Table 3. Average Hammoude metric for sub-sections of SPOT5s with fixed parcel size (in units). The value of H for parcels of 8x8 units is 0.046.

An alternative way to analyse the segmentation performance as a function of the parcel size is to use plots, such as the example presented in Figure 5, where the four indices H', R, CR and J are plotted for rectangular parcels of x by 1 units. In this plot the inverted Hammoude index (H') is presented instead of H, so that an easier comparison between the various indices can be made. Figure 5 is illustrative of the relative performance of the four indices. The Rand index is not very useful, as it tends to rate highly almost all sub-windows tested, thus reducing its capability to discriminate. The values of the three other indices

(H, CR, and J) are generally consistent, all providing valuable information to evaluate the segmentation results.

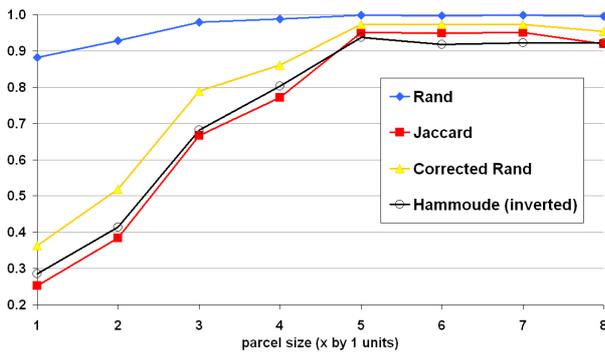


Figure 5. Segmentation performance for SPOT5s as a function of the parcel size (x by 1 units of 3x3 pixels each).

The performance of the segmentation process does not depend only on the parcel size, but also on its shape. Comparing groups of parcels with approximately the same area (for example 8 by 1, 4 by 2 and 3 by 3, or 8 by 2, 5 by 3 and 4 by 4) it can be seen that the performance varies consistently with the shape, with rectangles with lower ratios between the largest and smallest sides being favoured (see Table 3). One parameter that can characterise the shape of a parcel is the ratio between interior and border pixels ($r_{i/b}$). For example, for a single unit parcel of 3 by 3 pixels, there is only 1 interior pixels and 8 pixels on the border, thus the value of $r_{i/b}$ is 0.125. In the plot of Figure 5, not only the area increases with the scale parameter x, but also the values of $r_{i/b}$. This analysis reveals that the $r_{i/b}$ ratio seems to be more important than the parcel area in the segmentation performance. This is not a surprise, as the segmentation errors occur mostly in the objects edges, due to neighbouring parcels with close spectral signatures.

5.2 Test images from Landsat TM

The summary of results obtained for the Landsat TM test images is presented in Table 4. The global segmentation results exhibit the same relationship as for the SPOT test images, with the best scores attributed to TM5d, followed by TM8 and TM5s. However, the global results for TM are worst than for SPOT, particularly for the test image with 8 classes ($H=0.0662$ vs 0.0430) and TM5d ($H=0.0575$ vs 0.0025).

Class ID	TM8	TM5d	TM5s
No. objects	1810	1966	1543
H	0.0662	0.0575	0.0859
H'	0.9338	0.9425	0.9141
R	0.9957	0.9974	0.9880
CR	0.9694	0.9797	0.9424
J	0.9572	0.9670	0.9274

Table 4. Global segmentation results for the TM test images.

Another significant feature displayed in Table 4 is the total number of objects produced by the segmentation process. Image TM5s was segmented in 1543 objects (almost the same as for SPOT5s – 1545), but the number of objects in the segmented images TM8 and TM5d largely exceeds the expected (reference) number of 1600 objects. This fact can be explained by an evaluation of the segmentation performance for the various land cover classes. The average value of H computed for each land cover class is presented in Table 5 for the TM test images. There are some important differences between these

results and those presented in Table 3 for SPOT. The land cover class sand (ID=7), which had a perfect results with SPOT, is in fact the worst class for the TM test images. This is the main reason for the poor global performance of TM8 and TM5d. The parcels of sand are segmented into multiple objects in the TM images, possibly due to the additional short wave infrared bands available.

Class ID	TM8	TM5d	TM5s
1	0.0261	0.0136	0.0750
2	0.0454	(*)	0.0639
3	0.0526	0.0452	(*)
4	0.0655	0.0455	0.0578
5	0.0696	(*)	0.1160
6	0.1008	(*)	0.1169
7	0.1697	0.1835	(*)
8	0.0000	0.0000	(*)

Table 5. Average Hammoude metric for each land cover class for Landsat TM test images. (*) Class not used.

5.3 Parameter setting for SPOT5s

The influence of the eCognition segmentation parameters Scale, Color and Smoothness were evaluated using the SPOT5s image.

5.3.1 Scale. The number of objects produced by the segmentation process greatly varies with the values assigned to the scale parameter, from 1087 (scale=20) to 25990 objects (scale=4). The closest to the reference value occurs for the default value of 10 (1545 objects). The global indices also vary considerably with this parameter. The best value of H is 0.0821 for scale=10 and the best CR is 0.9910 for scale=8.

The average value of H for image sub-windows with parcels of 2x2, 4x4, 6x6 and 8x8 units is presented in Figure 6. The dotted line is the global value of H. The plot shows that the best choice for the scale parameter varies with the parcel size: 8 for parcels of 2x2 units, 10 or 12 for parcels of 4x4 and 16 to 20 for larger parcels.

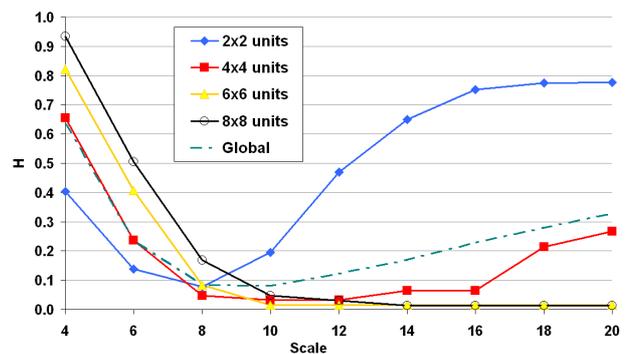


Figure 6. Average Hammoude metric for selected sub-sections of SPOT5s as a function of the scale parameter.

5.3.2 Color. The segmentation results do not vary as much with the parameter color as with scale. For the range of values tested, the number of objects increases with the value assigned to the color parameter, with a minimum of 1437 objects (color=0.50) and a maximum of 1592 objects (color=0.95). The best value of H is achieved for color=0.90 and the best value of CR is achieved equally for color=0.90 and color=0.95.

The average value of H for parcels belonging to each land cover type are presented in Figure 7, as a function of color, as well as the global value of H (dotted line). There is a general trend for H to decrease with the value assigned to the color parameter,

with the best results obtained for 0.75 (class 5), 0.85 (class 1), 0.90 (classes 2 and 3) and 0.95 (class 4).

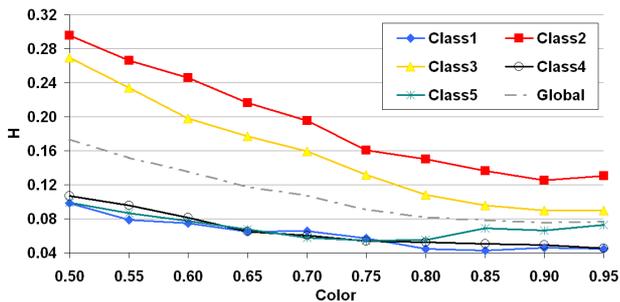


Figure 7. Average Hammoude metric for the land cover types of SPOT5s as a function of the color parameter.

5.3.3 Smoothness. The importance of the smoothness parameter in the final segmentation result is not as important as scale or color. The total number of objects ranges between 1526 and 1552. The global values of H and CR indicate that the best choice for this parameter is 0.55, with a slow but steady decrease in overall segmentation quality as the value assigned to color increases. This can be confirmed in Figure 8, where the plots of global and class averaged H are presented as a function of smoothness. The values of H are nearly unchanged for smoothness between 0.50 and 0.75, and then there is a slow increase in H for all classes, between 0.75 and 1.00.

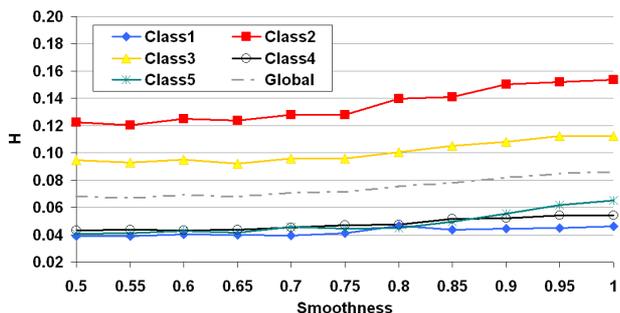


Figure 8. Average Hammoude metric for the land cover types of SPOT5s, function of the smoothness parameter.

6. CONCLUSIONS

The proposed methodology permits an evaluation of the performance of segmentation algorithms applied to a multi-spectral satellite images. As the approach is based on the production of multi-spectral synthetic images with controlled spatial and spectral characteristics, it is possible to overcome one of the main limitations in the evaluation of segmented satellite images, which is the lack of a reference segmentation. The methods used to compare segmentations were the Hammoude metric and the external similarity indices and, Corrected Rand and Jaccard. The Rand index was found to have a very limited range, but the other 3 indices were found to provide meaningful information consistently. The Hammoude metric is more convenient to evaluate the performance for distinct land cover types, while the similarity indices are more effective for parcel size evaluation. The Jaccard index has the advantage of being the easiest to compute. The experience carried out clearly revealed that the segmentation result obtained for a multi-spectral image not only depends on the geometric properties of the objects present in

the image, but also on their spectral characteristics. In fact, the selection of land cover types was found to be the single most important factor that affected the segmentation results.

The method proposed can be used to fine tune the parameters of a segmentation algorithm to a specific dataset and application, as the example presented illustrates. In the example with the e-cognition software the parameter scale was found to be the most important, and the parameter smoothness the less important.

The software developed to create synthetic images and to compute the Hammoude metric and external similarity indices is freely available at www.fc.up.pt/sitef. The images used in this paper are also available in that web site.

ACKNOWLEDGMENTS

The authors would like to thank the OASIS programme (n° 307) for providing the SPOT satellite image and Isabel Poças for the selection and identification of land cover classes.

REFERENCES

- Baatz, M., Benz, U., Dehghani, S., Heynen, M., Höltje, A., Hofmann, P., Lingenfelder, I., Mim-ler, M., Sohlbach, M., Weber, M., Willhauck, G., 2001. eCognition Object Oriented Image Analysis, User Guide, Definiens Imaging GmbH, Munich, Germany, 426pp.
- Baatz, M., Schape, A., 2000. Multiresolution Segmentation: an optimization approach for high quality multi-scale image segmentation, In: Strobl, J., Blaschke, T. (Eds.) *Angewandte Geographische Informationsverarbeitung*, vol. 12. Wichmann-Verlag, Heidelberg, pp. 12–23.
- Chalana, V., Kim, Y., 1997. A Methodology for Evaluation of Boundary Detection Algorithms on Medical Images. *IEEE Transactions on Medical Imaging* 16, 642-652.
- Dubes, R.C., 1987. How many clusters are best? – An experiment. *Pattern Recognition* 20, 645-663.
- Hammoude, A., 1988. Computer-assited endocardial border identification from a sequence of two-dimensional echocardiographic images. Unpublished Ph.D. dissertation, University of Washington, Seattle, WA.
- Hay, G.J., Castilla, G., 2006. Object-Based Image Analysis: Strengths, Weaknesses, Opportunities and Threats (SWOT). In: *OBIA, 2006: The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*.
- Hay, G.J., Castilla, G., Wulder, M. A., Ruiz, J.R., 2005. An automated object-based approach for the multiscale image segmentation of forest scenes. *International Journal of Applied Earth Observation* 7, 339–359.
- Hubert, L., Arabie, P., 1985. Comparing Partitions. *Journal of Classification* 2, 193-218.
- Moller, M., Lyburner, L., Volk, M., 2007. The comparison index: A tool for assessing the accuracy of image segmentation. *International Journal of Applied Earth Observation* 9, 311-321.
- Rand, W.M., 1971. Objective Criteria for Evaluation of Clustering Methods. *Journal of the American Statistical Association* 66, 846-850.