COMPARISON OF THREE SEGMENTATION METHODS FOR GROVES RECOGNITION IN VERY HIGH RESOLUTION SATELLITE IMAGES

MOUGEL Baptiste[1], LELONG Camille[1], NICOLAS Jean-Marie[2]

1 UMR TETIS – CIRAD - Maison de la Télédétection – 500 Rue JF Breton Montpellier, France
2 Departement TSI – ENST Paris – 46, rue Barrault – 75634 Paris cedex 13, France

ABSTRACT

This study is dedicated to the automatic recognition and mapping of tree crops by remote sensing, using very high resolution multi-spectral satellite images (0.7 m).
Our goal is to segment the images in order to perform an independent classification according to a set of pre-determined land use types: apple groves, vineyards, miscellaneous young and old groves, pastured and cropped fields, food crop, fallow lands and forests. In this article, we compare three methods of segmentation that seem to provide suitable units for the resolution of our problem: SxS, eCognition and watersheds. A set of criteria are defined to quantitatively analyze the efficiency of these segmentations. We then try to select the more relevant method in terms of subsequent classification operability.

1. INTRODUCTION

This study is part of the ORFEO methodological program led by CNES (the French Space Agency) and several research institutes for the development of algorithms dedicated to image processing of the future Very High Spatial Resolution (VHSR) PLEIADES sensor. This part of the project aims at developing automatic tools for the recognition of landscapes elements, such as groves and other tree plantations.
Since 2001 a new generation of satellite sensors delivers more accurate details and information of the Earth surface. We are now able to distinguish individual trees in VHSR satellite images. They should allow a better identification of the landscape units based on their content like, for the groves, the identification of the species or the crop system. Although current treatment of this type of images goes back to airborne photographs and are based on visual photo-interpretation, this technique is time consuming and it is thus necessary to develop new tools based on computer processing for more automatic extraction of spatial information.
We propose to develop a VHSR-image processing sequence in an attempt to improve current practices in this field. The first step consists in a segmentation of the image in homogeneously textured units using tools previously published in the literature. The second step corresponds to an independent classification of the obtained units into a set of pre-defined classes; the segmentation preparing the best data set to be efficiently classified.
In this paper, we present a first step of this treatment.
As the number of operational segmentation tools proposed in the literature is considerably high, we first select a set of three of them that seem well adapted to the resolution of our problem. This paper presents the results of these segmentations in the case of a typical agricultural area, and compares them in terms of classification and grove recognition potentialities.
Then some perspectives are set for the classification of the resulting units into expected themes.

2. DATA-SET

The input data-set for this test-study is extracted from a Quickbird [1] image acquired in July 2005 with a spatial resolution of 0.7 m in the panchromatic mode and 2.5 m in the multispectral mode (three spectral bands in the visible (blue, green and red) and near-infrared domains). These data were merged using a Brovey transform [2] to provide a multispectral image at 0.7 m spatial resolution, similar to that of the future PLEIADES products.
The test-site is located in the South of France, between the cities of Nîmes and St Gilles. It is composed of a large variety of land use types, including several types of groves, orchards and forests. The analyzed image frame dimensions are 2411 pixels x 2122 pixels (2.5 km^2) to allow quick computing.

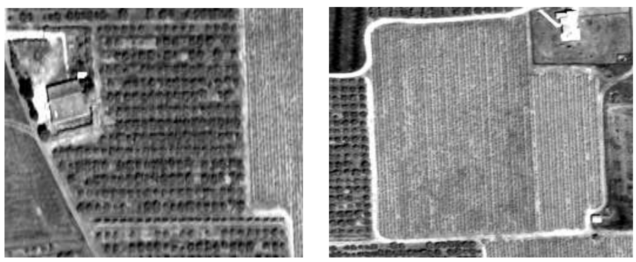


Fig 1: Test-image sample: peach grove (left) and vineyard (right)

3. METHODOLOGY

Here are introduced the three different methods chosen to be tested during the segmentation step of this study and the means proposed to compare their results.

3.1. Multistep segmentation

eCognition® [3][4] is a software developed for segmentation and classification by the Definiens company. This software allows a quick and easy, step by step segmentation. At each segmentation step, a new adjustment is carried out taking the result of the previous step as an additional constraint for the new segmentation step.
Several parameters can be tuned to modify visual segmentation and the process is repeated until a satisfactory result is obtained. The different parameters that can be adjusted are:
-Scale Parameters, for parcel size,
-Homogeneity Criterion, composed of two sub-parameters to regulate homogeneity according to color and scale.
The main disadvantage of eCognition is its "black box" nature, where one does not really control the segmentation but only tunes some poorly documented parameters as long as the result is not good enough.

3.2. Hierarchical segmentation

SxS [5] is a software in CeCILL license, developed by Guigues at IGN (French Geographic Institute). SxS uses a multi-scale approach for images segmentation. The principle of this method is to establish a hierarchy based on the scale that merges the more coherent areas.

3.3. Classical segmentations

OTB (ORFEO ToolBox) [7] is a library developed by CNES (French National Space Studies Center) for VHSR satellite image processing. This toolbox comprises several traditional methods of segmentation such as: watersheds, connected threshold segmentation, etc....[8]. This library allows us to quickly test traditional segmentation methods even if they do not seem to fit the problem at first sight.

3.4. Methodology of evaluation

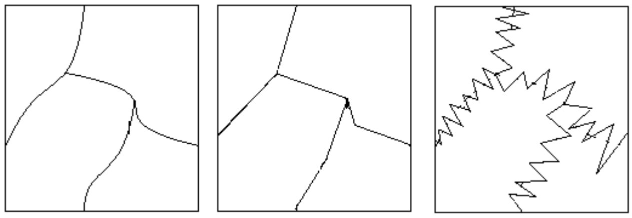
To evaluate the 3 segmentations we will calculate several indicators extracted from result images where each parcel's pixel contains a single label, the same for the whole given parcel. These images will be called "label images". In addition, a label image is also created based on the ground truth. Segmentation results should then bring the closer than possible to the ground truth. To estimate this degree of concordance, we propose to calculate the following indicators.

3.4.1. Parcels number

This indicator is basic and very simple to derive. But it shows us easily if the method of segmentation under or over-segments the image.

3.4.2. Perimeter

The perimeter of the pieces is a very good indicator of the contour complexity of the segmented areas. Indeed, ,one will prefer to have the smaller perimeter than possible because it brings closer to what a human would trace to segment the image. This is illustrated on figure 2 where two segmentations in four pieces are shown: the preferred one will be the middle one.



A B C

Fig 2 : Illustration of perimeter complexity
A= Real B=small perimeter C= long perimeter

3.4.3. Inner-covering

This indicator is derived on a set of 2 label images of the same dimensions. In our case, the first image is the ground truth reference (A), the second is a label image resulting from one of the tested methods of segmentation (B).
The "covering matrix" C is then of a number of column equal to the number of different labels in A, and a number of lines equal to the number of different labels in B. The value of the pixel C(i,j) corresponds to the number of pixels in the whole image frame that have the label i in A and the label j in B.
From this matrix C we derive two indicators in the form of matrices:

- C1: the maximum of the percentages calculated in a single column, corresponding to the proportions of pixels segmented in the parcel j that actually belongs to the parcel i,
- C2: the maximum of the percentages calculated in a single line, corresponding to the dispersion of pixels segmented in the parcel j into the different actual parcels.

A B

1	1	2	2
1	1	2	2
1	1	3	3
4	4	4	4

covering matrix C

a	a	c	c
a	a	c	c
b	b	b	c
d	d	e	e

C1

	a	b	c	d	E
1	4	2	0	0	0
2	0	1	4	0	0
3	0	0	1	0	0
4	0	0	0	2	2

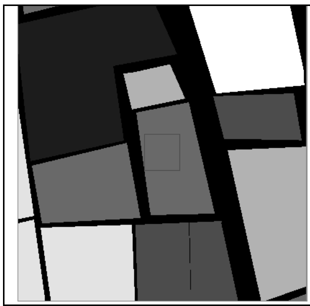
	a	b	c	d	E
1	100	66	0	0	0
2	0	33	80	0	0
3	0	0	20	0	0
4	0	0	0	100	100

Fig 3 : Images of indicators matrix
A= *Images of Ground-truth* B=*Label Image*

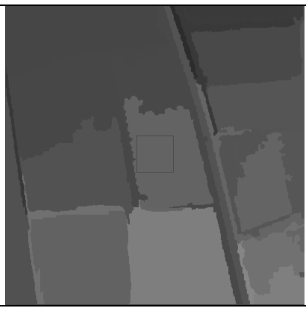
The maximum of percentage found in one column in C1 (maxC1) and in one line in C2 (maxC2) are relevant indicators of over or under-segmentation in the cases of very high and very low values. For instance, if both maxC1 and maxC2 are very high in the bin 90-100%, the result will be very close to the ground-truth.

4. RESULTS AND DISCUSSIONS

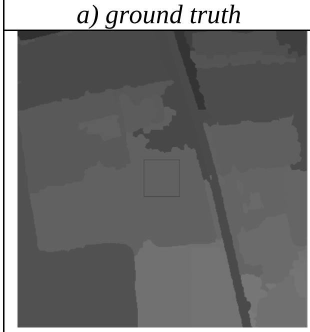
4.1. Result images



a) ground truth



b) eCognition



c) SXS

d) OTB : watersheds

e) Pir-Red-Green bands composition displayed in gray levels

Fig4 : Label-images obtained a) with the ground-truth data, b) with the segementation of eCognition, c) with the SXS algorithm and d) with the watershed segmentation of the OTB for the same extract of the multispectral image e).

The results show at first sight that we can quickly discard the traditional method of watersheds. Indeed it is obvious that this method doesn't segment effectively our case-image: a large group of parcels are not separated while some other areas are split into too many small "parcels" of no actual meaning. As a simple indicator, the number of segmented parcels is 268303 while there are actually 277 parcels in the ground-truth image. This is due to the fact that THRS images are strongly texturized, inducing a high local variation of energies that are taken into account in the watershed thresholding. Moreover, this method requires many adjustments, what is not easily compatible with semi-automatic data processing sequences. We thus choose not to explore further this method of segmentation with quantitative evaluation. Only the results of eCognition and SXS will be analyzed in the following

4.2. Result indicators

4.2.1. Parcels number

Parcels number in the label images			
Method	Ground-truth	eCognition	Sxs
Result	277	416	673

This first indicator shows that eCognition and Sxs find a more important parcels number than the ground-truth: 416 and 673 respectively, vs 277.
Considering that the ground truth map has been established to fit the end-user point of view, namely a classification of orchards and groves, it has been produced more like a classification ground-truth than a segmentation one. This led to the merging of several areas in a single parcel though they are heterogeneous in texture and radiometry. The main example is the network of roads and paths that are not the subject of study here (see the black areas in fig.4.a)). To analyse if this limit is the reason of the high number of segmented parcels with eCognition or Sxs compared to this ground-truth, we propose to mask to the label-images, discarding all the pixels included in the black areas of the ground-truth image.

Parcels number in the masked label images			
Method	Ground-truth	eCognition	Sxs
Result	276	409	516

The parcels number falls to 276 for the ground truth image vs 409 for eCognition and 516 for Sxs. This shows that the high number of parcels obtained with Sxs is effectively due in a large part to this character of the ground-truth image but not for eCognition. The resulting values indicate that both segmentations over-segment the image, but are quite acceptable, with a better mark for eCognition. But this indicator alone does not seem to be efficient enough to evaluate the results.

4.2.2. Perimeter

Method	Ground-truth	eCognition	Sxs
Result	317666	386245	320837

The perimeter indicator gives values quite similar to that of the ground-truth, Sxs being more efficient than eCognition. Indeed, the perimeter of Sxs' parcels (320837 pixels) are inferior to eCognition's (386245 pixels) and thus are simpler.

4.4.3. Inner-covering

- *Ground-truth vs eCognition*

%	00-10	10-20	20-30	30-40	40-50	50-60	60-70	70-80	80-90	90-100
C1	245	60	37	28	11	17	17	15	27	184
C2	367	165	70	57	31	34	46	49	92	167
%	00-10	10-20	20-30	30-40	40-50	50-60	60-70	70-80	80-90	90-100
max C1	94	14	17	7	5	4	3	2	6	126
max C2	93	21	13	11	8	12	18	23	74	144

- *Grounf-truth vs Sxs*

%	00-10	10-20	20-30	30-40	40-50	50-60	60-70	70-80	80-90	90-100
C1	324	74	35	36	23	21	19	12	27	169
C2	247	208	116	74	69	65	62	73	87	318
%	00-10	10-20	20-30	30-40	40-50	50-60	60-70	70-80	80-90	90-100
max C1	96	13	7	5	5	6	5	2	10	129
max C2	116	31	30	30	17	33	24	38	80	273

One can notice that the higher scores of parcels in C1 and C2 are for the extreme bins of percentages: 0-10% and 90-100%.

Possible cause of the large amount of small contributions to others parcels (0-10% bin) is the rectilinear segmentation of the ground-truth image that any automatic algorithm won't produce. The ground-truth label-image parcels borders are adjusted to the simplest limit shape (polygon) though in reality they are not rectilinear: radiometry and texture are not homogeneous at the borders of the parcels, which provokes automatic algorithm to skirt these heterogeneities. But at this stage of analysis it is impossible to attribute these values to this only cause, thus it is very difficult to evaluate if the two segmentations are efficient enough. Nevertheless, both having the same scores in the small bins 0-10%, their performance is identical in regards of this criteria.

On the other side, the high number of parcels in the 90-100% zone of C1 and C2 indicate that 1) a single actual parcel is highly covered by a single segmented parcel 2) a single segmented parcel is contained in only one actual parcel. The both criteria being fulfilled, the two segmentation methods fit thus quite correctly the ground-truth. Considering the higher score of Sxs in C2 and maxC2 (twice eCognition's scores) it seems that Sxs give more spatially coherent results.

5. CONCLUSION AND PERSPECTIVES

As a conclusion, Sxs' results are slightly better than eCognition's according to our indicators: the total perimeter of segments is lower and the spatial determination and cohesion are better even if it slightly more over-segments the image. In addition, Sxs method is also much simpler to use, more automatic and with fewer parameters to tune (the cut level in the hierarchical pyramid in Sxs vs 3 parameters for each stage of eCognition segmentation). Thus, it seems that Sxs is more efficient to segment our kind of image than eCognition.

Nevertheless, this work still needs some improvements and further analysis should be driven especially in the frame of validation/evaluation of the two segmentations. For instance, an indicator which would couple the information contained in C1 and C2 should be proposed in order to evaluate the position of the segmented parcels. More accurate and more intelligible indicators should also be defined to fix the results. Other pertinent segmentation methods could also be added to the comparison.

6. REFERENCES

[1] http://www.digitalglobe.com/about/quickbird.html

[2] C. Pohl and J.L. Van Genderen. Multisensor image fusion in remote sensing: Concepts, methods and applications. International Journal of Remote Sensing, 19 (5), 823-854 (1998).

[3] http://www.definiens-imaging.com/ecognition/

[4] C. Burnett and Thomas Blaschke. A multi-scale segmentation/objet relationship modelling methodology for landscape analysis. Ecological Modelling 168, 233-249 (2003).

[5] L. Guigues, H. Le Men and J.-P. Cocquerez. Scale-sets image analysis. In Proc. of IEEE Int. Conf. on Image Processing (ICIP'03), Barcelona, Spain, September 2003.

[6] L. Guigues. Modèles Multi-Echelles pour la Segmentation d'Images. PhD thesis of *Cergy-Pontoise University*, 287p. (2003).

[7] The ORFEO Tool Box Software Guide. Second Edition Updated for OTB-1.0.1 page 164-170 (2006).

[8] http://www.itk.org

8. AKNOWLEDGEMENTS

This study was funded by the Centre National d'Etudes Spatiales (CNES), the Institut Géographique National (IGN), the Centre de Coopération Internationale en Recherche Agronomique pour le Développement (CIRAD) and the Programme National de Télédétection Spatiale. The authors want to thank Agnès Bégué, Didier Boldo, Jordi Inglada and Michel Deshayes for helpful discussions concerning this work.