ASSESSMENT OF REMOTE SENSING IMAGE SEGMENTATION QUALITY

M. Neubert *, H. Herold

Leibniz Institute of Ecological and Regional Development (IOER), Weberplatz 1, D-01217 Dresden, Germany - (m.neubert, h.herold)@ioer.de

Commission IV, Working Group IV/4

KEY WORDS: Comparison, segmentation, quality, evaluation methods, software, remote sensing, IKONOS, high resolution

ABSTRACT:

Image segmentation is a crucial step within the object-based remote sensing information retrieval process. As a step prior to classification the quality assessment of the segmentation result is of fundamental significance for the recognition process as well as for choosing the appropriate approach and parameters for a given segmentation task. Thus, this research is also related to the topic of object based accuracy assessment. In this paper we present some methodical extensions of the segmentation quality evaluation process based on our previous studies. The main focus is set to increasing automation, new metrics and higher regard to spatially explicit metrics.

Object differences have been analyzed as topological and geometric relationships between the segment and the according reference object. Thus, the overlapping area was calculated (absolute and percentage) to describe the area concurrence. Furthermore, the accordance of the outlines was evaluated using buffer zones around reference objects by means of proportion inside specific buffer zones. This makes it possible to draw conclusions about the geometrical correctness of the segmented outlines. In addition to that we investigated other published assessment metrics such as the Area-Fit-Index. Results of several segmentation programs have been assessed and compared using identical imagery. The software tested is: ENVI Feature Extraction Module 4.4, BerkleyImgseg 0.54, EDISON, EWS 1.0, Definiens Developer 7 and InfoPack 2.0. Some newly available programs point out new possibilities for object-based image analysis.

Conclusions from a methodical and users point of view will be given. In combination with the previous studies, in total 24 segmentation programs or its releases have been evaluated. The results of all segmentations are displayed at the website www.ioer.de/segmentation-evaluation.

1. INTRODUCTION

Image segmentation is recently widely used in remote sensing especially since the availability of very high resolution imagery and it is leading to a new object-oriented paradigm (Navulur, 2007). Since we published previous studies (Neubert et al., 2006, Neubert et al., 2008) the number of segmentation algorithms is still growing. Thus, the evaluation of is still helpful to compare the suitability of the algorithms regarding the user needs as well as the resulting quality.

2. EVALUATED SEGMENTATION SOFTWARE

There is a permanent growing variety of implemented segmentation algorithms using very different concepts. Within this paper we extend the previous evaluations (published in Neubert et al., 2008) by two new algorithms (see table 1 for details):

- *ENVI Feature Extraction Module 4.4* (ITT Visual Information Solutions, Boulder, USA);
- *BerkleyImgseg 0.54* (BETI - Berkeley Environmental Technology International, LLC, Berkeley, USA).

Some further algorithms have been tested without success: ImageSeg v0.1 (Texture-Based Image Segmentation, ITT Visual Information Solutions) and SAGA (System for Automated Geoscientific Analyses, University of Göttingen). The first fails due to the image size even if the input file is prepared as required (image must be square and have an edge length with factor 2^n). The latter produces extremely over-segmented results with the implemented grid segmentation module. The region-growing module library described in Böhner et al. (2006) is currently not available for testing.

Furthermore, three recently developed algorithms are announced to be included in the comparison as soon as the results are available.

3. EXTENDED EVALUATION METHODS

In extension to the overview given in Neubert et al. (2006) and Neubert et al. (2008) we are focusing on the design of new spatially explicit metrics and the implementation of the Area-Fit-Index (AFI), suggested by Lucieer (2004). A detailed description of the applied evaluation methodology can be found in our previous studies.

3.1 Design for spatially explicit evaluation metrics

Previously, the quality of the segmentation results was tested by comparing segmented objects with reference objects using formal properties only. Thus, we looked for simple but effective metrics for the spatially explicit evaluation of the segmentation results to extend the existing method.

* Corresponding author.

Segmentation program		BerkleyImgseg 0.54	ENVI Feature Extraction Module 4.4
Fundamentals	Developer	Berkeley Environmental Technology International, LLC	ITT Visual Information Solutions
Fundamentals	Website	www. berkenviro.com/ berkeleyimgseg/	www.ittvis.com/ envi/ featureextraction/
Fundamentals	Algorithm	Region merging	Edge-based (Full Lambda-Schedule algorithm for region merging)
Fundamentals	Field of application	Remote sensing	Remote sensing
Fundamentals	Fundamental reference	(see website)	ITT Visual Information Solutions (2007)
Implementation	State of development	05/2008	10/2007
Implementation	Operating system	Win	Win
Implementation	System environment	Stand-alone	ENVI 4.4
Implementation	Number of parameters	3	1 (2 more for refinement)
Implementation	Ca. runtime[1,2]	20 min	1 min
Implementation	Reproduce-ability[3]	No	No
Implementation	Classification support	Yes	Yes
In- and Output	Max. image size [ca. Pixel][1]	No Limitations	No limitation
In- and Output	Max. bit depth	32	64 (no limitation)
In- and Output	Input formats	Raster (TIF, JPEG, etc.)	BSQ, BIP, BIL
In- and Output	Vector output format	Shapefile	Shapefile
In- and Output	Use of external data	Yes	Yes (raster only)
	Availability	Commercial (30-day full-featured trial)	Commercial (14-day full-featured trial)

[1] Specification heavily depends on system resources, particularly main memory; [2] Specifications for the used imagery (2,000 by 2,000 Pixel); [3] When image extend is modified.

Table 1. Outline of evaluated segmentation software.

The object differences can be analyzed as topological and geometric relationships between the segment and the according reference object. Therefore, both the area coverage and the outline situation are important. For a good segmentation the congruence of both factors should be maximized. The aim is to calculate the percentage of concordant area as well as the proportion of correctly delineated outline.

Thus, the overlapping area is calculated (absolute and percentage) to describe the area concurrence. Furthermore, the accordance of the outlines was evaluated using buffer zones around reference objects by means of the proportion inside specific buffer zones. This makes it possible to draw conclusions about the geometrical correctness of the segmented outlines.

In a first step the reference objects are buffered using 1 m buffer zones (see figure 1). Since the image resolution of the used pan-sharpened IKONOS imagery is 1 m we suggest calculating three buffer zones. Using other imagery the buffer distance should be set according to the image resolution. We assume that the situation of the segment outline within a 2 m buffer can be rated as good since most of the segmentation results are raster-oriented in opposition to vectorized reference outlines.

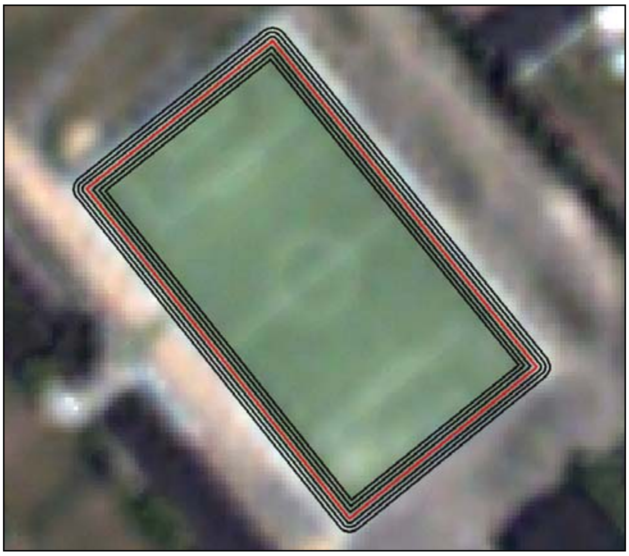

Figure 1. A soccer field as an example reference object (red outline) and three buffer zones in 1 m distances (black)

The segmentation result is than fused with the buffered reference objects (see figure 2). Like in the existing method all partial segments with at least 50 % area within a reference object were merged before comparison. Then the proportions of the overlapping area and the outline length within the inside and the buffer zones are calculated (absolute and percentage each).

For the example shown in figure 2 the resulting proportions are given in the following tables. The results of this example demonstrate that the area overlap of the reference and the segmented object is 92.8 % while 99.0 % (sum of area within inside buffer) of the segment is situated inside the reference (see table 2). Thus, 7.2 % (383 m²) of the reference object is not covered or segmented properly respectively (area gap). The area protruding beyond the reference outline is only 1 % and almost situated in the 1 m outside buffer what is an acceptable result.

The delineation quality evaluation shows that only 47.9 % of the outline is situated within the 1 m buffer. After all, 15.4 % of the outline has a variance of more than 3 m (table 3). If the outline situation within a 2 m buffer is tolerated 70.7 % can be treated as well-segmented. The overall outline length difference is 4.7 % (426 m reference vs. 447 m segmented).

For comparing the different segmentations the average values for all reference areas will be cumulated in further analyses.

This method is currently going to be implemented in our GIS-based evaluation tool.

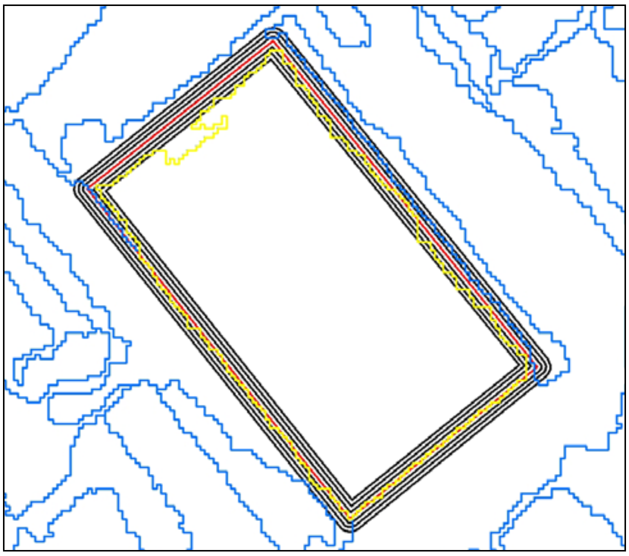

Figure 2. Segmentation result (blue) and selected segmented object for comparison (yellow)

	Area proportions of the segmented object	Percentage	Percentage compared to the area of the reference object
Inside	4400	87.8	82.3
3 m buffer inside	236	4.7	4.4
2 m buffer inside	188	3.8	3.5
1 m buffer inside	137	2.7	2.6
1 m buffer outside	45	0.9	0.8
2 m buffer outside	4	0.1	0.1
3 m buffer outside	0	0.0	0.0
Outside	0	0.0	0.0
Total	**5010**	**100.0**	**93.7**

Table 2. Area proportions within the buffer zones for the object shown in figure 2

	Line proportion [m]	Percentage [%]
1 m buffer	214	47.9
2 m buffer	102	22.8
3 m buffer	62	13.9
Inside	69	15.4
Outside	0	0.0
Total	**447**	**100.0**

Table 3. Outline proportions within the buffer zones for the object shown in figure 2

3.2 Area-Fit-Index (AFI)

To quantify the fit of each of the reference objects with the largest segment overlapping them, the Area-Fit-Index (AFI) was introduced by Lucieer (2004) and is defined as follows:

$$AFI = \frac{A_{\text{reference object}} - A_{\text{largest segment}}}{A_{\text{reference object}}}$$

where A is the Area of the reference and the largest segment of the result respectively. If the AFI equals 0 a perfect segmentation is indicated. The index was tentatively integrated in our evaluation procedure. The mean AFI and the AFI range were calculated using the 20 reference objects and are shown in Table 4 for the recently tested algorithms.

4. RESULTS AND DISCUSSION

4.1 Visual Quality Assessment and software specifics

ENVI Feature Extraction Module 4.4: The main advantage of this ENVI extension is the ability to preview results of each step in real time (esp. the impact of chosen parameter settings) what reduces the amount of time needed to optimize parameters. A stepwise procedure (segmentation, refinement, attribute computation) is used by the approach and a minimum set of parameters is needed (normally one per work step). The processing speed is very fast and the embedding of the algorithm in an image processing environment is beneficial. The classification features have not been subject of the evaluation.

The result appears to be very good in overall but it tends to over-segmenting bright image areas (see figure 3). This is maybe reducible by the implemented thresholding. Sometimes small island polygons are created. An internal tiling algorithm seems to be applied by the software since image objects are divided sometimes (2000 pixels left and below of the image border). The evaluated result is an intermediate result of the segmentation process (raw pixel outlines). The software has the ability of smoothing the output in a step after the classification.

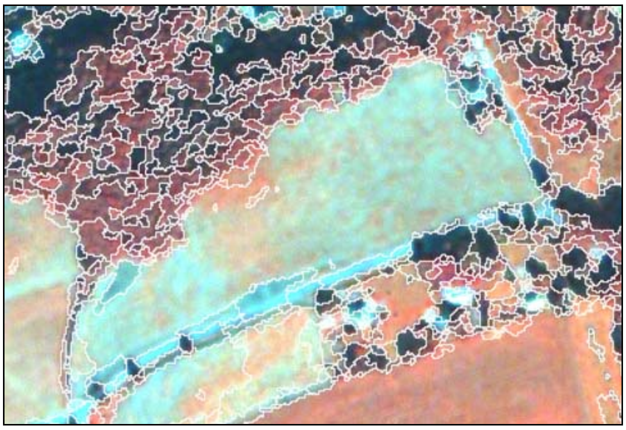

Figure 3. Segmentation result of ENVI Feature Extraction Module 4.4.

BerkleyImgseg 0.54: The segmentation result of Berkley-Imgseg 0.54 fit most of the object classes well; especially inhomogeneous forested and agricultural areas were good delineated (see figure 4). Finding optimal parameters is a tradeoff between the very good delineation of compact natural regions of different size and the segmentation of man-made objects such as streets and long buildings. For this reason, different parameter sets have to be used for the rural and the urban scene.

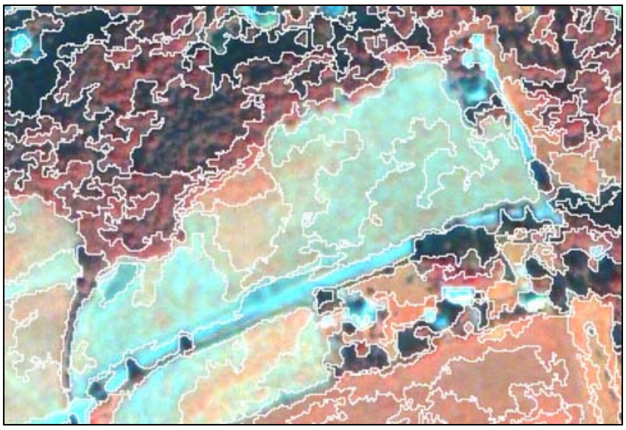

Figure 4. Segmentation result of BerkleyImgseg 0.54

The software allows the user to provide training segmentations and contains additional classification support such as Kmeans, K-Nearest Neighbor, and Neural Networks. It also offers a feature to convert the raster segmentation result into ESRI Shape format (vectorization).

The descriptions of the other evaluated algorithms showed in table 4 are to be found in Neubert et al. (2008).

4.2 Comparison Based on Reference Areas

In addition to the visual assessment, all segmentations were quantitatively (objectively) evaluated by means of 20 reference areas. The overall results are cumulated and compared in table 4. In addition to the recently tested algorithms selected results of our previous study have been added. It can be recognized that the values of the tested algorithms are almost in the same range. Whereas the ENVI Feature Extraction Module 4.4 shows remarkable low derivations in area and Shape Index differences, BerkleyImgseg tends to much less over-segmentation.

In contrast to objective quantitative measures the new implemented Area-Fit-Index corresponds not with the subjective visual rating. However, it might a suitable indicator for rating under- and over-segmentation, as it also emphasizes the differences between the reference objects.

5. CONCLUSIONS AND OUTLOOK

In this study the design of new spatially explicit metrics on image segmentation quality assessment has been suggested. Furthermore, two recently released segmentation algorithms have been tested using the conventional methodology and their results where contrasted to some previously evaluated software. The previously used evaluation procedure was extended by the Area-Fit-Index (AFI).

In combination with the previous studies, in total 24 segmentation programs or its releases have been evaluated. The results of all segmentations are displayed at the website www.ioer.de/segmentation-evaluation. Is has been shown, that there is more than one interesting approach in this dynamic field of research. The evaluation will be continued, it is planned to extend the quality assessment procedure itself e.g. by the herein proposed spatially explicit outline delineation quality measures or the Comparison Index proposed by Möller et al. (2007).. The evaluation process is open to further algorithms.

Segmentation program	**ENVI Feature Extraction Module 4.4**	**Berkley-Imgseg 0.54**	**Definiens Developer 7.0**	**EDISON**	**EWS 1.0**	**InfoPACK 2.0**	**Comparison to previously tested programs[1]**	
							Minimum[2]	Maximum[2]
Number of reference areas	20	20	20	20	20	20	10	20
Average difference of area [%]	6.9	12.3	15.9	11.5	24.6	17.0	8.2	2,100.3
Average difference of perimeter [%]	12.3	22.2	17.2	13.8	18.1	29.6	10.0	475.6
Average difference of Shape Index [%]	10.8	21,1	16.2	12.4	15.4	46.0	10.0	87.1
Average number of partial segments	15.9	5,4	1.8	13.4	3.8	21.4	1.8	134.6
Average quality, visual evaluated [0…2][3]	0.8	0.7	0.9	0.8	0.7	0.6	0.0	1.0
Area-Fit-Index [-∞…1]	-0.04	-0.14	0.08	-0.18	-0.12	-0.04	-	-
Range of Area-Fit-Index	0.54	1.19	1.94	1.19	2.25	0.52	-	-

[1] Meinel and Neubert (2004). [2] Values do not represent one algorithm, but the overall minimum and maximum values of each criterion. [3] 0 - poor, 1 - medium, 2 - good.

Table 4. Cumulated results of all 20 reference areas.

REFERENCES

Böhner, J., Selige, T., Ringeler, A., 2006. Image Segmentation using Representativeness Analysis and Region Growing. – In: Böhner, J., McCloy, K. R., Strobl, J. (Eds.). *SAGA – Analyses and Modelling Applications*. Göttinger Geogr. Abh. 115: p. 29-38

ITT Visual Information Solutions, 2007. ENVI Feature Extraction Module User's Guide. 48 p.

Lucieer, A., 2004. Uncertainties in Segmentation and Their Visualisation. PhD Thesis, Utrecht University, ITC Dissertation 113, Enschede, 174 p. http://www.itc.nl/library/Papers_2004/phd/lucieer.pdf (accessed 10 April 2007)

Meinel, G., Neubert, M., 2004. A Comparision of segmentation programs for high resolution remote sensing data. *Int Arch Photogram Rem Sens Spatial Inform Sci* XXXV-B4, pp. 1097-1102. http://www.isprs.org/istanbul2004/comm4/papers/506.pdf (accessed 02 May 2008)

Möller, M., Lymburner, L., Volk, M., 2007. The comparison index: A tool for assessing the accuracy of image segmentation. *Int. Journal of Applied Earth Observation and Geoinformation,* 9, pp. 311-321.

Navalur, K., 2007. *Mulispectral image analysis using the object-oriented paradigm.* CRC Press, Boca Raton, 163 p.

Neubert, M., Herold, H., & Meinel, G., 2008. 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 application*s. Springer, Berlin (in press)

Neubert, M., Herold, H., Meinel, G. 2006. Evaluation of Remote Sensing image Segmentation Quality – Further Results and concepts. *Int Arch Photogram Rem Sens Spatial Inform Sci* XXXVI-4/C42, 6 p. http://www.commission4.isprs.org/obia06/Papers/10_Adaption and further development II/OBIA2006_Neubert_Herold_Meinel.pdf (accessed 02 May 2008)

Neubert, M., Meinel, G., 2003. Evaluation of segmentation programs for high resolution remote sensing applications. In: Schroeder, M., Jacobsen, K., Heipke, C. (Eds.). Proceedings Joint ISPRS/EARSeL Workshop "High Resolution Mapping from Space 2003", Hannover, Germany, October 6-8, 2003 (8 p., CD-ROM). http://www.ipi.uni-hannover.de/fileadmin/institut/pdf/neubert.pdf (accessed 02 May 2008)

ACKNOWLEDGMENTS

The authors thank ITT Visual Information Solutions *(Boulder, USA)* as well as CREASO *(Gilching, Germany)* for the availability of a test version of the ENVI Feature Extraction Module and Dr. Ian McConnell (InfoSAR Limited) for the collaboration and the processing of our sample data.