# EVALUATION OF REMOTE SENSING IMAGE SEGMENTATION QUALITY – FURTHER RESULTS AND CONCEPTS

M. Neubert, H. Herold*, G. Meinel

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

**KEY WORDS:** Segmentation, Quality, Software, Methods, Comparison, High resolution, IKONOS

**ABSTRACT:**

Primarily due to the progresses in spatial resolution of satellite imagery, the methods of segment-based image analysis for generating and updating geographical information are becoming more and more important. In the studies of Neubert and Meinel (2003) and Meinel and Neubert (2004) the capabilities of available segmentation programmes for high resolution remote sensing data were assessed and compared. This paper intends to supplement the preceding studies by considering recently available software. Moreover, a self-implemented optimised segmentation algorithm for the image processing software HALCON is included in the test. The achieved segmentation quality of each programme is evaluated on the basis of an empirical discrepancy method using pan-sharpened multi-spectral IKONOS data. Furthermore, an overview of further methods for quantitative image segmentation quality evaluation is given. Finally, the qualitative and quantitative outcomes are compared and contrasted to the previously tested software solutions. The stated results provide an approach to determine each programme's performance and appropriateness for specific segmentation tasks.

## 1. INTRODUCTION

Image segmentation means the partitioning of an image into meaningful regions based on homogeneity or heterogeneity criteria, respectively (Haralick and Shapiro 1992). It represents the interface between image pre-processing and image understanding (object recognition). Image segmentation techniques can be differentiated into the following basic concepts: pixel-oriented, contour-oriented, region-oriented, model-oriented, and hybrid. Detailed descriptions as well as mathematical backgrounds and evaluations of these algorithms can be found for instance in Haralick and Shapiro (1992), Pal and Pal (1993), and Gonzalez and Woods (1993). This paper considers a more application-oriented comparison based on real remote sensing data.

Image objects in remotely sensed imagery are often homogeneous and can be delineated by segmentation. Thus, the number of elements as a basis for a following image classification is enormously reduced. The quality of classification is directly affected by segmentation quality. Hence the quality assessment of segmentation is within the main focus of this study on different presently available segmentation software. Recent investigations have shown that a pixel-based analysis of such high resolution imagery has explicit limits. Using segmentation techniques some problems of pixel-based image analysis could be overcome (e.g. Meinel et al., 2001). Feature extraction programmes, which perform selective image segmentation, will not be considered in this study.

## 2. EVALUATED SEGMENTATION SOFTWARE

### 2.1 Overview

There is a large variety of implemented segmentation algorithms using very different concepts. They are distributed commercially or are freely available for scientific use. For the evaluation only approaches were considered that are able to perform a full (so-called multi-region) image segmentation in an operational way. Furthermore, the choice of approaches was based on the suitability to segment remote sensing imagery. In addition to the results presented in Meinel and Neubert (2004), the following algorithms and programs were included in the comparison (see table 1 for details):

- *HalconSEG* (Adapted *Lanser*-segmentation algorithm for HALCON, MVTec GmbH, Munich, Germany);
- *Imagine WS* for Erdas Imagine (Austrian Academy of Sciences, Vienna, Austria);
- *PARBAT* (International Institute for Geo-Information Science and Earth Observation, Enschede, Netherlands);
- *RHSEG* (NASA, Goddard Space Flight Center, Greenbelt, MD, USA);
- *SEGEN* (IBM Haifa Research Labs, Haifa, Israel);
- *SegSAR 1.0* (National Institute for Space Research, São José dos Campos, Brazil).

### 2.2 Optimized segmentation algorithm HalconSEG

While all algorithms were tested in their implemented version, the proposed HalconSEG was developed on the basis of the segmentation approach described in Lanser (1993). The original algorithm is a combination of an edge-detection and a region-growing procedure. It was originally designed for the research on segmenting natural images on mobile devices. This approach was adapted and optimized in order to handle and process high resolution remote sensing data by adding various parameterisation opportunities (e.g. for the egde detection filters and the morphological operators), a region-merging algorithm, a hierarchical extension and a GIS-interface proposed in Herold (2005).

* Corresponding author.

| | Segmentation program | **HalconSEG** (Extended Lanser algorithm for HALCON) | **Imagine WS** (Erdas Imagine extension) | **PARBAT 0.32** | **RHSEG 1.0** | **SEGEN** | **SEGSAR 1.0** |
|---|---|---|---|---|---|---|---|
| Fundamentals | Developer | TU Munich/ IOER Dresden | Austrian Academy of Sciences, Commission for Scientific Visualization | Lucieer 2004 | NASA, Goddard Space Flight Center | IBM Haifa Research Labs | INPE, National Institute for Space Research |
| | Website | www9. informatik.tu-muenchen.de/ research/horus/ | www.viskom. oeaw.ac.at/ ~milos/page/ wshed/wshed. html | www.parbat.net | tco.gsfc.nasa. gov/RHSEG/ | www.haifa.ibm. com/projects/ image/segen/ index.html | www.dpi.inpe.br |
| | Algorithm | Hybrid (edge/ region oriented) | Hierarchical watershed | Region growing | Hierarchical region growing | Region growing | Hybrid (edge/ region oriented) |
| | Field of application | Colour images, mobile Systems | Remote sensing | Remote sensing | Remote sensing | Colour images | Remote sensing, esp. radar data |
| | Fundamental reference | Lanser 1993, Herold 2005 | Sramek and Wrbka 1997 | Lucieer 2004 | Tilton 2003 | Gofman 2006 | Sousa et al. 2003 |
| Implementation | State of development | 09/2005 | 01/2003 | 05/2004 | 02/2005 | 05/2006 | 03/2005 |
| | Operating system | Win, Linux, Unix | Win, Linux, Unix | Win, Linux, Unix | Win, Unix (Solaris) | Linux, Win, Unix (AIX) | Win, Linux, Unix |
| | System environment | HALCON 7.0 | Xite, Xite/Erdas | Stand-alone | Stand-alone | Stand-alone | ENVI 4.0 |
| | Number of parameters | 5 (3)[1] | 4 (2)[1] | 3 (2)[1] | 16 (4)[1] | 5 (2)[1] | 5 |
| | Ca. runtime[2,3] | < 2 min | 2 min | 2 min | Several hours | < 1 min | 1 h |
| | Reproduce-ability[4] | No | No | No | No | No | No |
| | Classification support | No | Yes | Yes | Yes (HSEG Viewer) | No | No |
| In- and Output | Max. image size [ca. Pixel][2] | 5,000 x 5,000 | No Limitations | 1,300 x 1,300 | 8,000 x 8,000 | No Limitations | ≥ 2,400 x 2,400 |
| | Max. bit depth | 8 bit | 32 bit | 16 bit | 16 bit | 16 bit | 16 bit |
| | Input formats | Raster (TIFF, RAW) | Raster (IMG) | Raster (BSQ)[5] | Raster (RAW) | Raster (PPM, TIFF) | Raster |
| | Vector output format | ASCII (GEN) | No (Erdas export: Coverage) | No (external conversion) | No (external conversion) | No (external conversion) | EVF |
| | Use of external data | No | No | No | No | No | No |
| | Availability | On request | On request | Freeware | License Agreement | On request | On request |

[1] Parenthesised number: especially relevant segmentation parameters; [2] Specification heavily depends on system resources, particularly main memory; [3] Specifications for the used imagery (2,000 by 2,000 Pixel); [4] When image size is modified; [5] With GDAL-plug-in installed other formats useable, e. g. GeoTIFF, IMG.

Table 1. Outline of evaluated segmentation software.

A further extension allows to import various manually generated segments. The algorithm automatically optimizes the parameterisation to fit the result best to the reference (internal evaluation). It is a contribution to minimize the time needed to find the optimal segmentation parameters. Additionally there will be an extension to visualize and export the so-called uncertainty of segmentation (the boundary stability index) which is proposed in Lucieer (2004). In order to improve the usability in a further version the complete algorithm will run with a GUI using the HALCON/COM interface.

## 3. EVALUATION METHODS

### 3.1 Overview and Related Work

The qualitative and quantitative assessment of segmentation results is very important for further image processing as well as for choosing the appropriate approach for a given segmentation task. The accuracy significantly affects the recognition and classification as well as the derived conclusions. Evaluation studies either intend to compare various segmentation approaches (e.g. Estrada and Jepson, 2005) or different parameterisations of one algorithm (e.g. Palus and Kotyczka,

2001). Only very few studies employ their evaluation on remote sensing data, e.g. Carleer et al. (2005), Karantzalos and Argialas (2003). Mostly natural color images or artificially generated images are used.
Similar to the segmentation theory itself there is no established standard procedure for the evaluation of its results. In literature exists a multitude of very different approaches. A general classification of evaluation methods has been proposed by Zhang (1996), categorising three variants: analytic methods, empirical goodness methods, and empirical discrepancy methods. In recent studies, empirical goodness methods are also referred to as unsupervised evaluation methods, empirical discrepancy methods are denoted as supervised or stand-alone evaluation methods (e.g. Zhang et al., 2005). While most existing approaches are supervised methods using discrepancy measures between a reference and the segmentation, recently much effort is put into the development of empirical goodness methods, which do not require any reference (*a priori* knowledge). However, when comparing different approaches, these methods show a strong bias towards a particular algorithm (Everingham et al., 2002). For this reason in this evaluation an empirical discrepancy method using the *relative ultimate measurement accuracy* has been applied.

### 3.2 Approaches to Quantitative Segmentation Evaluation

As stated before, there is a variety of additional concepts and methods for evaluating image segmentation results. Here, a brief introduction to some prevailing algorithms is presented. The vast majority of the quantitative approaches are basically empirical discrepancy methods, analysing the number of misclassified pixels in relation to reference segmentations. In contrast, other algorithms directly address over- and under-segmentation by considering the number of segments, e.g. the Fragmentation Index *FRAG* (Strasters and Gerbrands, 1991), the Area-Fit-Index *AFI* (Lucieer, 2004) and the Precision/Recall Measure described in Estrada and Jepson (2005).
Similar to the evaluation employed in this paper Yang et al. (1995) used shape features to quantify the differences between segmentation and reference regions. Based on the study of Villegas et al. (1999), Mezaris et al. (2003) presented a distance weighted error measure for misclassified pixels. A *Hausdorf*-distance-based evaluation method for arc-segmentation algorithms is proposed by Liu et al. (2001). A map-algebra-based evaluation approach is introduced in Hirschmugl (2002). An intersection image of the segmentation result and a morphologically dilated binary reference segmentation is used to quantify the number of misclassified pixels. A combined vector-raster-based procedure for assessing the precision of cadastral data using fractal box dimension is introduced by Schukraft and Lenz (2003). Assuming a given reference segmentation, an adapted version of this algorithm is a promising approach to evaluate segmentation quality.
Other evaluation approaches are designed to minimize or exclude the *a priori* knowledge and the subjective (human) bias added to the evaluation by manually created references. Instead of using reference segmentations various objective evaluation criteria such as the intra-regional uniformity of segments are introduced (unsupervised evaluation methods). Cavallaro et al. (2002) present a *perceptual spatio-temporal quality measure* which allows an automated and objective evaluation by considering human perception criteria. However, it only applies to video sequence segmenting. Borsotti et al. (1998) presented an evaluation function which uses the colour uniformity within the segmented regions as criterion. Zhang et al. (2004) introduced a new entropy based evaluation approach, which leads to a very stable assessment measure using different segmentations.
An approach that comprises both analytical and empirical criteria is presented in Everingham et al. (2002) by defining a multidimensional fitness-cost-space instead of a single discrepancy-parameter-space. A promising co-evaluation framework which combines the results of various evaluation approaches using a machine learning approach is proposed in Zhang et al. (2005).
Further and partially older approaches to quantitative evaluation of segmentation results can be found in Yasnoff (1977), Levine and Nazif (1985), Haberäcker (1995), Yang et al. (1995), Schouten and Klein Gebbinck (1995), Zhang (1996), and Letournel et al. (2002). Table 2 provides an overview to recently proposed quantitative evaluation methods within the classification framework given in Zhang (1996).

### 3.3 Applied Evaluation Method

According to the procedure proposed and applied in Neubert and Meinel (2003) firstly all results came under an overall visual survey. General criterions, like the delineation of varying land cover types (e. g. meadow/forest, agriculture/meadow, etc.), the segmentation of linear objects, the occurrence of faulty segmentations and a description of the overall segmentation quality were in the focus of this first step.
Furthermore, a detailed comparison based on visual delineated and clearly definable reference areas was carried out. Therefore 20 different areas (varying in location, form, area, texture, contrast, land cover type etc.) were selected and each was visually and geometrically compared with the segmented pendants. The geometrical comparison is a combination of morphological features (area $A_i$, perimeter $P_i$, and Shape Index $SI_i$)

$$SI_i = \frac{P_i}{4\sqrt{A_i}} \tag{1}$$

of the region $i$ and the number of segments or partial segments in the case of over-segmentation. The Shape Index comes from landscape ecology and addresses the polygon form. For all features the variances to the reference values were calculated. As partial segments all polygons with at least 50 % area in the reference object were counted. Additionally the quality of segmentation was visually rated (0 poor, 1 medium, 2 good). A good segmentation quality is reached, when the overall differences of all criteria between the segmentation results and the associated reference objects are as low as possible.

## 4. RESULTS AND DISCUSSION

### 4.1 Visual Quality Assessment

**HalconSeg**: The proposed adapted segmentation algorithm for the image processing software HALCON offers satisfying results. As observed with all approaches the result is highly depending on the parameter settings. A significant influence to the segmentation quality could be seen for the parameterisations of the applied edge detection filter and the morphological operators. Linear structures are represented very well in the segmentation result.
**Imagine WS:** The ERDAS Imagine add-on Imagine WS is based on a watershed algorithm. Due to a sophisticated region-merging algorithm, over-segmentation is reduced, but still apparent.

| Evaluation Approach / Reference | Method Type[1] | Equation | Description |
|---|---|---|---|
| Fragmentation *(FRAG)* Strasters and Gerbrands (1991) | ED | $FRAG = \frac{1}{1 + p \cdot \lvert T_N - A_N \rvert^q}$ where $T_N$ is the number of objects in the image and $A_N$ the number of regions in the reference; $p$ and $q$ are scaling parameters | addresses over-/under-segmentation by analysing the number of segmented and reference regions |
| Area-Fit-Index *(AFI)* Lucieer (2004) | ED | $AFI = \frac{A_{\text{reference object}} - A_{\text{largest segment}}}{A_{\text{reference object}}}$ | |
| Geometric features Circularity Yang et al. (1995) | ED | $Circularity = \frac{4\pi A}{P}$ where $A$ is the area and $P$ is the perimeter | addresses the shape conformity between segmentation and reference regions (scaling invariant shape feature) |
| Geometric features Shape Index Neubert and Meinel (2003) | | $ShapeIndex = \frac{P}{4\sqrt{A}}$ where $A$ is the area and $P$ is the perimeter | |
| Empirical Evaluation Function Borsotti et al. (1998) | EG | $Q(I) = \frac{1}{1000(N \cdot M)} \sqrt{R} \sum_{i=1}^{R} \left[ \frac{e_i^2}{1 + \log A_i} + \left( \frac{R(A_i)}{A_i} \right)^2 \right]$ where $N{\cdot}M$ is the size of the image $I$, $e_i$ is the colour error of the region $i$ and $R(A)$ the number of regions of the size $A$ | addresses the uniformity feature within segmented regions (colour deviation) |
| Entropy-based evaluation function and a weighted disorder function Zhang et al. (2004) | EG | $E = H_l(I) + H_r(I)$ where $H_l$ is the layout entropy and $H_r$ is the expected region entropy of the image $I$ | addresses the uniformity within segmented regions (luminosity) using the entropy as a criterion of disorder within a region |
| Fitness function Everingham et al. (2002) | A, ED | Probabilistic hull, Potential accuracy $f(a,I)$ Multidimensional fitness-cost-space | addresses multiple criteria and parameterizations of algorithms by a probabilistic Fitness/Cost Analysis |

[1] according to classification proposed in Zhang (1996): Analytical (A), Empirical Goodness, unsupervised (EG), Empirical Discrepancy, supervised (ED)

Table 2. Approaches to quantitative evaluation of segmentation results.

**PARBAT**: The segmentations generated by the Parbat region growing algorithm provide good contour representations, but also a lot of very small scattered segments. Another drawback is the maximum processable scene size of 1.300 x 1.300 pixels. The large image support has been announced for updated versions.

**RHSEG**: The RHSEG software produces as a result of different hierarchy and resolution levels a set of segmentations. Additionally it allows the most extensive parameter settings of all programs. The results show both over- and under-segmentation within the same segmentation. Well-contrasted boundaries between main land cover classes were correctly represented. Areas of low contrast were often not reproduced properly. Previously presented results (Tilton, 2003, NASA, 2005) are related to low resolution LANDSAT-TM-data. The IKONOS imagery seems to be too complex for the algorithm. Due to the multitude of parameter settings there is still a need for optimization regarding the usability.

**SEGEN:** The segmentation software SEGEN also shows good results. Within objects it tends to a general over-segmentation due to parallel multi-contours. The software has a very good processing performance – it takes only 35 seconds to segment a 2.000 x 2.000 pixel scene (Gofman, 2006).

**SegSAR:** The predominantly for radar data developed programme SegSAR produces segmentations of very good quality. Like eCognition, it shows very uniform regions. Despite a moderate over-segmentation, objects are reproduced properly by the segment borders. Especially linear objects are reproduced properly. This is caused by the combination of different segmentation techniques and the focus on radar data, which are immanently noisy.

### 4.2 Comparison Based on Reference Areas

Additionally 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 3. SegSAR, HalconSEG and Imagine WS are reaching the best average region conformity (lowest Shape Index differences), whereas SEGEN shows the best visual result. The low perimeter deviations of HalconSEG are the result of smoothed segment boundaries due to the application of a morphological image processing. PARBAT and SEGSAR reproduce comparable results, but show higher area and perimeter deviations. It can be seen that in some cases there is a certain discrepancy between the objective quantitative measures and the subjective visual rating.

For better comparability table 3 shows – additionally to the recently tested algorithms – the minimum and maximum values of each indicator of all tested programs. It shows that none of the presently tested programmes quantitatively performs better than the previously tested ones. Despite of some good results of SegSAR, SEGEN and Imagine WS, the low values for partial segments (over-segmentation) produced by eCognition 3.0 as well as the low area and perimeter deviations of SPRING 4.0 and the *Minimum Entropy Approach* could not be reached.

| Segmentation program | Halcon-SEG | Imagine WS | PARBAT 0.32 | RHSEG 1.0 | SEGEN | SEGSAR 1.0 | 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 [%] | 27.5 | 19.5 | 19.6 | 52.4 | 14.4 | 15.9 | 8.2 | 2,100.3 |
| Average difference of perimeter [%] | 11.6 | 9.8 | 56.0 | 98.5 | 37.0 | 17.4 | 10.0 | 475.6 |
| Average difference of Shape Index [%] | 12.7 | 9.3 | 47.6 | 83.7 | 31.6 | 15.9 | 10.0 | 87.1 |
| Average number of partial segments | 5.7 | 4.8 | 23.6 | 37.0 | 2.8 | 4.6 | 1.8 | 134.6 |
| Average quality, visual evaluated [0…2][3] | 0.8 | 0.8 | 0.6 | 0.6 | 1.1 | 1.0 | 0.0 | 1.0 |

[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 3. Cumulated results of all 20 reference areas.

## 5. CONCLUSIONS

This paper has presented an overview and some theoretical background on segmentation techniques and their quality assessment. It has been shown that there is no established standard evaluation method. However, there exist various ad hoc approaches. For this study an empirical discrepancy method was used to compare recently available segmentation programs. Due to the diversity of implemented algorithms the segmentation results are varying remarkably. The appropriateness of each programme is still highly depending on the specific segmentation task. Beside a suboptimal segmentation quality some of the implemented algorithms still face technical issues such as a lack of process stability and robust import routines concerning image size and format, radiometric resolution, data structure and projection parameters. For this reason, almost all algorithms are still under development. Another optimization aspect is the minimization of segmentation parameters. As it could have been noticed, the algorithms are very sensitive to slightly differing parameterisations, their number should be diminished in order to reduce the time needed to obtain an optimal segmentation (mostly *trial-and-error*). Implemented evaluation methods (like proposed in HalconSEG) could iteratively support the user reaching the optimal settings. For the extraction of geo-information from segmentation results, integrated segment-based classification methods are desirable. Referring to this PARBAT and RHSEG already offer simple resources.

Despite suboptimal results, segmentation offers an important approach to semi-automated image analysis. Particularly in combination with presented evaluation methods and existing GIS-data image segmentation algorithms already are indispensable resources to retrieve geo-information from remote sensing imagery.

In combination with the previous study in total 13 segmentation programmes have been evaluated. Is has been shown, that there is more than one interesting approach in this dynamic field of research. The evaluation will be continued, e.g. using the new algorithms of Definiens 5.0 (former eCognition), InfoPACK 2.0, EDISON and MATLAB implementations. Furthermore, it is planned to extend the quality assessment procedure itself by some of the presented evaluation methods. The evaluation is still open for further algorithms.

## REFERENCES


Borsotti, M.; Campadelli, P.; Schettini, R. (1998): Quantitative evaluation of color image segmentation results. Pattern Recognition Letters, 19(8), pp. 741–747.

Carleer, A. P.; Debeir, O.; Wolff, E. (2005): Assessment of very high spatial resolution satellite image segmentations, PE&RS 71(11), pp. 1285-1294.

Cavallaro, A.; Gelasca, E. D.; Ebrahimi, T. (2002): Objective Evaluation of Segmentation Quality Using Spatio-temporal Context. In: Proc. IEEE International Conference on Image Processing, New York, 4 p. [http://www.elec.qmul.ac.uk/staffinfo/andrea/papers/cavallaro_icip02.pdf - 21.06.2006].

Estrada, F.J.; Jepson, A.D. (2005): Quantitative evaluation of a novel image segmentation algorithm. Computer Vision and Pattern Recognition 2, pp. 1132-1139.

Everingham, M.; Muller, H.; Thomas, B. (2002): Evaluating image segmentation algorithms using monotonic hulls in fitness/cost space. In: Cootes, T., Taylor, C. (Eds.): Proc. 12th British Machine Vision Conference, pp. 363-372.

Gofman, E. (2006): Developing an Efficient Region Growing Engine for Image Segmentation. Proc. 13th Int. Conf. on Image Processing (ICIP), Atlanta, GA, USA, 4 p (in press).

Gonzales, R. C.; Woods, R. E. (1993): Digital Image Processing. Addison-Wesley, Reading, 716 p.

Haberäcker, P. (1995): Praxis der digitalen Bildverarbeitung und Mustererkennung. Hanser Verlag, München, 350 p.

Haralick, R. M.; Shapiro, L. G. (1992): Computer and robot vision. Vol. I, Addison-Weasley, Reading, 672 p.

Herold, H. (2005): Beiträge zum Vergleich von Segmentierungsprogrammen fernerkundlicher Bilddaten, Master Thesis, Inst. f. Geographie, TU Dresden, 125 p.

Hirschmugl, M. (2002): Automatisierte Bildsegmentierung von Nutzungsstrukturen in der Fernerkundung. Master Thesis, Inst. f. Geographie u. Raumforschung, Universität Graz, Graz: 115 p.

Karantzalos, K. G.; Argialas, D. P. (2003): Evaluation of Selected Edge Detection Techniques in Remote Sensing. In: Serpico, S. B. (Ed.): Image and Signal Processing for Remote Sensing VIII, Proc. SPIE 4885, pp. 102-110.

Lanser, S. (1993): Kantenbasierte Farbsegmentation im CIE-Lab Raum. In: Pöppl, H. (Ed.): Mustererkennung, Informatik aktuell, Springer-Verlag, pp. 639-646.

Letournel, V.; Sankur, B.; Pradeilles, F; Maître, H. (2002): Feature Extraction for Quality Assessment of Aerial Image Segmentation. In: ISPRS Technical Commission III Symposium, Graz, 6 p. [http://www.isprs.org/commission3/proceedings02/papers/paper105.pdf - 21.06.2006].

Levine, M; Nazif, A. (1985): Dynamic Measurement of Computer Generated Image Segmentations. IEEE Transaction on Pattern Analysis and Machine Intelligence, 7(2), pp. 155-164.

Liu W.; Zhai J.; Dori D.; Tang L. (2001): A System for Performance Evaluation of Arc Segmentation Algorithms. In: Proc. CVPR Workshop on Empirical Evaluation in Computer Vision, 11 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 - 21.06.2006].

Meinel, G.; Neubert, M. (2004): A Comparision of segmentation programs for high resolution remote sensing data. Int. Arch. of Photogrammetry, Remote Sensing and Spatial Information Sciences XXXV-B4, pp. 1097-1102.

Meinel, G.; Neubert, M.; Reder, J. (2001): Pixelorientierte versus segmentorientierte Klassifikation von IKONOS-Satellitenbilddaten – ein Methodenvergleich. In: PFG (3), pp. 157-170.

Mezaris V.; Kompatsiaris, I.; Strintzis, M.G. (2004): Still Image Segmentation Tools for Object-based Multimedia Applications. Int. J. Pattern Recogn. 18(4), pp. 701-725.

NASA (2005): Recursive Hierarchical Segmentation (RHSEG) Pre-processing Software. [http://techtransfer.gsfc.nasa.gov/rhseg - 21.06.2006].

Neubert, M., Meinel, G. (2003): Evaluation of segmentation programs for high resolution remote sensing applications. In: Schroeder, M., Jacobsen, K., Heipke, C. (Eds.): Proc. Joint ISPRS/EARSeL Workshop "High Resolution Mapping from Space 2003", CD-ROM, 8 p.

Pal, N. R.; Pal, S. K. (1993): A Review on Image Segmentation Techniques. Pattern Recognition 26(9), pp. 1277-1294.

Palus, H.; Kotyczka, T. (2001): Evaluation of Colour Image Segmentation Results. In: Proc. 7. Workshop Farbbildverarbeitung 2001, 4 p. [http://kb-bmts.rz.tu-ilmenau.de/gcg/html/Vortr_01_pdf/Palus_Bild.pdf - 21.06.2006].

Schouten, T. E.; Klein Gebbinck, M. S. (1995): Quality Measures for Image Segmentation Using Generated Images. Proc. Image and Signal Processing for Remote Sensing II, pp. 411-422.

Schukraft, A.; Lenz, R. (2003): Ermittlung der Lagegenauigkeit von Vektordaten mit Hilfe der fraktalen Boxdimension. In: GeoBIT/GIS (3), pp. 15-19.

Sramek, M.; Wrbka, T. (1997): Watershed based image segmentation - an effective tool for detecting landscape structure. In: Digital Image Processing and Computer Graphics (DIP'97), Proc. SPIE 3346, pp. 227-235.

Strasters, K.; Gerbrands, J. (1991). Three-dimensional segmentation using a split, merge and group approach. Pattern Recognition Letters 12, pp. 307-325.

Tilton, J. C. (2003): Analysis of hierarchically related image segmentations. In: Proc. of IEEE Workshop on Advances in Techniques for Analysis of Remotely Sensed Data, pp. 60-69.

Villegas, P.; Marichal, X.; Salcedo, A. (1999): Objective Evaluation of Segmentation Masks in Video Sequences. In: Proc. Workshop on Image Analysis for Multimedia Interactive Services, 4 p.

Yang, L.; Albregtsen, F.; Lønnestad, T.; Grøttum, P. (1995): A Supervised Approach to the Evaluation of Image Segmentation Methods. In: Proc. CAIP, Lect. Notes Comput. Sc. 970, pp. 759-765.

Yasnoff W. A.; Mui, J. K.; Bacus J. W. (1977): Error measures for scene segmentation. Pattern Recogn. 9(4), pp. 217-231.

Zhang, H.; Fritts, J. A.; Goldman, S. A. (2004): An entropy-based objective segmentation evaluation method for image segmentation. SPIE Electronic Imaging - Storage and Retrieval Methods and Applications for Multimedia, pp. 38-49.

Zhang, H.; Fritts, J. A.; Goldman, S. A. (2005): A Co-Evaluation Framework for Improving Segmentation Evaluation. SPIE Signal Processing and Target Recognition XIV, pp. 420-430.

Zhang, Y. J. (1996): A survey on evaluation methods for image segmentation. Pattern Recognition 29(8), pp. 1335-1346.

Zhang, Y. J.; Gerbrands, J. J. (1994): Objective and quantitative segmentation evaluation and comparison. Signal Processing 39, pp. 43-54.

## ACKNOWLEDGMENTS

The authors thank Mr. Dr. C. Steger *(Munich University of Technology/MVTec GmbH, Germany)*, Mr. Dr. A. Lucieer *(Centre for Spatial Information Science, University of Tasmania, Australia)*, Mr. Dr. M. de Araújo Sousa Jr. *(Instituto National de Pesquisas Espacias, Brazil)*, Mr. Dr. J. C. Tilton *(Goddard Space Flight Center, National Aeronautics and Space Administration, USA)*, Mr. Dr. E. Gofman *(IBM Haifa Research Labs, Israel)*, Mr. Dr. H. Beissmann and Mr. Dr. M. Šrámek *(Austrian Academy of Sciences, Commission for Scientific Visualization, Austria)* for the collaboration, the technical support for the segmentation algorithms and the processing of sample data, respectively.