

Improvement of Image Segmentation Accuracy Based on Multiscale Optimization Procedure

T. Esch, M. Thiel, M. Bock, A. Roth, and S. Dech, *Member, IEEE*

Abstract—This letter proposes an optimization approach that enhances the quality of image segmentation using the software *Definiens Developer*. The procedure aims at the minimization of over- and undersegmentations in order to attain more accurate segmentation results. The optimization iteratively combines a sequence of multiscale segmentation, feature-based classification, and classification-based object refinement. The developed method has been applied to various remotely sensed data and is compared to the results achieved with the established segmentation procedures provided by the *Definiens Developer* software. The quantitative assessment of segmentation accuracy based on reference objects is derived from an aerial image, and a high-resolution synthetic aperture radar scene shows an improvement of 20%–40% in object accuracy by applying the proposed procedure.

Index Terms—High-resolution data, image segmentation, level fusion, object refinement, object-oriented image analysis.

I. INTRODUCTION

OBJECT-ORIENTED image analysis has constantly gained importance in Earth observation during the last few years [1]–[4]. A comparison of recent object-oriented image analysis algorithms applied in remote sensing is given in [5]–[9]. The most promising and commonly used segment-based approach is the Fractal Net Evolution concept of the *Definiens Developer* software [10], [11]. Here, the segmentation is controlled by the heterogeneity criteria color h_c (1) and shape h_s , with the shape being composed of the smoothness h_{ss} (2) and the compactness h_{sc} (3)

$$h_c = \sum_b w_b * \sigma_b \quad (1)$$

with w_c defining the weight and σ_c the standard deviation of band b , and

$$h_{ss} = \frac{l}{k} \quad (2)$$

Manuscript received December 17, 2007; revised February 5, 2008.

T. Esch was with the German Remote Sensing Data Center, German Aerospace Center (DLR), 82234 Oberpfaffenhofen, Germany. He is now with the DLR Department of Remote Sensing, Institute for Geography, University of Wuerzburg, 97074 Wuerzburg, Germany (e-mail: thomas.esch@uni-wuerzburg.de).

M. Thiel is with the DLR Department of Remote Sensing, Institute for Geography, University of Wuerzburg, 97074 Wuerzburg, Germany (e-mail: michael.thiel@uni-wuerzburg.de).

M. Bock, A. Roth, and S. Dech are with the German Remote Sensing Data Center, German Aerospace Center, 82234 Oberpfaffenhofen, Germany (e-mail: Michael.Bock@dlr.de; Achim.Roth@dlr.de; Stefan.Dech@dlr.de).

Color versions of one or more of the figures in this paper are available online at <http://ieeexplore.ieee.org>.

Digital Object Identifier 10.1109/LGRS.2008.919622

with l giving the actual length of the object's outline and k providing the shortest length of the bounding box, and

$$h_{sc} = \frac{l}{\sqrt{n}} \quad (3)$$

with l representing the length of the object outline and n giving the number of pixels of the object. After each merge process, the change of heterogeneity—which flows into the fusion value S_f —is calculated by

$$S_f = w_s h_c + (1 - w_s) h_s \quad (4)$$

with w_s presenting a user-defined weighting factor of the shape criterion. The fusion value is compared with a user-defined scale parameter which defines the maximum allowable heterogeneity of the image objects. By varying the scale parameter, arbitrary object levels with scale-specific segment sizes can be generated. As an alternative to multiresolution segmentation, chessboard, quadtree-based, and spectral difference segmentations are provided [12]. The chessboard algorithm results in a regular grid of segments with a predefined size. The quadtree segmentation splits an image domain into squares, whereas these squares are subdivided into smaller ones until the spectral heterogeneity of every element falls below a user-defined scale factor. This factor defines the maximum color difference inside the square image objects. The spectral difference segmentation merges adjacent objects that do not exceed a user-defined maximum threshold for a weighted spectral difference.

Achieving appropriate segmentation with *Definiens Developer* exhibits the difficulty to determine the optimal parameterization for the segmentation [13], [14]. Maxwell and Zhang [15] target this handicap by a fuzzy logic approach that identifies suitable object segmentation parameters. Although the proposed method achieved promising results, this technique still features a significant degree of complexity, which requires high technical skills and expertise of the user. Recently, Moeller *et al.* [16] proposed a validation algorithm that localizes and quantifies segmentation inaccuracies. It facilitates a relative comparison of segmentation results. Nevertheless, this approach requires the generation of an extensive basis of segmentations which have to be compared *ex post*.

This letter discusses a method that directly adds to the optimization of image segmentation in *Definiens Developer* by reducing under- and oversegmentations. The basic concept of the approach is introduced in Section II. The third section discusses the results achieved on the basis of both optical and SAR images. Finally, the conclusions are drawn.

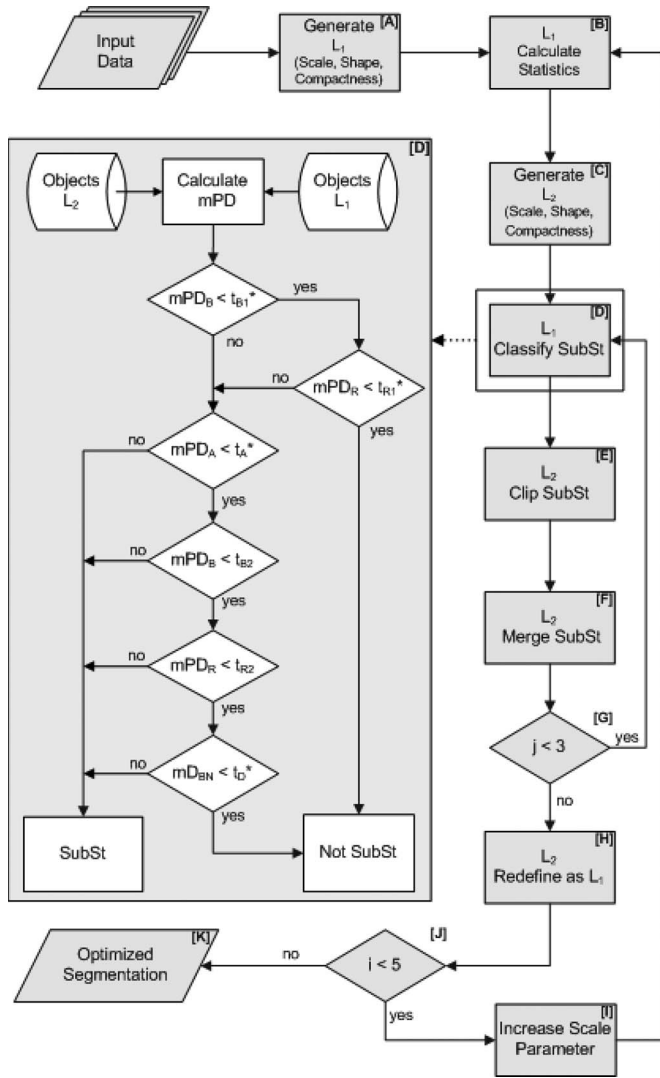


Fig. 1. Flowchart of SOP implemented in environment of Definiens Developer. User-defined settings are asterisked.

II. OPTIMIZATION PROCEDURE

Facing the difficulties associated with image segmentation described in [13] and [14], the proposed technique focuses on increased segmentation accuracy by minimizing over- and undersegmentations. Thus, we developed a segmentation optimization procedure (SOP) that combines the functionality of hierarchical segmentation on multiple scales with a classification-based image object refinement using clipping or merging of segments. Thereby, the user initially specifies a minimum tolerable difference between distinct objects by a maximum of four parameters. The SOP will then generate a segmentation that optimally meets the defined requirements.

The SOP starts with the generation of a basic segmentation (L_1) by means of the original multiresolution segmentation algorithm (Fig. 1, [A]). The settings for the scale parameter, color/shape, and compactness/smoothness are defined as constants so that they need not be set by the user anymore. The segments of L_1 are suited to represent the smallest object types to be classified in a later image analysis phase—e.g., houses or trees. Based on the generated segments, statistical parameters,

such as global minimum and maximum of brightness, are calculated [B]. They are required for the automated determination of optimal thresholds for some of the parameters used in the following optimization process. Next, a second level (L_2) with increased object sizes is segmented atop of L_1 —again by means of the original multiresolution algorithm and predefined settings for the segmentation parameters [C]. Since the objects at L_2 must consider the boundaries of the L_1 segments, the objects of L_2 result from a fusion of the objects provided by L_1 [10], [11].

Next, the segments of L_1 —termed subobjects—are assessed with the classification functionality of Definiens Developer whether they significantly differ from the spectral properties of their superordinate segments—termed superobjects—located at L_2 [D]. The quantification of the spectral similarity between sub- and superobjects is described by the mean percentage difference (mPD) (5) in their weighted brightness $B(\text{mPD}_B)$ and their spectral signature $R(\text{mPD}_R)$

$$\text{mPD} = \frac{|v_{L_1} - v_{L_2}|}{v_{L_2}} \quad (5)$$

with v calculated at L_1 and L_2 for the weighted brightness B (6) and the ratio of reflectivity R between the single bands (7)

$$B = \frac{1}{MJ} \sum_{m=1}^M \sum_{j=1}^J w_m p_{m,j} \quad (6)$$

with B representing the weighted brightness, M providing the number of channels, $p_{m,j}$ giving the gray value of the pixels, J providing the number of pixels per object, and w_m defining a weighting factor for the bands. The ratio R between the reflectivity of the adjacent spectral bands is calculated by

$$R = \frac{p_i}{p_{i+1}} \quad (7)$$

with p_i giving the gray value in channel i . This ratio is generated for all pairs of adjacent channels. Thus, when I is the number of image bands, $I - 1$ ratios are compared in order to characterize the difference in the spectral signature of sub- and superobjects. As soon as mPD_B or mPD_R exceed a threshold (t_{B1}, t_{R1}), the corresponding segment at L_1 is classified as a distinct substructure (SubSt) [D].

Optionally, the user can define specific conditions for small objects in order to regulate the tolerable amount of local detail such as noise effects or cars on streets. Thereby, segments whose mPD to the superobject in their area (mPD_A) is smaller than threshold t_A have to fulfill additional conditions [D]. They will only be classified as SubSt if mPD_B or mPD_R are larger than more rigorous thresholds t_{B2} or t_{R2} , respectively, or if the mean absolute difference in brightness to their neighboring objects (mD_{BN}) (8) exceeds threshold t_D .

$$\text{mD}_{BN} = \frac{1}{N} \sum_{n=1}^N |B_n - B_E| \quad (8)$$

with N providing the number of neighbors of object E . B is the weighted brightness of the segments as given in (6).

TABLE I
QUANTITATIVE COMPARISON OF SEGMENTATION ACCURACY ACHIEVED WITH SEGMENTATION ALGORITHMS PROVIDED BY DEFINIENS DEVELOPER (Q-SD, M-SD) AND THE OPTIMIZED SEGMENTATION PROCEDURE (SOP)

| Feature | Q-SD ^a | M-SD ^b | SOP ^c |
|--|-------------------|-------------------|------------------|
| Aerial image | | | |
| Number of reference objects | 70 | 70 | 70 |
| Correctly segmented objects [%] ^d | 40 | 44 | 63 |
| Incorrectly segmented objects | | | |
| Subdivided [%] (no. of sub-objects) | 88 (5) | 54 (3) | 77 (4) |
| Incorrectly segmented objects | | | |
| Merged [%] (Ø StndDev [%]) | 12 (8) | 46 (13) | 23 (10) |
| Deviation in area [%] ^e | 15 | 11 | 7 |
| Deviation in perimeter [%] ^e | 12 | 7 | 6 |
| Deviation in shape index [%] ^e | 22 | 7 | 10 |
| Deviation in band 1 [%] ^e | 3 | 2 | 1 |
| Deviation in band 2 [%] ^e | 2 | 2 | 1 |
| Deviation in band 3 [%] ^e | 2 | 1 | 1 |
| E-SAR image | | | |
| Number of reference objects | 60 | 60 | 60 |
| Correctly segmented objects [%] ^d | 42 | 53 | 85 |
| Incorrectly segmented objects | | | |
| Subdivided [%] (no. of sub-objects) | 60 (6) | 50 (4) | 67 (4) |
| Incorrectly segmented objects | | | |
| Merged [%] (Ø StndDev [%]) | 40 (9) | 50 (15) | 33 (15) |
| Deviation in area [%] ^e | 5 | 4 | 8 |
| Deviation in perimeter [%] ^e | 3 | 6 | 8 |
| Deviation in shape index [%] ^e | 3 | 4 | 6 |
| Deviation in band 1 [%] ^e | 2 | 1 | 2 |

^a Settings: Aerial: Q = 18, SD = 11; E-SAR: Q = 20, SD = 15

^b Settings: Aerial: M: scale = 16, shape = 0.2, comp = 0.7, SD = 8; E-SAR: M: scale = 25, shape = 0.2, comp = 0.7, SD = 8

^c Settings: Aerial: $t_{B1} = 0.7$, $t_{R1} = 0.047$, $t_A = 0$, scale = 10, 20, 50, 80, 110, 180 shape = 0.3, comp = 0.8; E-SAR: $t_{B1} = 0.44$, $t_A = 0.01$, $t_{B2} = 0.925$, $t_D = 35$, scale = 10, 20, 50, 80, 110, 180 shape = 0.3, comp = 0.8

^d A segment was assigned as correctly segmented if its area and perimeter showed a maximum deviation of 20% compared to its reference object

^e Calculated on the basis of the correctly represented image objects (Deviation = $|(V_{ref} - V_{seg})/V_{ref}|$) with V_{ref} providing the considered variable of the reference objects and V_{seg} giving the according value of the automatically segmented objects

Subsequent to the described classification process, all segments identified as SubSt at L_1 are clipped from the superobjects at L_2 [E]. By doing so, the shape of the SubSt is transferred to L_2 . In addition, all adjacent SubSt with a similar weighted brightness (6) are merged at L_2 [F]. The similarity (sim) of two adjacent objects ob1 and ob2 in terms of their weighted brightness is calculated as

$$\text{sim}_{ob2,ob1} = \begin{cases} 1, & 0.9B_{ob1} \leq B_{ob2} \leq 1.1B_{ob1} \\ 0, & \text{else.} \end{cases} \quad (9)$$

The processes of classifying, clipping, and merging those objects identified as SubSt run in a loop (j) with three cycles [G]. This is due to the changing statistics of the superobjects, when clipping and merging one or more substructures. After three iterations, the first loop of the SOP ends with the deletion of L_1 and the redefinition of former L_2 as new L_1 [H].

The second iteration of the SOP starts with the generation of a new L_2 , whereas the scale is increased again [I]. All subsequent steps are identical to the first iteration. The optimization continues while constantly increasing the size of segments for L_2 until the generated objects attain a spatial extent that adequately represents the largest objects to be analyzed—e.g., agricultural fields or water bodies. The range of scale parameters from the smallest to the largest objects is realized within five iterations (i) [J].

The final output of the SOP provides a single segmentation which results from a fusion of differently scaled segmentation levels [K]. Large segments of homogeneous nature are extracted from coarser scale levels, whereas small-scale structures such as trees and buildings are preserved from the finer segmentations. Hence, the algorithm minimizes over- and undersegmentations that are characteristic for the standard segmentation modes of Definiens Developer and other segmentation approaches [5]. Thereby, each step of the SOP is stored

and performed in a so-called process tree within Definiens Developer. This is particularly suitable for a combined segmentation of various data—e.g., synthetic aperture radar (SAR) and optical imagery. For that purpose, two or more optimization procedures are run in series, whereas each input layer is optimized individually. The resulting segmentation for the first data set directly serves as the optimization level in the subsequent SOP of the second image data set. By doing so, the spectral and spatial structures of all input layers are finally represented in a single object level.

Those parameters of the SOP that can be defined by the user are provided in the form of a graphical user interface (GUI) in the Definiens Architect. This includes four parameters in the case of multiband data (t_{B1} , t_{R1} , t_A , and t_D) and three parameters for single-band imagery (t_{B1} , t_A , and t_D). If there is no requirement to define a minimum area, the number of parameters even decreases to two (t_{B1} and t_{R1}) and one (t_B), respectively. The variables t_{B2} and t_{R2} are determined automatically by calculating the 70% quantile of t_{B1} and t_{R1} based on all objects fulfilling the conditions defined by t_{B1} , t_{R2} , and t_A .

III. RESULTS AND DISCUSSION

In order to analyze the characteristics and functionality of the SOP, we quantitatively assessed the performance on the basis of an aerial image and a high-resolution SAR scene by comparing the outcomes to the results achieved with the standard algorithms provided by Definiens Developer. The aerial image represents 8-bit RGB data with a spatial resolution of 50 cm, whereas the single-polarized 16-bit E-SAR scene features a spatial resolution of 1.5 m. For the assessment, 70 and 60 reference objects were manually digitized based on the aerial data and the E-SAR image, respectively. The spectral and shape-related properties of the reference objects were then compared to the characteristics of the automatically generated segments (Table I). The reference objects represented spatial entities such as houses, trees, shadows, streets, lawns, or open area. The best results related to the algorithms provided by Definiens Developer were obtained by applying a combination of quadtree and spectral difference segmentation (Q-SD) and by a combination of multiresolution and spectral difference segmentation (M-SD).

Regarding the segmentation of the aerial photograph, 63% of the reference objects are represented correctly when applying the SOP. Q-SD and M-SD achieve accuracies of 40% and 44%. Looking at the incorrectly represented objects, SOP and Q-SD tend to oversegmentation—77% or 88% of the reference objects are subdivided into an average of four or five segments, respectively. Concerning the spectral and shape-related properties, the SOP exhibits the least discrepancies to the characteristics of the reference objects. Here, the average variation reaches up to 4%. With 85% of correctly delineated objects in terms of the E-SAR data, the SOP exceeds the accuracy of the established procedures by 32% (M-SD) and 43% (Q-SD). SOP and Q-SD again lean toward oversegmentation. The spectral and shape-related properties of the generated objects which are correctly represented differ from the reference ob-

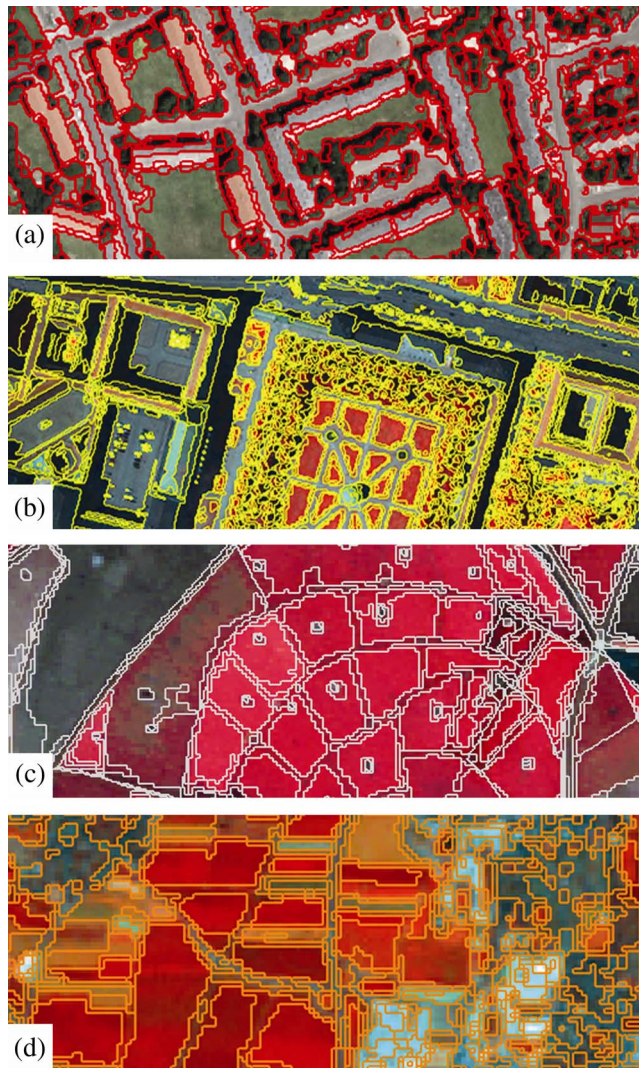


Fig. 2. Result of SOP segmentation based on (a) aerial image (50 cm, 8-bit), (b) Ikonos data (1-m resolution merge, 11-bit), (c) SPOT-5 image (2.5-m resolution merge, 8-bit), and (d) Landsat-7 data (25 m, 8-bit).

jects by an average of 3% for Q-SD, 4% for M-SD, and 6% for SOP.

In addition to the quantitative assessment, we visually evaluated the results achieved by applying the proposed approach to data of various sensors. Most convincing results were achieved with very high resolution optical imagery such as aerial data (8-bit, 50 cm) and Ikonos (11-bit, 1 m) or SPOT-5 (8-bit, 2.5 m) images [Fig. 2(a)–(c)]. The segmentation of IRS (8-bit, 5.8 m) or Landsat (8-bit, 30 m) [Fig. 2(d)] data appears sufficient to a certain extent. The latter can be attributed to the increasing amount of mixed pixels which involve a smooth transition of different land cover types—a fact that hampers their distinct delineation based on spectral properties as implemented in the SOP. Hence, transition zones between distinct land cover types are partly lined with smaller objects representing the area of mixed pixels. Due to the high variance of SAR data, the segmentation of radar images results in a significant degree of subdivisions in highly structured image regions such as built-up areas or woodlands (Fig. 3).

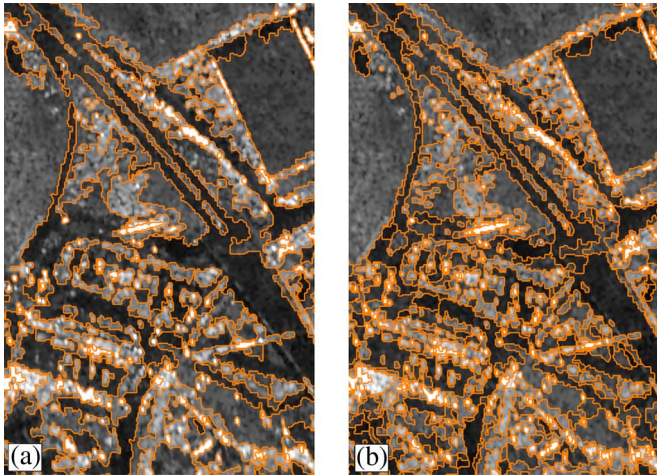


Fig. 3. Segmentation of high-resolution E-SAR image (1.5 m, 16-bit, HH) generated by (a) established M-SD algorithm and (b) proposed SOP approach.

Limitations in the usability of the proposed procedure arise when a superobject consists of a few equally sized subobjects. In this case, the spectral property of the corresponding superobject results from an almost equal averaging of the spectral characteristics of the subsegments. Thus, none of these subobjects significantly differs from the respective superordinate segment located at L_2 , and the superobject remains unmodified. Moreover, the SOP is solely based on spectral characteristics. Hence, heterogeneously composed regions such as built-up areas and woodlands or highly dynamic data sets, e.g., SAR imagery, are subdivided into their spectral and structural primitives. Based on these primitives, it is impossible to calculate—for instance—meaningful textual features. If such features are required for classification, these should be supplemented by a superior segmentation level generated with the multiresolution algorithm and featuring large-area objects.

IV. CONCLUSION

Quantitative assessment and visual inspection of the results show that the adaptive segmentation procedure facilitates a more accurate and robust image segmentation with Definiens Developer. This improvement is mainly due to the definition of the segmentation settings that apply for all spatial entities of the image—a fact that significantly reduces under- and oversegmentations. The complete procedure is implemented in a GUI, whereas the user has to define a maximum of four parameters—the mean percentage difference in brightness and spectral signature and, optionally for small objects, the mean percentage difference in the area and the spectral difference to neighboring segments. These attributes can directly be queried for each single segment by the object information provided in Definiens Developer. Thus, the complete process of object generation becomes more transparent and targeted. Moreover, there is no requirement anymore to determine the appropriate number of scale levels and their individual optimum parameterization in terms of scale parameter, the color/shape, and smoothness/compactness.

Nevertheless, due to the multiple segmentation and classification processes included in the proposed segmentation procedure,

the processing time is significantly increased by a factor of 20–50 compared to the standard methods provided by Definiens Developer. Nevertheless, the improvement in segmentation accuracy between 20% and 40% and the accelerated achievement of an accurate segmentation that fits all spatial entities of an image compensate this drawback.

ACKNOWLEDGMENT

The authors would like to thank the German Federal Ministry of Economics and Technology (BMWi) for their financial support in the research project “TerraSAR-X urban.”

REFERENCES

- [1] T. Blaschke, S. Lang, E. Lorup, J. Strobl, and P. Zeil, “Object-oriented image processing in an integrated GIS/remote sensing environment and perspectives for environmental applications,” in *Umweltinformation für Planung, Politik und Öffentlichkeit*, vol. 2, A. Cremers and K. Greve, Eds. Marburg, Germany: Metropolis Verlag, 2000, pp. 555–570.
- [2] G. Meinel, M. Neubert, and J. Reeder, “The potential use of very high resolution satellite data for urban areas—First experiences with IKONOS data, their classification and application in urban planning and environmental monitoring,” in *Proc. Regensburger Geographische Schriften 35: Remote Sens. Urban Areas*, C. Jürgens, Ed., 2001, pp. 196–205.
- [3] T. Blaschke and J. Strobl, “What’s wrong with pixels? Some recent developments interfacing remote sensing and GIS,” *GIS—Zeitschrift für Geoinformationssysteme*, no. 6, pp. 12–17, 2001.
- [4] U. Benz, P. Hofmann, G. Wilhauck, I. Lingenfelder, and M. Heynen, “Multi-resolution, object-oriented fuzzy analysis of remote sensing data for GIS-ready information,” *ISPRS J. Photogramm. Remote Sens.*, vol. 58, no. 3/4, pp. 239–258, 2004.
- [5] G. Meinel and M. Neubert, “A comparison of segmentation programs for high resolution remote sensing data,” *Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci.*, vol. XXXV, pt. B4, pp. 1097–1102, 2004.
- [6] J. Schiewe, “Segmentation of high-resolution remotely sensed data—Concepts, applications and problems,” *Int. Arch. Photogramm. Remote Sens.*, vol. XXXIV, no. 4, pp. 380–385, 2002.
- [7] M. Neubert, H. Herold, and G. Meinel, “Evaluation of remote sensing image segmentation quality—Further results and concepts,” *Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci.*, vol. XXXVI, no. 4/C42, pp. 6–11, 2006.
- [8] A. P. Carleer, O. Debeir, and E. Wolff, “Assessment of very high spatial resolution satellite image segmentations,” *Photogramm. Eng. Remote Sens.*, vol. 71, no. 11, pp. 1285–1294, 2005.
- [9] G. Hay, T. Blaschke, D. Marceau, and A. Bouchard, “A comparison of three image-object methods for the multiscale analysis of landscape structure,” *ISPRS J. Photogramm. Remote Sens.*, vol. 57, no. 5/6, pp. 327–345, 2003.
- [10] M. Baatz and A. Schäpe, “Multiresolution segmentation—An optimization approach for high quality multi-scale image segmentation,” in *Angew. Geographische Informationsverarbeitung XII*, J. Strobl *et al.*, Eds., Karlsruhe, Germany: Verlag Herbert Wichmann Verlag, 2000, pp. 12–23.
- [11] M. Baatz, A. Schäpe, G. Schmidt, M. Athelougou, and G. Binnig, “Cognition network technology: Object orientation and fractal topology in biomedical image analysis. Method and applications,” in *Fractals in Biology and Medicine*, vol. 4, pt. 1, G. A. Losa, D. Merlini, T. F. Nonnenmacher, and E. R. Weibel, Eds. Basel, Switzerland: Birkhäuser Verlag, 2005, pp. 67–73.
- [12] *Definiens Developer 7, Reference Book*, Definiens AG, Munich, Germany, 2007, pp. 15–24.
- [13] J. Schiewe, L. Tufte, and M. Ehlers, “Potential and problems of multi-scale segmentation methods in remote sensing,” *Geo-Informationssysteme*, vol. 6, pp. 34–39, 2001.
- [14] D. Flanders, M. Hall-Beyer, and J. Pereverzoff, “Preliminary evaluation of eCognition object-based software for cut block delineation and feature extraction,” *Can. J. Remote Sens.*, vol. 29, no. 4, pp. 441–452, 2003.
- [15] T. Maxwell and Y. Zhang, “A fuzzy approach to supervised segmentation parameter selection for object-based classification,” *Proc. SPIE*, vol. 5909, pp. 528–538, 2005.
- [16] M. Moeller, L. Lyburner, and M. Volk, “The comparison index: A tool for assessing the accuracy of image segmentation,” *Int. J. Appl. Earth Obs. Geoinf.*, vol. 9, no. 3, pp. 311–321, 2007.