ELSEVIER

International Journal of Applied Earth Observation and Geoinformation 9 (2007) 311–321

INTERNATIONAL JOURNAL OF APPLIED EARTH OBSERVATION AND GEOINFORMATION

www.elsevier.com/locate/jag

The comparison index: A tool for assessing the accuracy of image segmentation

M. Möller [a,*], L. Lymburner [b], M. Volk [c]

[a] *Geoflux Gbr, Paracelsusstraße 6, 06114 Halle (Saale), Germany*
[b] *Australian Center for Tropical Freshwater Research, James Cook University, Townsville, Qld 4811, Australia*
[c] *UFZ—Center for Environmental Research, Department of Applied Landscape Ecology, Permoserstr. 15, 04318 Leipzig, Germany*

Received 2 November 2005; accepted 20 October 2006

Abstract

Segmentation algorithms applied to remote sensing data provide valuable information about the size, distribution and context of landscape objects at a range of scales. However, there is a need for well-defined and robust validation tools to assessing the reliability of segmentation results. Such tools are required to assess whether image segments are based on 'real' objects, such as field boundaries, or on artefacts of the image segmentation algorithm. These tools can be used to improve the reliability of any land-use/land-cover classifications or landscape analyses that is based on the image segments.

The validation algorithm developed in this paper aims to: (a) localize and quantify segmentation inaccuracies; and (b) allow the assessment of segmentation results on the whole. The first aim is achieved using object metrics that enable the quantification of topological and geometric object differences. The second aim is achieved by combining these object metrics into a 'Comparison Index', which allows a relative comparison of different segmentation results. The approach demonstrates how the Comparison Index CI can be used to guide trial-and-error techniques, enabling the identification of a segmentation scale H that is close to optimal. Once this scale has been identified a more detailed examination of the CI–H- diagrams can be used to identify precisely what H value and associated parameter settings will yield the most accurate image segmentation results.

The procedure is applied to segmented Landsat scenes in an agricultural area in Saxony-Anhalt, Germany. The segmentations were generated using the 'Fractal Net Evolution Approach', which is implemented in the eCognition software.

Keywords: Segmentation; Landsat; Field detection; Validation; Accuracy; Object metric

1. Introduction

Image segmentation algorithms such as those contained within eCognition™ are increasingly popular for a wide range of image processing tasks, and the advantages of working with image segments, rather than individual pixels are widely recognized (Fortin et al., 2000; Shi et al., 2005). However, there is a wide range of variables to manipulate, whereas segmenting an image and identifying an 'optimal' result can be difficult. The tools developed in this paper aim to make this process more objective and rigorous.

In this study, image segmentation algorithms were applied to Landsat TM data for an agricultural region in Saxony-Anhalt, Germany. The accuracy of the field boundary delineation was highly important because of its impact on the accuracy of the sediment/pollutant transport model that uses the field boundaries as model input (Van Oost et al., 2000; Takken et al., 2001). In addition, measures of soil protection legislation like

* Corresponding author. Fax: +49 341 2353939.
E-mail address: moeller@geoflux.de (M. Möller).

0303-2434/$ – see front matter
doi:10.1016/j.jag.2006.10.002

'European Common Agricultural Policy' or 'Good Farming Practice' (EU, 2002) refer to up-to-date field shapes. However, the existing public vector data sets provided by land surveys are not able to define field boundaries accurately. This is mainly due to the large temporal variations of crop structures (Mysiak et al., 2004). Satellite sensor data have the capacity to fill this 'temporal gap' thereby providing up-to-date field boundary information.

To assess the accuracy of the field boundary delineation we focus on the geometric quality of a single class (cp. Zhan et al., 2005) rather than classification accuracy (sensu Foody, 2002). The image segmentation was carried out using the 'fractal net evolution approach' (FNEA) (Benz et al., 2004). FNEA is constrained using a set of user-defined parameter settings, which affect the segmentation results. In this paper, we investigate the influence of segmentation parameter settings on segmentation results with a view to identifying the parameter settings that provide optimal segmentation results for a specific target. The main objectives of this study are:

(1) the development of a user-friendly evaluation procedure to visualize and quantify segmentation inaccuracies. Inaccuracies refer to over-segmentations or under-segmentations, which stand for generating too many or too few segments (Delves et al., 1992). This level is referred to as local validation, because single objects are considered;
(2) the assessment of segmentation results on the whole, by which we want to achieve an optimal parameter setting of the applied segmentation method. This level is referred to as global validation, because the entire image is considered.

These objectives were achieved by visualizing two object metrics (local validation) and using the Site Comparison Method SICOM (Deumlich et al., 2006) which aggregates the classified object metrics to a map complexity metric Comparison Index CI (global validation). The extrapolation of the complexity metric to the whole study area was realized using random sampling methodology (Congalton and Green, 1990; Stehmann, 1992).

2. Methods

2.1. Image segmentation

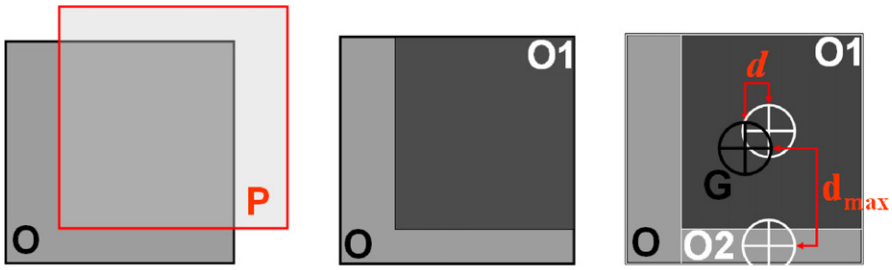
There are various methods for automatic field detection that are based on the application of segmentation algorithms to remote sensing data (e.g., Fuller et al., 2002; Evans et al., 2002; Betenuth, 2004; Mueller et al., 2004; Devereux et al., 2004). Image segmentation is a spatial clustering technique, which leads to a complete image sub-division into non-overlapping regions or segments. The wide variety of segmentation approaches can be distinguished in two broad categories (Fortin et al., 2000; Muñoz et al., 2003; Mueller et al., 2004). The first category uses boundary techniques that apply edge detection methods to locate boundary elements, which are then filtered by using element attributes like angle measures or minimum length. After that, the remaining elements are connected into segments or objects. The second category uses region-growing algorithms that rely on 'seed' pixel groups (local and/or global minima), which grow until an abortion criterion is fulfilled (e.g., homogeneity, meeting of another boundary).

The functionality of the FNEA-algorithm used in this paper is described in detail by Baatz and Schäpe (2000) and Benz et al. (2004). The hierarchical region growing algorithm is widely accepted in the remote sensing community so that a multitude of references emerged in the last few years (see http://www.definiens.com/documents/publicationsearth.php). The crucial parameters are the homogeneity criteria H (scale parameter) and the weight parameters w_{color} and w_{shape} which allows adaptation of the heterogeneity definition to the desired target objects. While H affects the heterogeneity for each segmentation level, w_{color} and w_{shape} balance the spectral and shape heterogeneity ($w_{\text{color}} + w_{\text{shape}} = 1$). w_{shape} can be influenced by the weight parameters compactness w_{compt} and smoothness w_{smooth} ($w_{\text{compt}} + w_{\text{smooth}} = w_{\text{shape}}$). As the name implies, the higher w_{compt} the more segmentation results tend to compact shapes.

2.2. Segmentation validation

2.2.1. Object metrics

The quality of a segmentation result is connected with data quality (e.g., noise, spatial and spectral resolution) (Fortin et al., 2000) as well as the optimal customization of parameter settings, which enable the adaptation of segmentation results on target objects (Delves et al., 1992). The problem is that the customization is often a result of trial-and-error procedures (Hay et al., 2003; Stein and de Beurs, 2005). Thus, in recent years various object validation techniques were developed for assessing uncertainties in segmentation-based object extraction (Shi et al., 2005). Pixel metrics are appropriate to validate the quality of detected edges or boundaries (Delves et al.,



(a) Overlap (b) Containment (c) Containment and differences between object centers and gravity center G

Fig. 1. Topological and geometric relationships between two object levels.

1992; Prieto and Allen, 2003; Lucieer and Stein, 2002). However, Zhan et al. (2005) suggest object metrics for the validation of segments.

Object-based segmentation validation can be described as 'the problem of matching objects' (Zhan et al., 2005) where at least two hierarchical object-levels have to be considered. Object differences can be specified by, (1) topological and (2) geometric relationships (Molenaar, 1998; Ragia and Winter, 2000; de Bruin et al., 1999; Zhan et al., 2005):

(1) the topological relationships of interest are 'containment' and 'overlap' (Fig. 1). In Fig. 1a object P overlaps object O. Both object levels are not hierarchically connected. If a topological overlay GIS-operation was carried out, the resulting object $O1$ is contained in the primary object O (Fig. 1b). $O1$ and O are hierarchically coupled if a 'part of'-relation was created (de Bruin et al., 1999). As a result, $O1$ is a sub-object of the superior or super-object O. Metrics of containment arise from the comparison of object sizes between related super- and sub-object;
(2) geometric object differences can be determined by the comparison of object positions. Common metrics arise from distances between the gravity centers or skeletons as well as super- and sub-objects (Ragia and Winter, 2000; Zhan et al., 2005). Fig. 1c shows the gravity center G of the super-object O and the centers of sub-object $O1$ and $O2$.

In this study, we calculated the hierarchical object metrics 'relative area in super-object' RA_{SO} and 'relative position to super-object' RP_{SO}. Both metrics are implemented in the eCognition software (Benz et al., 2004). RA_{SO} expresses the topological relationship of two objects connected by a 'part of'-relation and arises from the ratio of the area size A of the object of interest $O1$ and the area size covered by its super-object O (Fig. 1b; Eq. (1)). The metric values are within the values 0 and 1. The value 1 indicates a complete match between sub- and super-object whereas values smaller than 1 represent sub-objects that are smaller than their super-objects.

$$RA_{SO} = \frac{A_{O1}}{A_O} \quad \text{with } RA_{SO} \in [0,1] \tag{1}$$

The hierarchical metric RP_{SO} relies on the gravity center G of the super-object and is calculated by dividing the distance d from the center of the object of interest C_{O1} to G by the distance d_{max} of the center of the most distant image object, which has the same super-object (Eq. (2); Fig. 1c: object $O2$). The metric value tends toward 0 if d reaches the minimum i.e., if the centers of both objects are in the same location. The metric value comes up to 1 if the distance between the centers of gravity of both objects is large.

$$RP_{SO} = \frac{d}{d_{max}} \quad \text{with } RP_{SO} \in [0,1] \tag{2}$$

2.2.2. Integral consideration of object metrics

Zhan et al. (2005) emphasized that topographical and geometric metrics have to be considered integrally in order to reach a more accurate detection of mismatched objects. A simple option is a threshold-based classification of the above-defined object metrics. Since a classification schema does not exist, we carried out a qualitative grouping of both metrics RA_{SO} and RP_{SO} by means of the K-means clustering algorithm (see Bishop, 1995; McGarigal et al., 2002) within the statistica-environment (http://www.statsoft.com; option 'maximum distance between clusters'). An advantage of this approach is that the two-dimensional metric feature space can be structured in an automatic manner. The resulting clusters show which objects represent the best

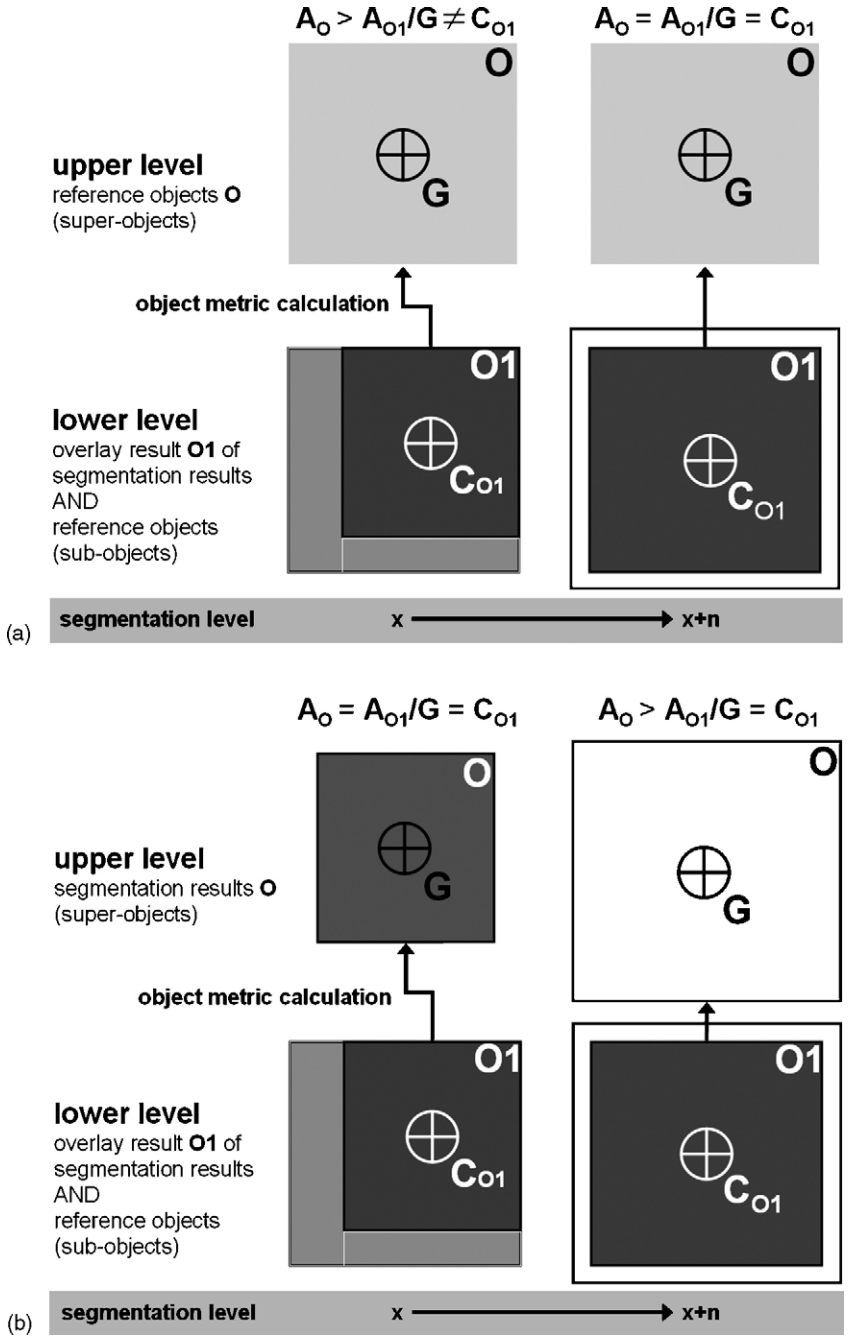


Fig. 2. Principle of local validation (a) test for over-segmentation; (b) test for under-segmentation.

geometrical (high cluster means of RA_{SO}) and topological (low cluster means of RP_{SO}) match of super- and sub-objects.

2.2.3. Local validation

Lucieer and Stein (2002) demonstrated that the application of topological object metrics enables the reference-based and object-specific validation of under- and over-segmentation. Our approach is similar, but uses a topological metric in combination with the geometric metric that was described in the last section. The calculation of both these metrics is carried out from a lower level to an upper level. The lower level represents the overlay result of reference objects and segmentation results (sub-objects). The upper level corresponds to either (1) reference objects or (2)

segmentation results (super-objects). Fig. 2 shows the two cases where we distinguish the segmentation levels x and $x + n$:

(1) over-segmentation can be identified by considering the reference objects O as the upper level. Over segmentation has occurred when the size of segmented objects is smaller than the size of reference objects ($A_O > A_{O1}$) (Fig. 2a: level x) or where object centers G and C_{O1} are different. If over-segmentation has occurred the resulting RA_{SO}-metrics are less than 1 and the RA_{PO}-metrics are greater than 0.

If the segmented object size is equal to or exceeds A_O then both A_O is identically to the size of overlay result A_{O1} and the locations of G and C_{O1} are congruent. That means that $RA_{SO} = 1$ and $RA_{PO} = 0$. Under-segmented objects cannot be detected in this instance because the reference object is now contained within the segmentation result. This means that A_O and A_{O1} are congruent (Fig. 2a: level $x + n$);

(2) under-segmentation can be identified by considering segmentation results as upper level (Fig. 2b). The topological and geometric similarity between A_O (level x) and A_{O1} lasts until A_O (level $x + n$) exceeds A_{O1}. The under-segmented objects can be identified using RA_{SO}- and RA_{PO}-metrics. For under-segmented objects RA_{SO} tends to 0 and RA_{PO} to 1 for larger A_O (level $x + n$).

By the simultaneous visualizing of the clustered object metric values (see Section 2.2.2) it is possible to estimate the balance between under- and over-segmentation related to each reference object.

2.2.4. Global validation

In contrast to the local validation that refers to single objects, the global validation is related to the entire image in order to deduce an optimal parameter setting H_{opt} and a segmentation accuracy CI_A. H_{opt} and CI_A are derived from (1) the aggregation of the calculated local validation results (clustered object metrics RA_{SO} and RP_{SO} in Section 2.2.3) and (2) the creation of CI–H-diagrams:

(1) the aggregation is described by the map complexity metric 'Comparison Index' CI. Complexity metrics are common in landscape analysis (O'Neill et al., 1986; McGarigal and Marks, 1994; Gustavson, 1998). In the connection with segmentation validation, Stein and de Beurs (2005) applied various semantic metrics to measure the complexity of segmentation results in order to quantify the semantic object accuracy. In our case the complexity metric is applied to assess the heterogeneity of the clustered object metrics related to the reference space. The reference space covers the spatial extent of all reference objects.

CI is calculated using Eq. (3). C_i is the comparison class, which represents clustered and ranked object metrics. The ranking is carried out by the highest value of RA_{SO} and the lowest value of RP_{SO} (highest ranking). A_{Ci} is equivalent to the proportion of C_i within the reference space. A complete topological and geometric match (i.e., image objects are identical to reference objects) is achieved if CI equals 100.

$$CI = \frac{\sum_{i=1}^{n}(C_i \times A_{Ci})}{n} \quad \text{with } CI \in [0, 100] \tag{3}$$

(2) In the CI–H-diagrams, the resulting CI values and the corresponding scale parameter H are plotted. Because of the consideration of two calculation directions (see Section 2.2.3) the resulting graphs intersect where over- and under-segmentation are balanced. CI_A and H_{opt} correspond to the intersection point on the CI- and H-axis (as shown in Fig. 6).

3. Study area

The study area (435 km^2) is situated in the south of the German state Saxony-Anhalt near the city of Halle (Saale) (Fig. 3). The accurate delineation of agricultural field boundaries is important because this area is a study site of a major soil erosion study within a project about 'Integrated River Basin Management on the example of the Saale River Basin' (Rode et al., 2002). The field boundaries are used here as model inputs into empirical and physics-based erosion modelling (cp. Merritt et al., 2003). In this paper, we use the 'Comparison Index' CI (Eq. (3)) to disaggregate the thematic class 'intensive agriculture' of the digital biotope and land use types in a scale of 1:10,000 (FANC, 2002; Rosenberg et al., 2003) with multi-temporal Land-sat Thematic Mapper (TM) and Landsat Extended Thematic Mapper (ETM) imagery. The images with a spatial resolution of 30 m were supplied by Eurim-age (http://www.eurimage.com) and cover three significant acquisition dates of the growth season for the year 1999 (April 30, July 3 and September 13). That means that 18 multi-temporal Landsat bands were available. Thermal infrared and panchromtic bands were not considered.

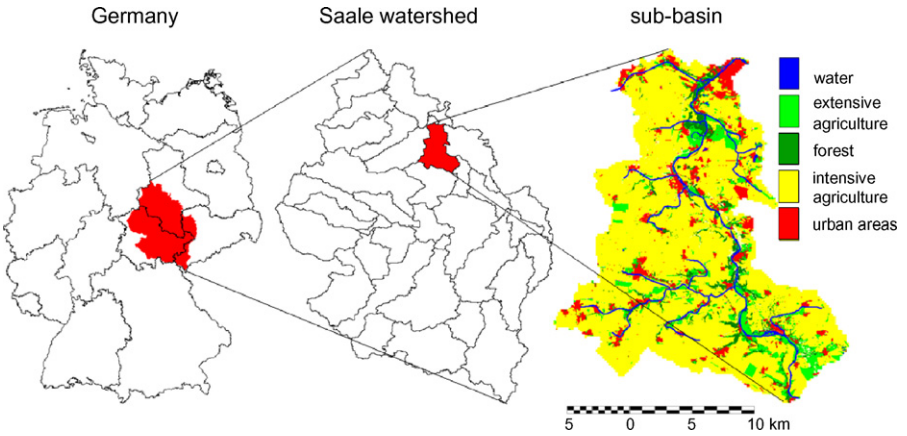


Fig. 3. Position of study area: sub-basin within the German Saale watershed.

Geo-referencing (to the Gauss Kruger projection, with Potsdam datum and the Bessel ellipsoid) was carried out for the images using 200 ground control points (GCPs) taken from topographic maps at a scale of 1:25,000. A linear resampling with the nearest neighbour algorithm was applied using ERDAS Imagine 8.5 (http://gis.leica-geosystems.com). The resampling was done using an output pixel size of 30 m with a RMSE of less than one pixel. The validation objects (agricultural fields) used in this study were selected from the Landsat imagery using stratified random sampling (Stehmann, 1992). The buffering of sample points produced sample areas which provide the basis for a manual on-screen digitizing of field parcels within the ArcInfo -environment (http://www.esri.-com). Like Devereux et al. (2004), independent interpreters collected samples from the test image on the basis of visual interpretation. This approach was used to ensure that the validation objects were derived from the same image resolution as the image segments.

4. Results

4.1. Reference objects and segmentation

The results of manual field detection are visualized in Fig. 4a. Reference objects (400) were digitized. The validation procedure will be exemplified by the reference object shown in red in the south of the study area (Fig. 4b). The round field boundaries are a function of the circular buffer of 1000 m that we used around the randomly selected points.

The segmentation parameter settings are listed in Table 1. The H-, w_{compt}-and w_{shape}-parameters were generated using trial-and-error tests to narrow the

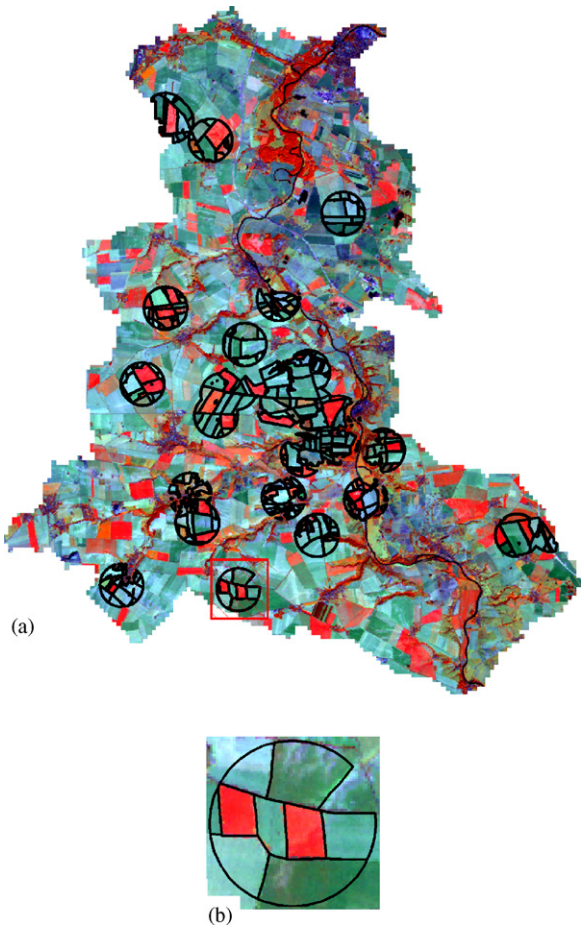


Fig. 4. Positions of reference data within the thematic land-use class 'intensive agriculture' (a) and reference example objects (b); Landsat ETM image from September 13, 1999 (band combination 4-5-3).

Table 1
Parameter setting versions for segmentations of Lands at imagery with FNEA-algorithm

Parameter	Version				
	*S*1	*S*2	*S*3	*S*4	*S*5
w_{shape}	0.2	0.4	0.4	0.4	0.5
w_{compt}	0.5	0.5	0.3	0.7	0.3
H	10, 20, . . ., 100				

parameter range. It is common for all parameter setting versions that ten segmentation levels were produced whereas each level is represented by a specific scale parameter *H* value (10, 20, . . ., 100). From the multi-temporal data set all 18 Landsat bands were used as input data.

4.2. Local validation

The local validation procedure is illustrated by means of segmentation results of *S*1 version (cp. Table 1). Fig. 5e–h show the influence of scale parameter *H*. A greater *H* value affects more heterogeneous and greater segmentation objects.

As explained in Section 2.2.3, the calculation of object metrics and the *K*-means clustering was carried out in two directions. The cluster means refer to all reference objects.

The clustered and ranked object metrics RA_{SO} and RP_{SO} calculated using the manually digitized reference objects are presented according to the corresponding scale parameter *H* in Fig. 5a–d. The RA_{SO} and RP_{SO} results calculated using the segmentation results are

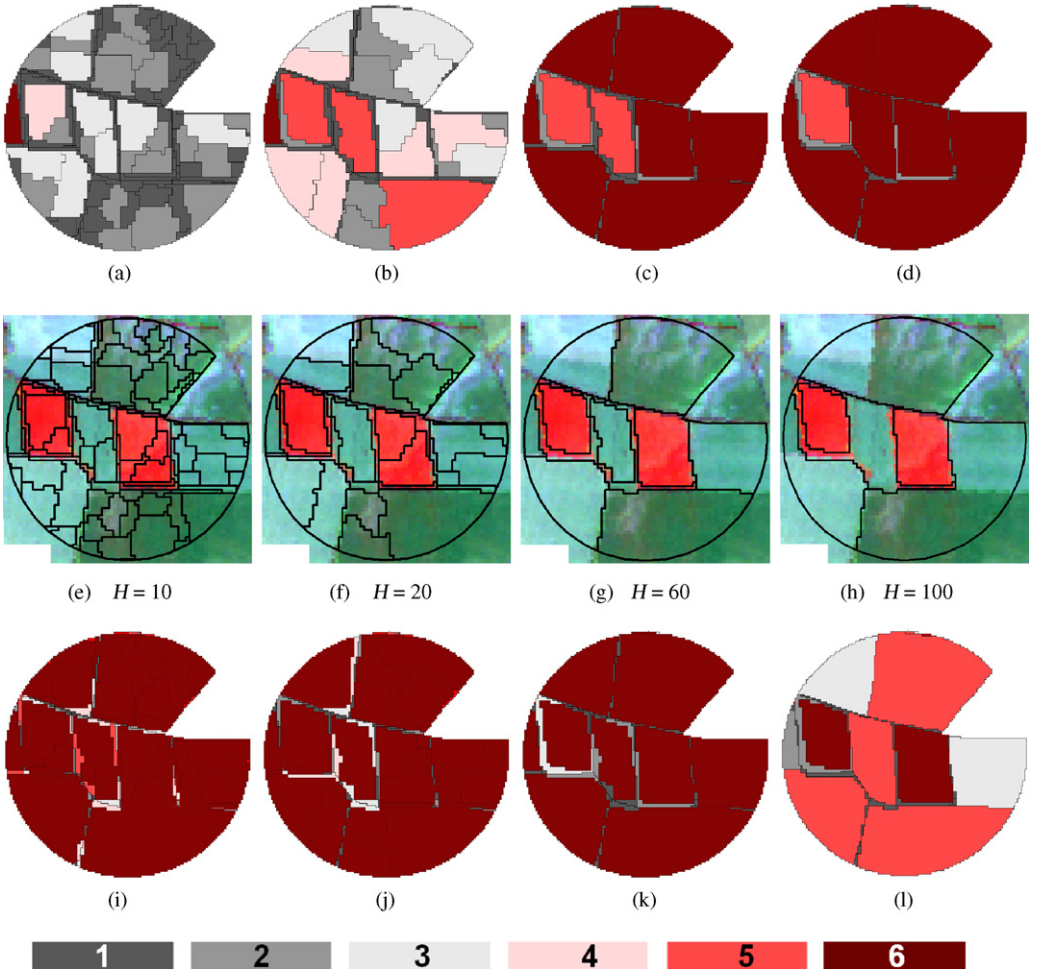


Fig. 5. Visualization of clustered and ranked object metrics RA_{SO} and RP_{SO} by example of segmentation version *S*1 (cp. Table 1). e–h: segmentation results layed over the Landsat ETM image from September 13, 1999 (band combination 4-5-3); a–d: metric calculation relative to reference objects (cp. Fig. 4b) = test for over-segmentation; i–l: metric calculation relative to segmentation results = test for under-segmentation. Attributes of corresponding comparison classes are listed in Table 2.

Table 2
Ranked *K*-means cluster means *C*, class proportions A_C (%), 'Comparison Indices' CI of object metrics and corresponding scale parameters *H* for segmentation version SI (see Table 1)

H	object metric	comparison class 0		1		2		3		4		5		*CI*
		A_C	C	A_C	C	A_C	C	A_C	C	A_C	C	A_C	C	
10^A	RA_{SO}	11.0	0.01	33.8	0.05	30.2	0.07	0.8	0.33	19.9	0.41	4.2	0.91	39
	RP_{SO}		0.98		0.68		0.35		0.95		0.32		0.05	
20^A	RA_{SO}	3.0	0.01		0.05	3.3	0.08	1.1	0.33	36.6	0.46	26.8	0.89	65
	RP_{SO}		0.99	8.3	0.71		0.35		0.96		0.35		0.05	
40^A	RA_{SO}	5.7	0.01		0.03	6.1	0.09	0.8	0.35	17.7	0.57	66.4	0.93	**84**
	RP_{SO}		0.99	3.2	0.75		0.41		0.96		0.30		0.03	
50^A	RA_{SO}	5.2	0.01		0.03	5.9	0.11	0.9	0.36	13.2	0.63	72.8	0.94	**87**
	RP_{SO}		0.99	1.9	0.76		0.42		0.94		0.25		0.03	
60^A	RA_{SO}	5.0	0.01	2.0	0.03	2.2	0.11	1.0	0.36	10.0	0.61	79.7	0.94	**90**
	RP_{SO}		0.99		0.76		0.41		0.94		0.26		0.03	
70^A	RA_{SO}	4.6	0.01	1.6	0.03	2.2	0.11	0.8	0.36	7.3	0.61	83.6	0.94	**91**
	RP_{SO}		0.99		0.76		0.41		0.94		0.26		0.03	
100^A	RA_{SO}	4.4	0.01	1.5	0.03	1.7	0.11	0.8	0.36	5.1	0.61	86.5	0.94	92
	RP_{SO}		0.99		0.76		0.41		0.94		0.27		0.03	
10^B	RA_{SO}	2.9	0.05	1.2	0.11	1.0	0.13	2.1	0.36	5.1	0.60	87.8	0.98	94
	RP_{SO}		0.99		0.66		0.28		0.96		0.28		0.02	
20^B	RA_{SO}	3.7	0.02	1.5	0.08	1.0	0.10	1.5	0.35	4.7	0.58	87.7	0.96	93
	RP_{SO}		0.99		0.68		0.29		0.97		0.28		0.02	
40^B	RA_{SO}	4.4	0.01	2.9	0.07	0.9	0.09	1.1	0.35	6.3	0.58	84.5	0.95	**91**
	RP_{SO}		0.99		0.67		0.28		0.96		0.30		0.02	
50^B	RA_{SO}	4.4	0.01	1.1	0.07	4.2	0.08	1.4	0.35	10.2	0.56	78.7	0.95	**90**
	RP_{SO}		0.99		0.29		0.67		0.96		0.31		0.03	
60^B	RA_{SO}	4.5	0.01	1.6	0.07	4.4	0.08	1.6	0.35	11.1	0.55	76.7	0.95	**89**
	RP_{SO}		0.99		0.30		0.68		0.96		0.31		0.03	
70^B	RA_{SO}	4.6	0.01	1.6	0.09	4.8	0.09	1.4	0.33	16.1	0.56	71.4	0.95	**87**
	RP_{SO}		0.99		0.30		0.68		0.97		0.31		0.03	
100^B	RA_{SO}	4.7	0.01	1.9	0.07	6.0	0.08	3.0	0.35	19.7	0.55	64.6	0.95	85
	RP_{SO}		0.99		0.31		0.68		0.97		0.31		0.03	

A: Object metric calculation relative to reference objects = test for over-segmentation.
B: Object metric calculation relative to segmentation results = test for under-segmentation.

shown in Fig. 5i–l. The ranked cluster means correspond to colored comparison classes. The comparison classes symbolize high (= dark red) and low (= gray) topological and geometric similarities between reference and segmentation objects. In other words, image segments shown in dark red represent a geometrical and topological match with their reference object.

The comparison classes refer to Table 2 where the attributes of the displayed results are listed. For the scale parameter $H = 60$ for instance (see Fig. 5c and k and gray highlighted cells in Table 2) the topological and geometric similarity between reference objects and segmentation results is high in both calculation directions because of the high proportion of the comparison class 5 (see framed values in Table 2). The corresponding cluster mean values 0.94, 0.95 (=high topological similarity) and 0.03 (high geometric similarity) are underlined. A high topological and geometric similarity also means that under- and over-segmentation are balanced. In contrast, an unbalanced situation indicates under- or over-segmentation. Under-segmented objects are detected by object metrics calculated relative to reference objects (Fig. 5a–c, comparison classes 0–4 in Table 2). However, over-segmented objects are detected by object metrics calculated relative to segmentation results (Fig. 5i–l, comparison classes 0–4).

4.3. Global validation

The aggregation of comparison class proportions leads to the complexity metric 'Comparision Index' CI and CI–*H*-diagrams (see Section 2.2.4). The aggregation was carried out for each segmentation level (defined by scale parameter *H*) and each segmentation version (see Table 1). The CI-values were calculated from the comparison class proportions A_C (see Eq. (3)) for CI and *H* values associate with the point where

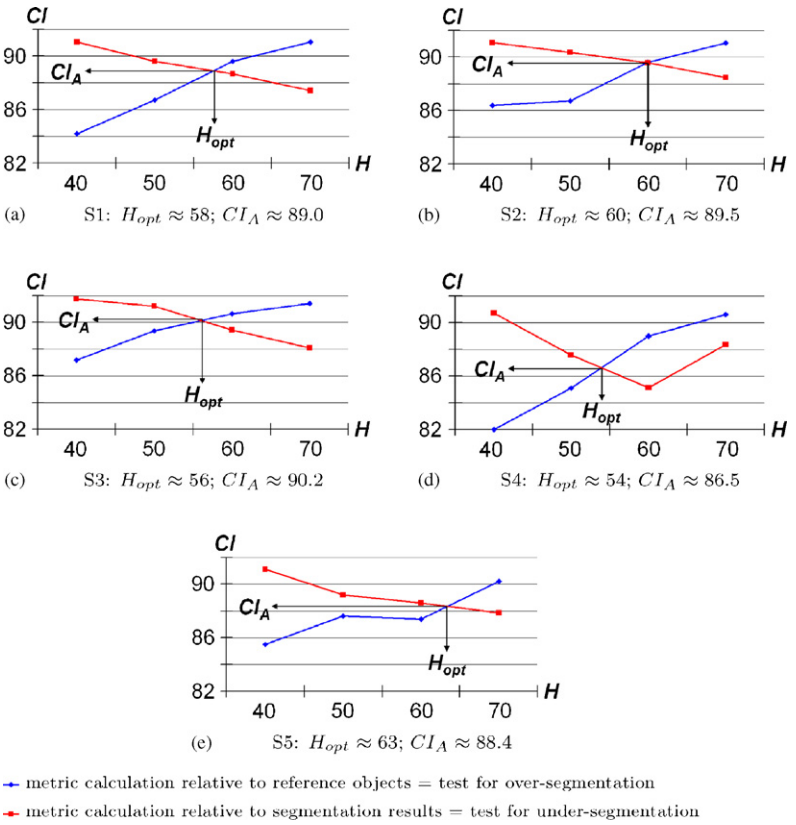


Fig. 6. Estimation of segmentation accuracy CI_A and optimal heterogeneity H_{opt} for segmentation versions of Table 1.

over- and under-segmentation are balanced. Segmentation tests showed that this point is achieved around the scale parameter $H = 60$. In order to meet the H_{opt}-value, the CI values for $H = 40, 50, 60, 70$ are calculated (bold highlighted CI values in Table 2). The resulting CI–*H*-diagrams show differences regarding the 'segmentation accuracy' CI_A (Fig. 6). CI_A values vary from 88.4 (*S*5) to 90.2 (*S*3). Accordingly, the results of *S*3 version show the best topological and geometric similarity. This means that the optimal scale parameter setting for detecting agricultural field boundaries in the study area is H_{opt} of 56, and this scale parameter should be used in combination with a w_{shape} parameter of 0.4 and a w_{compt} parameter of 0.3 (which were the settings used for the *S*3 segmentation).

5. Conclusions

The main objective of this paper was the determination of an optimal parameter setting of the FNEA-segmentation method in eCognition™ based on the assumption that an optimal parameter setting is reached when over-and under-segmentation are balanced. The approach developed in this paper demonstrates how a two-stage process can be used to identify a segmentation scale *H* that is close to optimal using trial-and-error tests in combination with the CI metric. Once this scale has been identified a more detailed examination of the CI–*H*-diagrams can be used to identify precisely what *H* value and which w_{shape} and w_{compt} values will yield the most accurate image segmentation results.

The proposed segmentation validation procedure uses object-based metrics and a complexity metric, which were already applied to other studies (e.g., Lucieer and Stein, 2002; Zhan et al., 2005; Stein and de Beurs, 2005). The novelty of our approach is that we combine both enabling (1) local and (2) global validation of segmentation results:

(1) the local validation is based on the consideration of both topological and geometric object metrics. The metrics characterize differences in size and position between segmentation results and reference objects. In the study, we calculated metrics for 400 reference objects. The statistical metric grouping by *K*-means clustering and ranking revealed which objects show

high and low topological and geometric similarities. The metric calculations were carried out with reference to manually digitized field boundaries and super-objects within the segmentation hierarchy. Thus, under- and over-segmented objects could be visualized and classified;

(2) global validation enables the assessment of segmentation results throughout the whole study area. For this purpose, the clustered and ranked metrics were aggregated to 'Comparison Indices' CI by comparing the proportions of cluster areas where the clusters were weighted according to their rank. Depending on the input data and parameter settings used, CI–*H*-diagrams allow the estimation of 'optimal segmentation results'. This way parameters of the used segmentation algorithm can be optimized.

As the name implies, the 'Comparison Index' enables a relative comparison of segmentation results. Here, different field detection results based on FNEA algorithm were compared with reference objects. Other possible fields of application are the investigation of different segmentation procedures (see Delves et al., 1992; Meinel and Neubert, 2004). The local validation procedure is suitable for the assessment of existing thematic data sets or change detection analysis of objects shapes (e.g., field sizes) which both also could help to improve the spatial input data sets for erosion and water quality models for more realistic simulations.

References

Baatz, M., Schäpe, A., 2000. In: Strobl, J., Blaschke, T. (Eds.), Multiresolution Segmentation: An Optimization Approach for High Quality Multi-scale Image Segmentation, vol. 12. Angewandte Geographische Informationsverarbeitung, Salzburg, pp. 12–23.

Benz, U.C., Hofmann, P., Willhauck, G., Lingenfelder, I., Heynen, M., 2004. Multi-resolution, object-oriented fuzzy analysis of remote sensing data for GIS-ready information. J Photogramm. Remote Sens. 58, 239–258.

Betenuth, M., 2004. Extraction of field boundaries and wind erosion obstacles from aerial imagery, in: Seyfert, E., (Ed.), Instrumentarien zur nachhaltigen Entwicklung von Landschaften. Vol. 13 of Publikationen der Deutschen Gesellschaft für Photogrammetrie, Fernerkundung und Geoinformation e. V. pp. 77–86.

Bishop, C., 1995. Neural Networks for Pattern Recognition. Oxford University Press, Oxford, England.

Congalton, R., Green, K., 1990. Assessing The Accuracy of Remotely Sensed Data: Principles and Practices. Lewis Publishers, Boca Raton.

de Bruin, S., Wielemaker, W., Molenaar, M., 1999. Formalisation of soil-landscape knowledge through interactive hierarchical disaggregation. Geo-derma 91, 151–172.

Delves, L.M., Wilkinson, R., Oliver, C.J., White, R.G., 1992. Comparing the performance of SAR image segmentation algorithms. Int. J. Remote Sens. 13 (11), 2121–2149.

Deumlich, D., Kiesel, J., Thiere, J., Reuter, H., Völker, L., Funk, R., 2006. Application of the Site Comparison Method (SICOM) to assess the potential erosion risk—a basis for the evaluation of spatial equivalence of agri-environmental measures. Catena 68 (2–3), 141–152.

Devereux, B., Amablea, G., Costa Posada, C., 2004. An efficient image segmentation algorithm for landscape analysis. Inter. J. Appl. Earth Observ. Geoinform. 6, 47–61.

EU, 2002. European communities—the Water Framework Directive (WFD) and tools within the Common Agricultural Policy (CAP) to support its implementation. Working Document–DG ENV.B.1/BB D(2002).

Evans, C., Jones, R., Svalbe, I., Berman, M., 2002. Segmenting multispectral Landsat TM images into field units. IEEE Trans. Geosci. Remote Sens. 40 (5), 1054–1064.

FANC, 2002. A system for the survey of biotope and land use types (survey guide)–Standard biotope and land use types for FCIR aerial photograph supported biotope and land use survey for the Federal Republic of Germany. Vol. 73 of Schriftenreihe für Landespflege und Naturschutz. Federal Agency for Nature Conservation, Bonn-Bad Godesberg.

Foody, G., 2002. Status of land cover classification accuracy assessment. Remote Sens. Environ. 80, 185–201.

Fortin, M.-J., Olson, R., Ferson, S., Iverson, L., Hunsaker, C., Edwards, G., Levine, D., Butera, K., Klemas, V., 2000. Issues related to the detection of boundaries. Landscape Ecol. 15, 453–466.

Fuller, R., Smith, G., Sanderson, J., Hill, R., Thompson, A., 2002. The UK land cover map 2000: construction of a parcel-based vector map from satellite images. Cartographic J. 39 (1), 15–25.

Gustavson, E., 1998. Quantifying landscape spatial pattern: what is the state of the art? Ecosystems 1, 143–156.

Hay, G., Blaschke, T., Marceau, D., Bouchard, A., 2003. A comparison of three image-object methods for the multiscale analysis of landscape structure. Photogramm. Remote Sens. 57, 327–345.

Lucieer, A., Stein, A., 2002. Existential uncertainty of spatial objects segmented from satellite sensor imagery. IEEE Trans. Geosci. remote sens. 40 (11), 2518–2521.

McGarigal, K., Cushman, S., Stufford, S., 2002. Multivariate Statistics for Wildlife and Ecology research. Springer, New York, Berlin, Heidelberg.

McGarigal, K., Marks, B., 1994. Fragstats: Spatial Pattern Analysis Program for Quantifying Landscape Structure—Program Documentation. Oregan State University, Corvallis.

Meinel, G., Neubert, M., 2004. A comparison of segmenation programs for high resolution remote sensing data. Int. Arch. ISPRS XXXV, Part B Commision 4, 1097–1105.

Merritt, W., Letcher, R., Jakeman, A., 2003. A review of erosion and sediment transport models. Environ. Modell. Software 18, 761–799.

Molenaar, M., 1998. An Introduction to The Theory of Spatial Object Modelling. Research Monographs in Geographical Information Systems, Tayler & Francis.

Mueller, M., Segl, K., Kaufmann, H., 2004. Edge- and region-based segmentation technique for the extraction of large, man-made objects in high-resolution satellite imagery. Pattern Recognit. 37, 1619–1628.

Muñoz, X., Freixenet, J., Cufí, X., Martí, J., 2003. Strategies for image segmentation combining region and boundary information. Pattern Recognit. Lett. 24, 375–392.

Mysiak, J., Rosenberg, M., Hirt, U., Haase, D., Petry, D., Frotscher, K., 2004. Uncertainty in the spatial transformation of socioeceonomic data for the implementation of the water framework directive. in: 10th EC GI & GIS Workshop: ESDI-State of the Art. Warsaw, Poland, June 23–25.

O'Neill, E., DeAngelis, D., Waide, J., Allen, T., 1986. A Hierarchical Concept of Ecosystems. Princeton University Press, Princeton.

Prieto, M.S., Allen, A.R., 2003. A similarity metric for edge images. IEEE Trans. Pattern Anal. Machine Intell. 25 (10), 1265–1273.

Ragia, L., Winter, S., 2000. Contributions to a quality description of a real objects in spatial data sets. J. Photogramm. Remote Sens. 55, 201–213.

Rode, M., Klauer, B., Krause, P., Lindenschmidt, K., 2002. Integrated river basin management: a new ecologically-based modelling approach. Ecohydrol. Hydrobiol. 2, 171–179.

Rosenberg, M., Haase, D., Augenstein, I., Erfurth, S., Schulz, G., 2003. Dokumentation zur Erstellung eines aggregierten Biotop- und Nutzungstypen-Datensatzes für das Einzugsgebiet der Saale auf der Basis der CIR-luftbildgestützten Biotop- und Nutzungstypenkartierung der Bundesländer Sachsen, Sachsen-Anhalt und Thüringen. Tech. rep., Centre for Environmental Research, Department of Applied Landscape Ecology, Permoserstr. 15, 04318 Leipzig, Germany, (in German).

Shi, W., Ehlers, M., Molenaar, M., 2005. Uncertainties in integrated remote sensing and GIS. Int. J. Remote Sens. 26 (14), 2911–2915.

Stehmann, S., 1992. Comparison of systematic and random sampling for estimating the accuracy of maps generated from remotely sensed data. Photogramm. Eng. Remote Sens. 58, 1343–1350.

Stein, A., de Beurs, K., 2005. Complexity metrics to quantify semantic accuracy in segmented Landsat images. Int. J. Remote Sens. 26 (14), 2937–2951.

Takken, I., Govers, G., Jetten, V., Nachtergaele, J., Steegen, A., Poesen, J., 2001. Effects of tillage on runoff and erosion patterns. Soil Tillage Res. 31 (1–2), 55–60.

Van Oost, K., Govers, G., Desmet, P., 2000. Evaluating the effects of changes in landscape structure on soil erosion by water and tillage. Landscape Ecol. 15, 577–589.

Zhan, Q., Molenaar, M., Tempfli, K., Shi, W., 2005. Quality assessment for geo-spatial objects derived from remotely sensed data. Int. J. Remote Sens. 26 (14), 2953–2974.