CONTRIBUTION TO THE ASSESSMENT OF SEGMENTATION QUALITY FOR REMOTE SENSING APPLICATIONS

Uwe WEIDNER

Institute of Photogrammetry and Remote Sensing, University Karlsruhe (TH), Englerstraße 7, 76128 Karlsruhe, Germany
uwe.weidner@kit.edu

KEY WORDS: Image processing, classification, image understanding, feature extraction, land cover, segmentation, accuracy assessment

ABSTRACT:

Within object-oriented or segment-based classification approaches the segmentation step is decisive, because the results form the basis for the following classification. Despite known investigations and approaches of quality evaluation for segmentations, the question of how to access this quality with respect to remote sensing applications is not yet completely answered. This contribution therefore addresses this topic.

1 INTRODUCTION

The first task within a segment-based classification is to derive a segmentation of the remote sensing data. This step is decisive, because the segments and their properties constitute the basis for the subsequent classification. Therefore, the quality of the segmentation has a direct impact on the quality of the classification. The most common problems are over- and undersegmentation combined with the problem, that the objects of interest may be scale dependent (Schiewe, 2002). Oversegmentation is normally of minor impact on the result when the segments are assigned to the correct classes, but problems may occur, if geometric properties of the segments are used within the classification step. Furthermore, the segments may not be useful for further analysis (Lemp and Weidner, 2005). Undersegmentation leads to problems in classification, because the segment inherent properties are distorted, thus instead of the *mixed-pixel* problem in pixel-based classification a *mixed-segment* problem occurs (Weidner and Bähr, 2007). Partly, these problems are overcome by a multiresolution hierarchical segmentation (Baatz and Schäpe, 1999) allowing for a classification of objects on different scales and thereby generalisation levels. Nonetheless, the segmentation on the different levels have to be meaningful with respect to the application. Despite known investigations the question of how to assess the quality of a segmentation of remote sensing data is not yet completely answered.

The paper discusses quantities and propose a framework for the quality assessment of segmentations focussing on remote sensing applications. For segmentation the software Definies Developer is used. The selected quantities are applied to evaluate the segmentations of different remote sensing data, namely Landsat and QuickBird data.

2 RELATED WORK

Frameworks for quality assessments have been proposed and published in the computer vision community, e.g. (Hoover et al., 1996) for range images, (Zhang, 2001) for optical images, and (Udupa et al., 2006) for voxel data sets. (Neubert et al., 2006) address the topic of segmentation quality for remote sensing in which context only a few investigations are published, see (Neubert et al., 2006) for a review. Their quality evaluation for different segmentation approaches is based on a qualitative visual and a quantitative evaluation based on geometrical features of the segments, e.g. area, perimeter, and shape, using manually derived ground truth. Partly, the results of the used quantities (average distances to ground truth) are difficult to interpret. Furthermore, some of them are correlated and the reliability of some are also dependent on the segment size, e.g. the shape index which is normally more reliable for larger segments. Evaluation quantities should also account for the uncertainty of the segments border as e.g. proposed in (Schuster and Weidner, 2003) and for ease of interpretation should have properties of metrics as e.g. given in (Unnikrishnan et al., 2007) or be normed to a fixed range of values as e.g. in (Correira and Pereira, 2003).

In the following section approaches and quantities for quality assessment of segmentations are discussed in more detail including the new approach in Section 3.3. Section 4 describes the data used for evaluation and presents the results of the recent work followed by conclusions and an outlook.

3 QUANTITATIVE QUALITY ASSESSMENT

3.1 Evaluation approaches

A segmentation of an image $\mathcal{I}$ is given by $\mathcal{I} = \bigcup_i \mathcal{S}_i$ where $\mathcal{S}_i$ are connected components with the conditions that segments are disjoint $\mathcal{S}_i \cap \mathcal{S}_j = \emptyset \quad \forall i, j$, each segment is homogeneous

$$H(\mathcal{S}_i) = \texttt{true} \quad \forall i \tag{1}$$

and the union of adjacent segments is not homogeneous

$$H(\mathcal{S}_i \cup \mathcal{S}_j) = \texttt{false} \quad \forall i, \forall j \quad \mathcal{S}_j \in \mathcal{N}(\mathcal{S}_i) \tag{2}$$

with respect to a homogeneity criterion H. For multiresolution hierarchical segmentations with 1:n parent-child relations (e.g. (Baatz and Schäpe, 1999)) furthermore

$$\mathcal{S}_{i.n} = \bigcup_j \mathcal{S}_{ij.m} \quad \text{with} \quad n > m \tag{3}$$

where n and m denote the levels and $m = 0$ is the pixel-level and

$$\mathcal{S}_{i.n} \cap \mathcal{S}_{jk.m} = \emptyset \quad \forall i, j, k \quad \text{with} \quad i \neq j \, , n > m \tag{4}$$

must hold.

(Zhang, 2001) categorised three groups of evaluation approaches for segmentations: (A) analytical evaluation of segmentation approaches based on their (theoretical) properties, and empirical evaluation using reference data and either (B) evaluating goodness – e.g. intra-segment uniformity related to (1) and inter-segment contrast related to (2) – or (C) discrepancy – e.g. number and/or position of missegmented pixels and attribute values of segments. The evaluation approach of (Radoux and Defourny, 2006) belongs to (B) and (C), because it evaluates the spectral information of the segments - thus the class separability - using the Bhattacharrya distance and the segments' geometry. In order to evaluate the geometry they analyse false positive and false negative areas compared to the reference segments taking the border length for normalisation. Evaluation of video sequences are proposed in (Correira and Pereira, 2003) and (Gelasca et al., 2004). Both references also belong to the above mentioned groups (B) and (C). They combine shape and partly inherent spectral / textural information of the segments. The latter reference also investigates the relation of their quantities to perceptual significance. Furthermore the changes within the segmentation sequence derived on the video sequence is considered. (Cardenes et al., 2007) also propose to use different normed criteria based on the geometry and spectral properties of segments and combine the results via the root of the squared mean. In case of just one combined quantity the interpretation with respect to different aspects of segment properties is difficult.

The focus of this contribution is on the evaluation of the geometry, because the usefullness of segments due to (1) with respect to class separability is evaluated by quality assessment for the final classification. Furthermore discrepancy methods (C) are considered to be more effective (cf. (Zhang, 2001), (Neubert et al., 2006)), thus those will be considered in the following.

Besides the categorisation above approaches take either the segment's borders or the entire segment as basis for the analysis. (Paglieroni, 2004) match boundaries of the extracted segments to those of the reference segments, but this matching may not guarantee symmetrical results, because it depends on the matching direction. Those approaches taking the entire segment often regard the segmentation evaluation as a binary classification evaluation (e.g. (Hoover et al., 1996), (Min et al., 2004), (Srivastava, 2006), (Van Droogenbroeck and Barnich, 2005), and (Udupa et al., 2006), who work on fuzzy sets) and some also include topological criteria (e.g. (Gelasca et al., 2004)).

3.2 Quantities for quality assessment

In order to discuss evaluation quantities, let $\mathcal{R}_{i[k]}$ denote a reference segment of class k and $\mathcal{S}_{j(k)}$ denote a derived segment of class k. The segments $\mathcal{S}_{j(k)}$ are assigned to a reference segment $\mathcal{R}_{i[k]}$ by the set

$$\mathcal{S}_{i[k]} = \bigcup_j \mathcal{S}_{j(k)} \quad \forall \mathcal{S}_{j(k)} \quad \text{with} \quad \frac{|\mathcal{S}_{j(k)} \cap \mathcal{R}_i[k]|}{|\mathcal{S}_{j(k)}|} > 0.5 \qquad (5)$$

where $|\mathcal{A}|$ denotes the number of pixels of $\mathcal{A}$ or its area respectively using their overlap with the reference segment as criterion. Furthermore let $\sharp\mathcal{A}$ denote the number of segments of set A. For the comparison of the two segmentations given by their segments $\mathcal{S}_{j[k]}$ and $\mathcal{R}_{i[k]}$ – the latter segments of a second segmentation or reference segments – quantities like (omitting the subscripts for ease of reading) the *detection rate*

$$\rho_d = \frac{|\mathcal{S} \cap \mathcal{R}|}{|\mathcal{R}|} \quad \text{with} \quad \rho_d \in [0,1] \qquad (6)$$

quantifying the rate of correctly segmented pixels, the *false positive rate*

$$\rho_{FP} = \frac{|\mathcal{S} \setminus \mathcal{R}|}{|\mathcal{R}|} \quad \text{with} \quad \rho_{FP} \geq 0 \qquad (7)$$

and the *false negative rate*

$$\rho_{FN} = \frac{|\mathcal{R} \setminus \mathcal{S}|}{|\mathcal{R}|} \quad \text{with} \quad \rho_{FN} \in [0,1] \qquad (8)$$

quantifying either the false positive or the false negative pixels with respect to the area of $\mathcal{R}$ can be used. Replacing the area of $\mathcal{R}$ by the area of the intersection $\mathcal{S} \cap \mathcal{R}$ leads to the *branch factor*

$$\rho_b = \frac{|\mathcal{S} \setminus \mathcal{R}|}{|\mathcal{S} \cap \mathcal{R}|} \quad \text{with} \quad \rho_b \geq 0 \qquad (9)$$

and the *miss factor*

$$\rho_m = \frac{|\mathcal{R} \setminus \mathcal{S}|}{|\mathcal{S} \cap \mathcal{R}|} \quad \text{with} \quad \rho_m \geq 0 \qquad (10)$$

respectively. Assuming that $\mathcal{R} \neq \emptyset$ the rates of *false positives* and *false negatives* have the advantage that they are always defined. Furthermore ρ_{FN} is normed to $[0,1]$ like the *detection rate* ρ_d. The sum of ρ_{FP} and ρ_{FN} leads to the *shape dissimilarity*

$$\rho_s = \frac{|(\mathcal{R} \setminus \mathcal{S}) \cup (\mathcal{S} \setminus \mathcal{R})|}{|\mathcal{R}|} \quad \text{with} \quad \rho_s \geq 0 \qquad (11)$$

evaluating both types of deviations - missed and additional pixels. This is also true for the *quality rate*

$$\rho_q = \frac{|\mathcal{S} \cap \mathcal{R}|}{|\mathcal{S} \cup \mathcal{R}|} = 1 - \frac{|(\mathcal{S} \setminus \mathcal{R}) \cup (\mathcal{R} \setminus \mathcal{S})|}{|\mathcal{S} \cup \mathcal{R}|} \qquad (12)$$

with $\rho_q \in [0,1]$. A comparison with the *detection rate* yields that these two quantities just differ in the denominator. The advantages of the *quality rate* compared to the *shape dissimilarity* is the symmetry with respect to $\mathcal{R}$ and $\mathcal{S}$ and its fixed range. The term $\delta_s = 1 - \rho_q$ has been used as similarity measure in computer vision (cf. e.g. (Keim, 1999)). Without loss of generality in the context of remote sensing applications the same class labels for the segmentation and the reference data can be assumed and therefore the *quality rate* fulfils the requirements for quantities for segmentation quality assessment defined by (Unnikrishnan et al., 2007). As example for this, $\mathcal{R} \neq \emptyset$ is assumed and three degenerated cases are considered: (a) $\mathcal{S} = \emptyset$, (b) $\mathcal{S} = \mathcal{I} \setminus \mathcal{R} = \mathcal{R}^*$ - thus being the complement of $\mathcal{R}$, and (c) $\mathcal{S} = \mathcal{I}$ For cases (a) and (b) the $\rho_q = 0$ showing that the quantity is meaningful also for these degenerated cases. In case (c) $\rho_q = \frac{|\mathcal{R}|}{|\mathcal{I}|}$ and therefore directly dependant on the area of $\mathcal{R}$. Thus, if $\mathcal{R} \to \emptyset$ then $\rho_q \to 0$ and if $\mathcal{R} \to \mathcal{I}$ then $\rho_q \to 1$ respectively.

3.3 New approach for quality assessment

The quantities discussed in the previous section are based on binary consideration of the deviations between the two sets $\mathcal{S}$ and $\mathcal{R}$. In order to increase the influence of larger deviations between the two sets on the quality measure ρ_q the weighted quality rate

$$\rho_{qw} = 1 - \frac{A}{|\mathcal{R} \cap \mathcal{S}| + A} \quad \text{with} \quad \rho_{qw} \in [0,1] \qquad (13)$$

is defined, where

$$A = \sum_{x \in (\mathcal{S} \setminus \mathcal{R})} w(d(x,\mathcal{R})) + \sum_{x \in (\mathcal{R} \setminus \mathcal{S})} w(d(x,\mathcal{S}))$$

and

$$d(x, \mathcal{A}) = \inf\{\rho(x, a) : a \in \mathcal{A}\} \tag{14}$$

denotes the distance and $w(x)$ a weighting function like the linear functions

$$w(d(x, \mathcal{A})) = \frac{1}{\Delta_d} d(x, \mathcal{A}) \tag{15}$$

or

$$w_T(d(x, \mathcal{A})) = \begin{cases} 0 & d \leq d_T \\ \frac{1}{\Delta_d}(d(x, \mathcal{A}) - d_T) & \text{else} \end{cases} \tag{16}$$

The function given by (16) allows to introduce a buffer around the sets. Therefore the accuracy of the boundaries can be taken into account. Alternative weighting functions and a discussion can be found in (Schuster and Weidner, 2003). The weighted quality rate ρ_{qw} is symmetric with respect to $\mathcal{R}$ and $\mathcal{S}$ like the quality rate according to (12). Analysing the three cases mentioned above yields $\rho_{qw} = 0$ for cases (a) and (b) and for case (c) where $\mathcal{S} = \mathcal{I}$ is assumed

$$\rho_{qw} = \frac{|\mathcal{R}|}{|\mathcal{R}| + \sum_{x \in (\mathcal{S} \setminus \mathcal{R})} w(d(x, \mathcal{R}))}$$

Again, if $\mathcal{R} \to \emptyset$ then $\rho_{qw} \to 0$ and if $\mathcal{R} \to \mathcal{I}$ then $\rho_{qw} \to 1$ respectively. The quantity given in (13) is also used by (Cardenes et al., 2007) with the squared Euclidean distance as weight. In order to relate the quality rate in (12) to the distances from the border $\partial\mathcal{R}$ of the reference $\mathcal{R}$ for both sets $\mathcal{S} \setminus \mathcal{R}$ and $\mathcal{R} \setminus \mathcal{S}$ the weighted quality rate

$$\rho^*_{qw} = 1 - \frac{A^*}{|\mathcal{R} \cap \mathcal{S}| + A^*} \quad \text{with} \quad \rho^*_{qw} \in [0, 1] \tag{17}$$

where

$$A^* = \sum_{x \in (\mathcal{S} \setminus \mathcal{R})} w(d(x, \mathcal{R})) + \sum_{x \in (\mathcal{R} \setminus \mathcal{S})} w(d(x, \mathcal{R}^*))$$

as new quantity can be defined. The three cases (a) to (c) yield analogue results for this quantity as for (13). All discussed quantities can also be used for the evaluation of voxel data segmentations (c.f. (Udupa et al., 2006) for medical image processing applications).

Evaluating the discussed pros and cons the *detection rate* ρ_d, the *false positive rate* ρ_{FP}, and the *false negative rate* ρ_{FN} are used as they can be easily interpreted. Furthermore the *weighted quality* ρ^*_{qw} is considered as it takes the distance weighted deviations $\mathcal{S} \setminus \mathcal{R}$ and $\mathcal{R} \setminus \mathcal{S}$ into account. Besides these quantities the distribution of distances can be analysed (mean, median, maximum). All of the above mentioned evaluation criteria can be applied for each reference segment, for all reference segments of class k and the entire set of reference segments. E.g. the *detection rate* computed for the entire set of reference segments relates to the *overall accuracy* used to evaluate the classification accuracy via confusion matrix.

All quantities evaluate just one aspect of segmentation properties at once. (Cardenes et al., 2007) propose to combine normed quantities via

$$\rho_g = \sqrt{\frac{1}{n}\sum_i^n \rho_i^2} \tag{18}$$

thus a simple mean of the squared sum. For illustration consider the quality rate ρ_q and the connectivity coefficient

$$\rho_{cc} = \frac{2\min(\sharp\mathcal{S}, \sharp\mathcal{R})}{\sharp\mathcal{S} + \sharp\mathcal{R}} \tag{19}$$

used by (Cardenes et al., 2007). In case of a squared segment $\mathcal{R}_{i[k]}$ and a set of 3×3 squared segments $\mathcal{S}_{i[k]}$ exactly overlapping $\mathcal{R}_{i[k]}$ $\rho_q = 1$, $\rho_{cc} = 0.1$ and $\rho_g = 0.07$. The segmentation is an example for oversegmentation, but the segments may still be appropriate for classification. At this step a listing and of the selected quantities is given and – with respect to the counter example above – a single combined quantity is not defined. Possibly based on the experiences made within the current work rule sets for the combination of quantities may be derived with respect to application schemes.

4 RESULTS

4.1 Data and experimental set up

For the presented investigations two different data sets - a mid-resolution and a high-resolution data set – are used. The mid-resolution data set consists of a subset of two Landsat scenes (the first taken in spring, the second in summer) of a rural area in the south-western part of Germany (Kaiserstuhl) with a ground sampling distance (GSD) of 30 m. Besides some larger forest and settlement areas most of the land use patterns consist of small parcels compared to the resolution of the Landsat sensor. The high-resolution data set is a subset of a QuickBird scene of Karlsruhe, thus an urban area with a GSD of about 2.4 m for the multispectral data. For both data sets reference polygons are compiled based on the satellite data. In our approach data given by a database may be used, but here digitisation of the reference polygons based on the input data for segmentation was preferred, because the focus is on the segmentation evaluation and discrepancies due to georeferencing are thereby avoided. The reference data for both data sets consist of polygons of selected classes, which can be easily identified and digitised. In case of the Landsat data (Fig. 1) polygons from the classes *forest* (dark green), *lake* (blue), *meadow* (light green), and *settlement* (red) and in case of the QuickBird data (Fig. 6) polygons from the classes *building* (red), *lawn* (light green) and *trees* (dark green) are used. For both data sets different segmentations with *Definiens Developer* based on different parameter settings were derived. The unclassified segments are used for the evaluation and assigned those segments which have an overlap of more than 50% of their area with the reference segments according to (5). In principle the classified segments can be assigned to the reference segments and the evaluation be based on those, but then the impact of the segmentation can not be separated from that of classification.

In the following the parameter settings for the *Definiens Developer* segmentation for each data set is shortly described. The parameter settings are not meant to be a recommendation, but are selected mainly in order to test the usefullness of the evaluation framework. Each section starts with a visual evaluation and continues with the quantitative evaluation.

4.2 Mid-resolution data: Landsat scene *Kaiserstuhl*

For this data set two multi-level segmentations were derived. Both use scale parameters from 5 to 30 with a step size of 5. For the first set *LA* the weight for *shape* was set to 0.2 for the first three segmentation levels *LA05 - LA15*, and to 0.0 for the higher segmentation levels *LA20 - LA30*, thus using only the spectral properties of the segments. The parameter *compactness* was set to 0.2 for scale parameter 5 *(LA05)*, to 0.5 for scale parameters 10

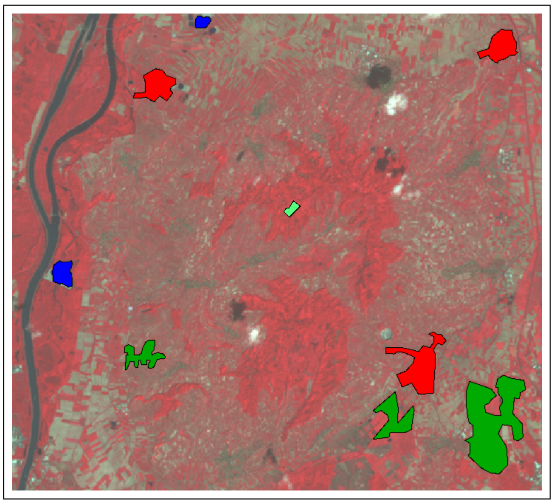

Figure 1: Landsat data Kaiserstuhl (432) and reference polygons

(LA10) and 15 *(LA15)*. For the second set *LB* the parameters *shape* and *compactness* were fixed to 0.3 and 0.5 for the segmentation levels *LB05 - LB15*. Again the parameter *shape* was set to 0.0 for the other levels *LB20 - LB30*.

Fig. 2 and Fig. 3 show selected levels of the two segmentation sets *LA* and *LB* for the large forest area in the south-eastern part of the data set. Some segmentation effects are clearly evident. The segmentations with scale parameter 5 lead to oversegmentation of the data with respect to the reference polygon. Although the settings for the parameters *shape* and *compactness* do not differ severely the segmentation set *LA* suffers from undersegmentation resulting in large segments which are the assigned to the reference segment via the used criterion in (5) compared to *LB*. Visually the higher levels of segmentation set *LB* with a parameter setting more emphasising the segment's geometry are preferable. Tab. 1 compiles the quantities for quality evaluation for this reference object. The discrepancies in segmentation results do not reveal in the detection rate ρ_d nor the false negative rate ρ_{FN}, because the most severe effect for the segmentations *LA20* to *LA30* is the undersegmentation and thus the false positives. Therefore these discrepancies are indicated by the rate of false positives ρ_{FP}, the quality rate ρ_q and even more by the weighted quality rate $\rho*_{qw}$. d_1 and d_2 denote the maximum distances from $\partial\mathcal{R}$ in the areas of false negatives $\mathcal{R} \setminus \mathcal{S}$ and false positives $\mathcal{S} \setminus \mathcal{R}$ showing the large deviations for this object. The results also constitute a counter example for using simple combined quality measures as discussed at the end of Section 3.3. Here the use of the combined quantity proposed by (Cardenes et al., 2007) using the weighted quality rate ρ^*_{qw} and the connectivity coefficient ρ_{cc} leads to $\rho_g = 0.52$ for *LA20*, $\rho_g = 0.74$ for *LA30*, $\rho_g = 0.64$ for *LB15* and $\rho_g = 0.93$ for *LB30*, thus in our opinion a false ranking with respect to the large area of $\mathcal{S} \setminus \mathcal{R}$.

The quality measures are not only computed for each reference object, but for all reference objects of each class and finally all reference objects. Tab. 2 and Tab. 3 compile these results for those segmentation levels also displayed in Figures 2 and 3 using the reference objects *forest* and all reference objects respectively as basis. Fig. 4 and Fig. 5 display examples for the segmentation of the entire data set. The weighted quality rate provides a similar ranking for the segmentations as discussed above. For the data set and with respect to the land use patterns a scale parameter of about 15 seems to be appropriate. This corresponds with previous investigations comparing pixel- and segment-based classification (Weidner and Bähr, 2007). The shown examples also indicate that

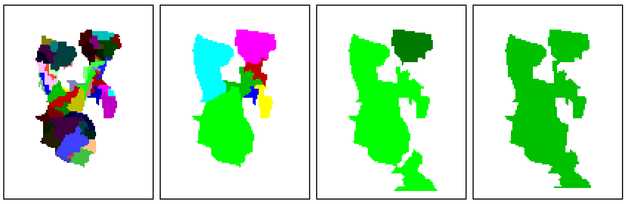

Figure 2: Segmentations *forest 02*: LA05[57], LA15[7], LA20[2], LA30[1] (from left to right with [] indicating number of segments)

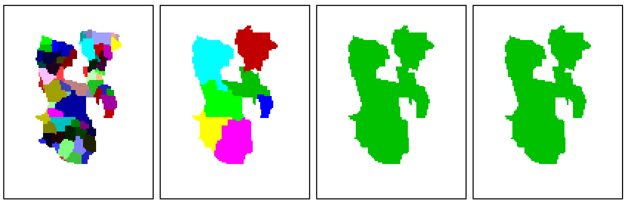

Figure 3: Segmentations *forest 02*: LB05[66], LB15[7], LB20[1], LB30[1] (from left to right with [] indicating number of segments)

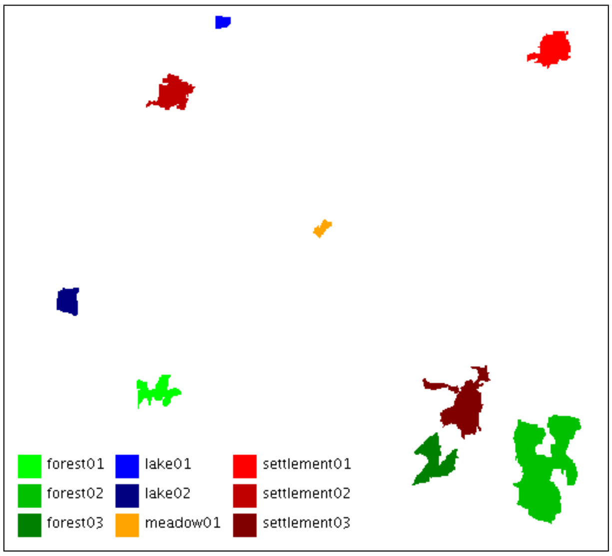

Figure 4: Segmentation LA05

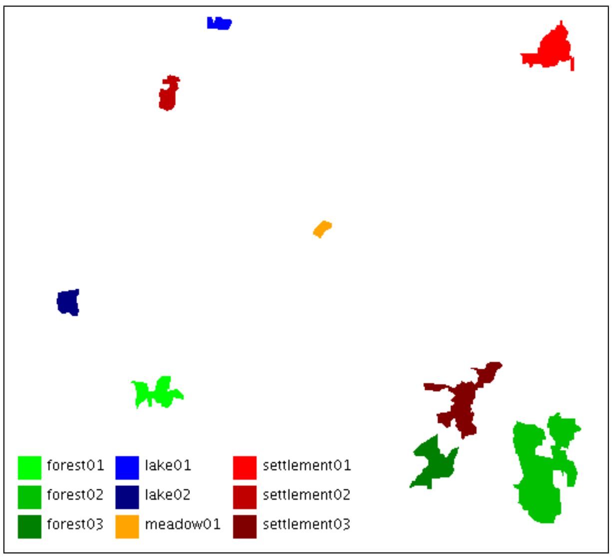

Figure 5: Segmentation LB30

	ρ_q	ρ^*_{qw}	ρ_d	ρ_{FN}	ρ_{FP}	d_1	d_2
LA05	0,92	0,91	0,97	0,03	0,05	72,4	102,4
LA10	0,91	0,89	0,97	0,04	0,06	114,9	102,4
LA15	0,89	0,86	0,96	0,04	0,08	114,9	162,4
LA20	0,78	0,31	0,93	0,07	0,19	180,0	933,4
LA25	0,77	0,31	0,94	0,07	0,21	180,0	933,4
LA30	0,77	0,31	0,94	0,07	0,21	180,0	933,4
LB05	0,92	0,91	0,97	0,03	0,05	72,4	102,4
LB10	0,91	0,88	0,95	0,05	0,05	180,0	102,4
LB15	0,90	0,87	0,96	0,05	0,06	180,0	102,4
LB20	0,90	0,87	0,96	0,04	0,06	180,0	102,4
LB25	0,90	0,87	0,96	0,04	0,06	180,0	102,4
LB30	0,90	0,87	0,96	0,04	0,06	180,0	84,9

Table 1: Landsat (Kaiserstuhl): Reference object *forest 02*

	ρ_q	ρ^*_{qw}	ρ_d	ρ_{FN}	ρ_{FP}	d_1	d_2
LA05	0,91	0,90	0,97	0,03	0,06	72,4	102,4
LA15	0,87	0,83	0,96	0,04	0,10	114,9	162,4
LA20	0,79	0,38	0,93	0,07	0,18	180,0	933,4
LA30	0,78	0,38	0,94	0,07	0,19	180,0	933,4
LB05	0,91	0,90	0,97	0,03	0,06	72,4	102,4
LB15	0,87	0,83	0,95	0,05	0,10	180,0	162,4
LB20	0,87	0,83	0,95	0,05	0,10	180,0	162,4
LB30	0,87	0,82	0,96	0,04	0,11	180,0	174,9

Table 2: Landsat (Kaiserstuhl): Reference objects *forest*

	ρ_q	ρ^*_{qw}	ρ_d	ρ_{FN}	ρ_{FP}	d_1	d_2
LA05	0,89	0,87	0,95	0,05	0,07	199,7	132,4
LA15	0,83	0,72	0,94	0,07	0,12	210,0	411,8
LA20	0,76	0,39	0,90	0,10	0,19	330,0	933,4
LA30	0,73	0,35	0,90	0,10	0,24	330,0	933,4
LB05	0,89	0,88	0,96	0,04	0,07	127,3	120,0
LB15	0,82	0,71	0,92	0,08	0,13	180,0	242,1
LB20	0,79	0,58	0,92	0,08	0,16	254,6	657.0
LB30	0,78	0,58	0,89	0,11	0,14	330,0	381,8

Table 3: Landsat (Kaiserstuhl): All reference objects

	ρ_q	ρ^*_{qw}	ρ_d	ρ_{FN}	ρ_{FP}	d_1	d_2
Q10	0,97	0,97	0,98	0,02	0,01	1,2	1,2
Q20	0,95	0,94	0,97	0,03	0,02	1,8	2,4
Q40	0,89	0,61	0,96	0,04	0,07	2,9	16,2

Table 4: QuickBird (Karlsruhe): Reference object *building 01*

	ρ_q	ρ^*_{qw}	ρ_d	ρ_{FN}	ρ_{FP}	d_1	d_2
Q10	0,95	0,94	0,98	0,03	0,03	2,3	2,3
Q20	0,92	0,9	0,96	0,04	0,04	3,6	3,15
Q40	0,86	0,72	0,95	0,05	0,10	9,0	16,2

Table 5: QuickBird (Karlsruhe): Reference objects *building*

	ρ_q	ρ^*_{qw}	ρ_d	ρ_{FN}	ρ_{FP}	d_1	d_2
Q10	0,95	0,94	0,97	0,03	0,03	2,3	2,3
Q20	0,92	0,90	0,96	0,04	0,04	4,0	3,5
Q40	0,86	0,64	0,94	0,06	0,09	9,0	34,5

Table 6: QuickBird (Karlsruhe): All reference objects

the criterion for the assignment of segments $\mathcal{S}_i$ to the reference segments may be improved. Up to now only the mean, median and maximum distances are computed. Although the comparison of these distances are already helpful, the distribution may contribute further valuable information in particular in the cases, where the evaluation quantities are computed for reference objects of a class or all reference objects.

4.3 High resolution data: QuickBird scene *Karlsruhe*

For the QuickBird data set three segmentations with scale parameters 10, 20, and 40 were generated. The parameter *compactness* was set to 0.5 for all levels, the parameter *shape* was fixed to 0.1 for the first levels *Q10* and *Q20* and to 0.3 for level *Q40*. As in the previous section we compiled the evaluation quantities for a single reference object (building with yard from center to northwest) in Tab. 4, all reference objects of class *building* in Tab. 5, and finally all reference objects in Tab. 6. Two segmentations are shown in Fig. 7 and Fig. 8. The results of the investigations with respect to the evaluation measures based on high-resolution remote sensing data confirm those of the previous section.

5 CONCLUSIONS

In this contribution evaluation approaches for image segmentation with respect to remote sensing data are reviewed. In our opinion the geometry and delineation of segments has more impact on the quality than other aspects like the number of segments. Oversegmentation is partly acceptable, undersegmentation is not, because it certainly leads to problems within the classification step of segment-based classification approaches. Therefore mainly geometric aspects of segmentation evaluation are considered. The evaluation problem is formulated as an evaluation of a binary classification leading to known quality measures like the quality rate. This measure was adapted by adding distance dependent weights. Nevertheless, one single measure is not able to cope with different requirements to segmentations. Therefore, a number of different measures is proposed and applied. Further work will aim at the verification and improvement of the approach, namely to consider a rule set analysing different quality measures dependent on the requirements of the application.

REFERENCES

Baatz, M. and Schäpe, A., 1999. Object-oriented and multi-scale image analysis in semantic networks. In: 2nd Intl. Symposium. Operationalization of Remote Sensing, 16 - 20 August, ITC, NL, Wichmann.

Cardenes, R., Bach, M., Chi, Y., Marras, I., de Luis, R., Anderson, M., Cashman, P. and Bultuelle, M., 2007. Multimodal evaluation for medical image segmentation. In: CAIP2007, LNCS, pp. 229 – 236.

Correira, P. and Pereira, F., 2003. Objective evaluation of video segmentation quality. IEEE Transactions on Image Processing 12(2), pp. 186 – 200.

Gelasca, E., Ebrahimi, T., Farias, M., Carli, M. and Mitra, S., 2004. Towards perceptually driven segmentation evaluation metrics. In: CVPR 2004 Workshop (Perceptual Organization in Computer Vision).

Hoover, A., Jean-Baptiste, G., Jiang, X., Flynn, P., Bunke, H., Goldgof, D., Bowyer, K., Eggert, D., Fitzgibbon, A. and Fisher, R., 1996. An experimental comparison of range image segmentation algorithms. IEEE Transactions on Pattern Analysis and Machine Intelligence 18(7), pp. 673–689.

Figure 6: QuickBird data Karlsruhe (432) and reference polygons

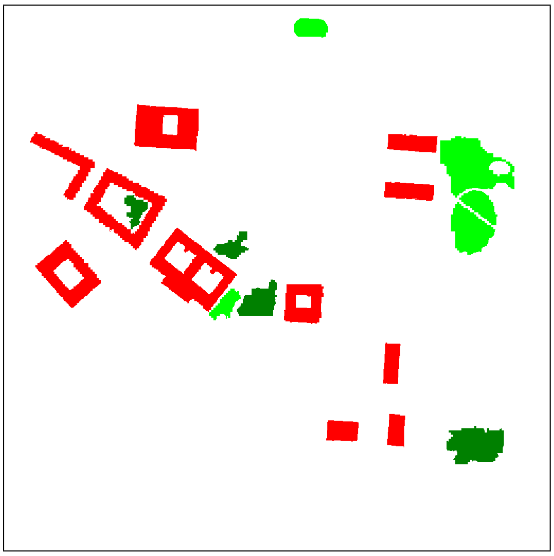

Figure 7: Segmentation Q10

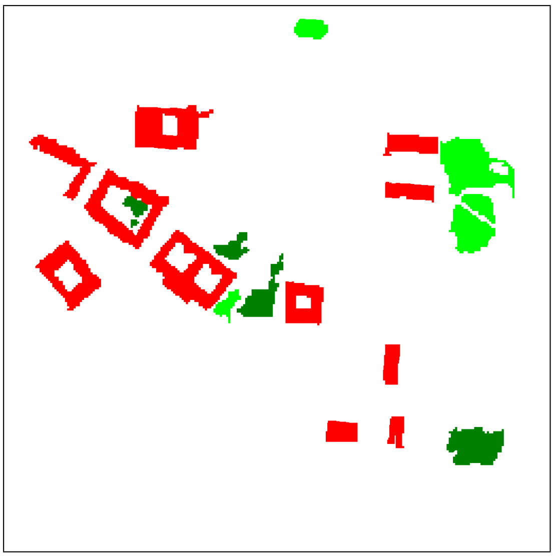

Figure 8: Segmentation Q40

Keim, D., 1999. Efficient geometry-based similarity search of 3d spatial databases. In: SIGMOD 99, Proceedings ACM SIGMOD International Conference on Management of Data, June 1-3, 1999, Philadelphia, PA, ACM Press, pp. 419–430.

Lemp, D. and Weidner, U., 2005. Improvements of roof surface classification using hyperspectral and laser scanning data. In: M. Möller and E. Wentz (eds), IAPRSIS, Vol. 36-8/W27: ISPRS Joint Conferences 3rd International Symposium Remote Sensing and Data Fusion Over Urban Areas (URBAN 2005) / 5th International Symposium Remote Sensing of Urban Areas (URS 2005).

Min, J., Powell, M. and Bowyer, K., 2004. Automated performance evaluation of range image segmentation algortihms. IEEE Transactions on Systems, Man, and Cybernetics 34(1), pp. 263 – 271.

Neubert, M., Herold, H. and Meinel, G., 2006. Evaluation of remote sensing image segmentation quality – further results and concepts. In: IAPRSIS, Vol. 26, Part 4/C42 - Conference on Object-based Image Analysis (OBIA 2006).

Paglieroni, D., 2004. Design considerations for image segmentation quality assessment measures. Pattern Recognition 37, pp. 1607 – 1617.

Radoux, J. and Defourny, P., 2006. Influence of image segmentation parameters on positional and spectral quality of the derived objects. In: IAPRSIS, Vol. 26, Part 4/C42 - Conference on Object-based Image Analysis (OBIA 2006).

Schiewe, J., 2002. Segmentation of high-resolution remotely sensed data – concepts, applications and problems. In: Symposium in Geospatial Theory, Processing and Applications, Ottawa.

Schuster, H.-F. and Weidner, U., 2003. A new approach towards quantitative quality evaluation of 3d building models. In: ISPRS Commission IV Joint Workshop *Challenges in Geospatial Analysis, Integration and Visualization II*, Stuttgart, pp. 156 – 163.

Srivastava, V., 2006. Evaluation of various segmentation tools for extraction of urban features using high resolution remote sensing data. In: ISPRS TC-IV Intl. Symposium on Geo Spatial Databases for Sustainable Development, Goa, India.

Udupa, J., LeBlanc, V., Zhuge, Y., Imielinska, C., Schmidt, H., Currie, L., Hirsch, B. and Woodburn, J., 2006. A framework for evaluating image segmentation algorithms. Computerized Medical Imaging and Graphics 20, pp. 75 – 87.

Unnikrishnan, R., Pantofaru, C. and Hebert, M., 2007. Toward objective evaluation of image segmentation algorithms. IEEE Transactions on Pattern Analysis and Machine Intelligence 29, pp. 929 – 944.

Van Droogenbroeck, M. and Barnich, O., 2005. Design of statistical measures for the assessment of image segmentation schemes. In: CAIP2005, LNCS, pp. 280 – 287.

Weidner, U. and Bähr, H.-P., 2007. Vergleich von pixel- und segementbasierter Klassifizierung am Beispiel des Kaiserstuhls. In: 27. Wissenschaftlich-Technische Jahrestagung der DGPF / Dreiländertagung SGPBF, DGPF und OVP, Muttenz, pp. 315 – 322.

Zhang, Y.-J., 2001. A review of recent evaluation methods for image segmentation. In: Intl. Symposium on Signal Processing and its Applications (ISSPA), 13 - 16 August 2001, Kuala Lumpur, Malaysia, pp. 148 – 151.