PFG 2014 / 2, 0091–0100
Stuttgart, April 2014

Article

Geometric Accuracy Assessment of Classified Land Use / Land Cover Changes

Markus Möller, Jens Birger & Cornelia Glässer, Halle (Saale)

Keywords: data quality, geometric accuracy, thematic accuracy, land use/land cover change, object based image analysis

Summary: European initiatives to harmonize geodata and the emergence of object-based image analysis techniques in remote sensing lead to increased demands regarding the quality assessment of thematic classification results. While common metrics for the thematic accuracy assessment have already been established for decades, there is a deficit in generally accepted geometric accuracy metrics enabling the assessment of two-dimensional thematic objects.

In this study, geometric accuracy metrics are introduced which base on differences in area and position between classified and reference objects. They are exemplarily calculated on classified and thematically verified agricultural fields in a German test site. We demonstrate how the metrics 1.) can be used for both the assessment of the total dataset and of single objects as well as 2.) can be coupled with thematic accuracy assessment results of a change detection analysis.

Zusammenfassung: *Genauigkeitsbewertung von klassifizierten Landnutzungs-/Landbedeckungsänderungen*. Europäische Initiativen zur Harmonisierung von Geodaten sowie das Aufkommen objektbasierter Bildanalysetechniken führen zu steigenden Anforderungen hinsichtlich der Qualitätsbewertung von thematischen Klassifikationsergebnissen. Während etablierte thematische Genauigkeitsmaße seit Jahrzehnten existieren, besteht ein Defizit an anerkannten geometrischen Genauigkeitsmaßen, die eine Bewertung von zweidimensionalen thematischen Objekten erlauben.

In der vorliegenden Studie werden geometrische Genauigkeitsmaße vorgestellt, die auf Differenzen zwischen Referenz- und klassifizierten Objekten hinsichtlich Position und Fläche basieren. Die Maße werden beispielhaft für klassifizierte und thematisch verifizierte Ackerschlagobjekte in einem deutschen Testgebiet berechnet. Wir zeigen, wie die Maße 1.) sowohl für die Bewertung ganzer Datensätze als auch einzelner Objekte verwendet werden können sowie 2.) mit den thematischen Validierungsergebnissen einer Veränderungsanalyse verknüpft werden können.

1 Introduction

In the context of the accuracy assessment of classified remote sensing data, commonly accepted thematic metrics like user's, producer's and overall accuracy have been established for decades. They are calculated from the well-known "error matrix" which aggregates point-related class assignments (Stehman 1997, Foody 2002, Zhan et al. 2005, Liu et al. 2007, Olofsson et al. 2013). The assessment of change detection accuracy results from the widely used binary "change/no change error matrix" (van Oort 2007).

The emergence of object-based image analysis (OBIA) techniques in remote sensing in recent years necessitates metrics for the geometric accuracy assessment of spatial objects which Blaschke (2010) has identified as one of the "hot" OBIA research topics. This is especially true for classified objects change detection techniques which are often used for the updating of maps or GIS layers (Chen et al. 2012, Hussain et al. 2013).

DOI: 10.1127/1432-8364/2014/0211 1432-8364/14/0211 $ 2.50

Geometric accuracy is related to the "problem of matching objects" (ZHAN et al. 2005). A complete matching can be assumed if there exists a one-to-one correspondence between reference and classified objects (CLINTON et al. 2010). Most common metrics result from the overlap of reference and classified objects (LUCIEER & STEIN 2002, ZHAN et al. 2005, MÖLLER et al. 2007, CLINTON et al. 2010, PERSELLO & BRUZZONE 2010, HERNANDO et al. 2012, SEBARI & HE 2013, MONTAGHI et al. 2013). Several authors have also compared the locations of objects's gravity centres (ZHAN et al. 2005, MÖLLER et al. 2007, CLINTON et al. 2010, WANG et al. 2010, SEBARI & HE 2013). Such metrics are strongly related to quality elements of the International Organization for Standardization (ISO) where – apart from the thematic accurracy – the positional accuracy is considered as most important for the sufficient accuracy assessment of thematic geodata. In this context, differences of gravity centres' locations can be seen as an indicator charaterizing two-dimensional spatial objects regarding their positional accuracy.

The combination of positional and areal metrics describing spatial extent and location enables an accurate geometric accuracy assessment of spatial objects (ZHAN et al. 2005). A combination of geometric metrics can be realized by arithmetic averaging, e.g. quadratic mean, whereas a normalized value range and value meaning is beneficially (CLINTON et al. 2010). Statistical clustering approaches allow an automatic structuring of a n-dimensional metrics' feature space independent of value ranges. For instance, MÖLLER et al. (2007) applied the k-means clustering algorithm on geometric metrics and used ranked cluster means to generate a comparison index which can be used to identify suitable metric combinations. However, PERSELLO & BRUZZONE (2010) have pointed out that a metrics' combination result can possibly lead to a lack of interpretability.

In this study, basic and combined geometric accuracy metrics are presented (sections 2.2 and 4.2) considering both requirements concerning ISO standards and OBIA. The metrics enable a local validation of single objects as well as an overall geometric validation of the entire geodata set. A precondition is the existence of independent reference objects (sections 2.3 and 3). The accuracy assessment procedure is exemplified on thematically tested and confirmed agricultural land use changes (sections 2.1 and 4.1) in a German test site (section 3).

The study was embedded in the national joint project DeCOVER 2 which has developed a methodological framework for the spatial and thematic updating of already existing land use data by analyzing multi-temporal RapidEye imagery (BUCK 2010). The RapidEye sensors belong to a new generation of small earth observation systems for the better observation of dynamic phenomena (SANDAU et al. 2010) representing a high spatial, temporal and up-to-date availability (TYC et al. 2005).

2 Methods

2.1 *Assessment of Change Detection Accuracy*

The "change/no change error matrix" distinguishes four error types whereas "true positive" and "true negative" stand for correctly classified changes and no-changes (Tab. 1). In contrast, the error types "false positive" and "false negative" characterize misclassifications representing errors of commission and ommission (BOSCHETTI et al. 2004).

Tab. 1: Change/no-change error matrix.

		reference	
		change	no change
class	change	**true positive**	false negative
	no change	true negative	**false positive**

2.2 *Geometric Accuracy Metrics*

In Fig. 1, object F overlaps object T. The intersection I of F and T corresponds to the Boolean AND operation according to (1). The "relative area metrics" RA_F and RA_T shown in (2) and (3) arise from the ratio of the object areas A_I and A_F or A_T.

$$I = T \cap F \quad (1)$$

$$RA_F = \frac{A_I}{A_F} \quad (2)$$

$$RA_T = \frac{A_I}{A_T} \tag{3}$$

Positional differences of two-dimensional objects can be expressed by the distances between objects' gravity centres. The normalized distances between the gravity centres of I and F ($dist$ (C_I, C_F)) or T ($dist$ (C_I, C_T)) are referred to as the "relative positional metrics" RP_F and RP_T ((4) and (5)). As normalization factor, the distance between C_I and the farthermost gravity centres $C_{F*,max}$ or $C_{T*,max}$ within the extent of F or T is used ($dist$ (C_I, $C_{F*,max}$)) or ($dist$ (C_I, $C_{T*,max}$)). F^* and T^* result from Boolean NOT operations according to (6) and (7). In Fig. 1b and c, three F^* and T^* complements arise in each case whereas the gravity centres between I and $F^{*,max}$ or $T^{*,max}$ are most distant.

$$RP_F = 1 - \frac{dist(C_I, C_F)}{dist(C_I, C_{F^*,max})} \tag{4}$$

$$RP_T = 1 - \frac{dist(C_I, C_T)}{dist(C_I, C_{T^*,max})} \tag{5}$$

$$F^* = \frac{F}{T} \tag{6}$$

$$T^* = \frac{T}{F} \tag{7}$$

The basic metrics M (RA_F, RA_T, RP_F and RP_T) represent a value range between 0 and 1. While the value 0 stands for no match, the value 1 indicates a complete correspondence of I with F or T regarding area or position. The normalized value ranges enable a combination which is realized by using the geometric mean according to (8). In doing so, small values are emphasized which reflects the fact that geometric quality is mainly determined by minimal overlaps or high positional distances of classified and reference objects.

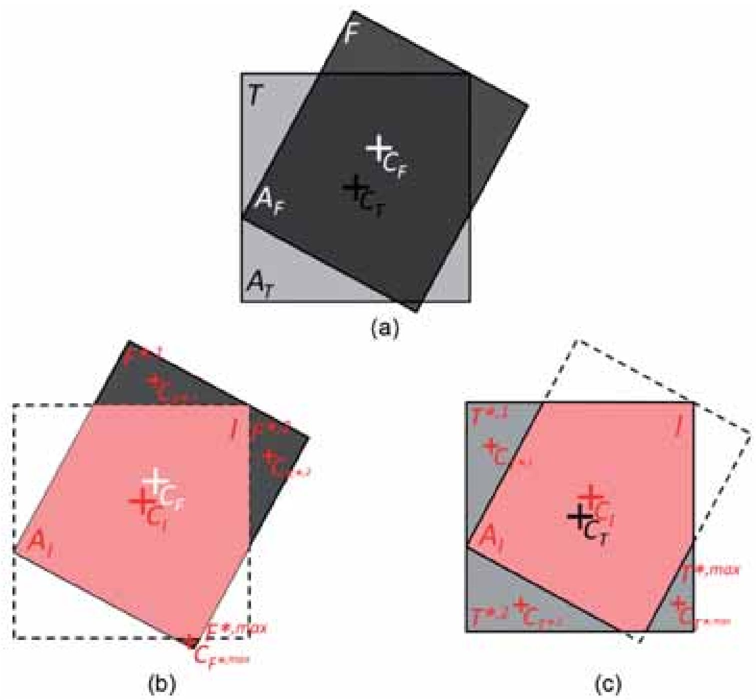

Fig. 1: Visualization of object features for the calculation of geometric accuracy metrics: (a) overlap of F and T, (b) comparison of I and F, (c) comparison of I and T (F – reference object | T – classified of to be tested object | I – intersection of F and T | F^*, T^* – relative complement of F or T and I | $C_{(T,F,I)}$ – gravity centres of T, F or I | $C_{F*,max}$, $C_{T*,max}$ – farthermost gravity centre from C_I within the extent of F or T).

The resulting combined metrics C are listed in Tab. 2. The combination of RA_F and RA_T as well as of RP_F and RP_T results in the metrics RA and RP which show the degree of aggregated positional and areal differences between reference and classified objects. Their mean is considered as "overall geometric accuracy" (OGA). The mean of RA_T and RP_T is referred to as "test object-related geometric accuracy" (TGA). The metric shows whether a classified object is under- ($OGA < TGA$) or over-sized ($OGA > TGA$).

$$
\begin{aligned}
&C = \left(\prod_{i=1}^{n} M_i \right)^{\frac{1}{n}} \\
&C \in \left\{ RA, RP, OGA, TGA \right\} \\
&M_i \in \left\{ RA_F, RA_T, RP_F, RP_T \right\} \\
&0 \le C, M_i \le 1
\end{aligned}
\tag{8}
$$

Tab. 2: Combination of basic geometric accuracy metrics M_i to combined metrics C according to (8).

	C	M_i
RA	combined relative area	RA_F, RA_T
RP	combined relative position	RP_F, RP_T
OGA	overall geometric accuracy	RA_F, RA_T, RP_F, RP_T
TGA	test object-related geometric accuracy	RA_T, RP_T

2.3 *Sampling of Reference Objects*

The samples used for the thematic validation also mark the locations of reference objects. In this study, independent interpreters digitized manually such objects on-screen on the basis of digital aerial photographs.

3 Study Site and Input Data

The geometric accuracy assessment procedure is demonstrated using a classification which was carried out in the study site Bitterfeld/Wolfen (Fig. 2a). The study site is one of five DeCOVER 2 project test sites, is situated in the south of the German Federal State of Saxony-Anhalt and covers 263 km². Due to the high land use dynamic caused by urban development and land use conversions, agricultural parcels of the *Land Parcel Identification System* (LPIS, Inan et al. 2010) have been geometrically and thematically updated in the course of a DeCOVER 2 test production by Buck et al. (2011). They have analyzed bi-temporal RapidEye scenes from 2010 (16.7. and 21.8.) which have been segmented and classified within the software environment eCognition (Trimble 2012) using the region growing segmentation algorithm Fractal Net Evolution Approach (FNEA, Benz et al. 2004) and applying an object-based and scale-specific change detection approach.

Publicly available digital aerial photographs from 2010 with a resolution of 0.4 m × 0.4 m have been used to create a reference data base by digitizing reference objects. Their locations are spatially associated with all samples representing confirmed classified geometric changes (Fig. 2b).

4 Results

4.1 *Change Detection Accuracy*

The change detection results are visualized in Fig. 2a. Accordingly, 567 field objects had to be tested which corresponds to an area of 5033 ha or 20 % of the total study area. In doing so, 148 objects (26 %) or 379 ha (7.5 %) were identified as changes. Using stratified random samples, a thematic accuracy assessment of changed and un-changed agricultural objects was carried out. While the error for the classification of un-changed objects is negligible, the accuracy of changed objects is 70 % (Tab. 3).

Tab. 3: Change detection accuracy results: number and proportion (%) of changed and unchanged agricultural land use objects.

		reference	
		change	no change
class	change	**72** (70%)	31 (30%)
	no change	1 (1%)	119 (99%)

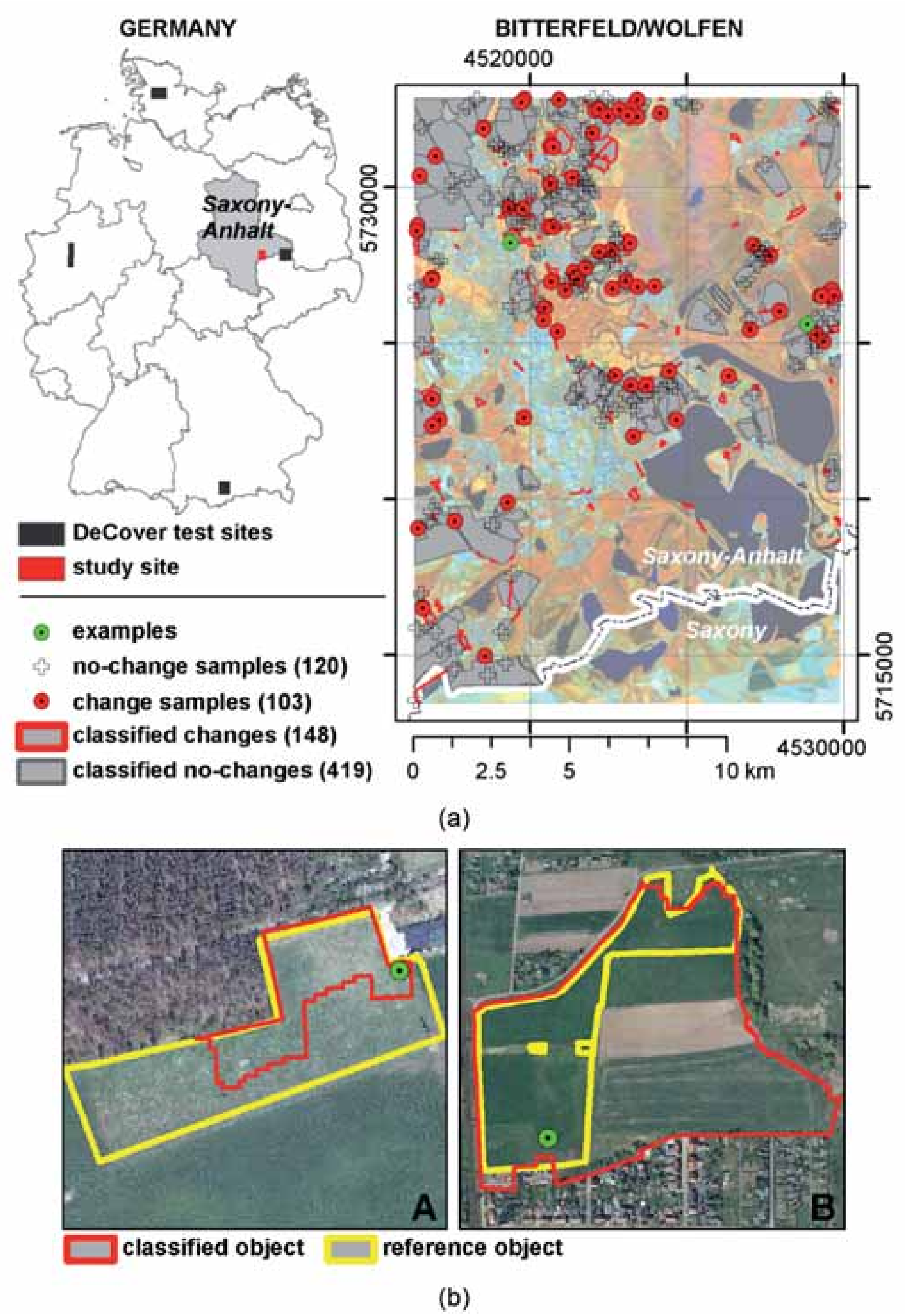

Fig. 2: (a) Test site location within Germany and Saxony-Anhalt and a RapidEye image from 21th August 2010 overlaid with the used datasets as well as (b) digital aerial photographs from 2010 overlaid with two examples of classified and reference objects.

4.2 *Geometric Accuracy Assessment*

The geometric accuracy assessment is applied to the 72 confirmed samples in the sense of an in-depth assessment of thematically verified object changes (see Tab. 3). The basic geometric metrics were calculated using the software package eCognition (Trimble 2012). The metrics' visualization was realized within the statistical environment R (R Core Team 2012).

The accuracy assessment is spatially related to the intersections of classified and reference objects (see Figs. 1b and c). An intersection operation leads to objects of different sizes. The determination of object size thresholds, which are considered as meaningful geometric changes, depends on user-specific needs like minimum mapping units (MMU) and affects the geometric accuracy results. This is illustrated in Fig. 3a where the relation between thresholds of object sizes as well as corresponding object numbers (*N*) and the overall accuracy metric *OGA* is shown. The median of the *OGA* distribution is used for the global assessment of the entire dataset. In this study, only object changes greater than 0.5 ha are considered which corresponds to the MMU within the DeCOVER 2 project. This is true for 57 objects and results in an *OGA* median of 0.75 (Fig. 3b). The *OGA* boxplot illustrates the broad value distribution which is characterized by the 25th percentile (Q1). Here, the Q1 value is 0.43. The relation of *RA* and *RP* medians clarifies which error type is domi-

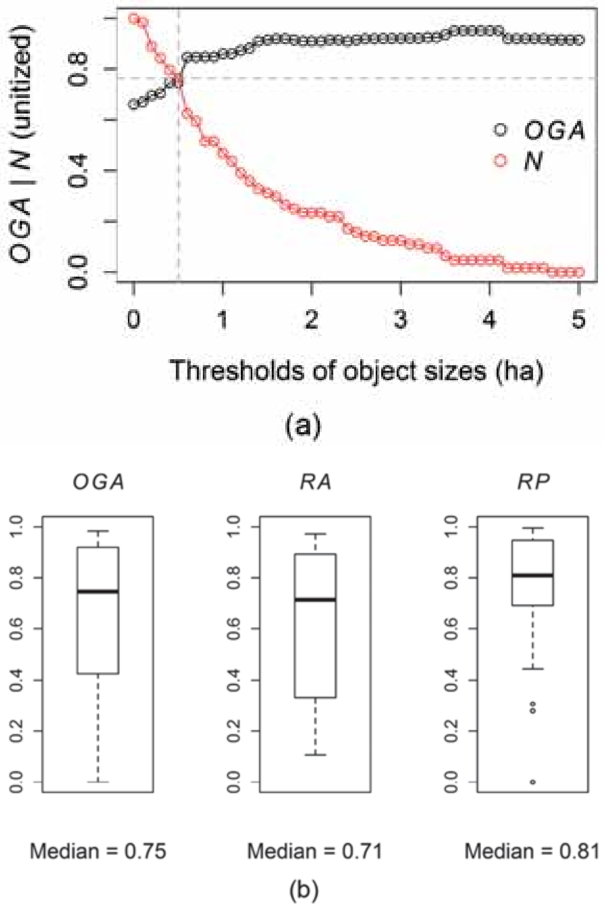

Fig. 3: Global geometric accuracy assessment results: (a) Relation between thresholds of object change sizes, overall geometric accuracies (*OGA*) and object numbers (*N*). (b) *RA*, *RP* and *OGA* boxplots for all 57 objects with an object size greater 0.5 ha.

Tab. 4: Local geometric accuracy results of the example objects A and B (see Fig. 2b).

	RA_R	RA_T	RP_R	RP_T	RA	RP	OGA	TGA
A	0.35	0.95	0.63	0.98	0.57	0.79	0.67	0.96
B	0.97	0.40	0.98	0.67	0.62	0.81	0.71	0.52

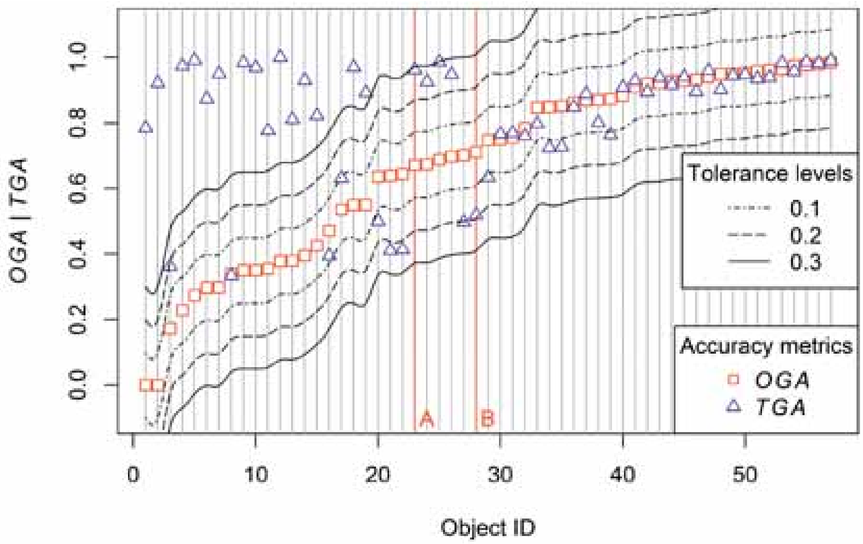

Fig. 4: Sorted *OGA* values of all 57 object changes greater than 0.5 ha (see Fig. 3b). The difference of an *OGA* and corresponding *TGA* value indicates whether a classified object is under-sized or over-sized. The classified object "A" is under-sized or smaller than the corresponding reference object which leads to a negative difference of -0.29. In contrast, the classified object "B" is over-sized or greater than the corresponding reference object. Thus, the resulting difference of 0.19 is positive (see Fig. 2b and Tab. 4).

nant. Here, the *RA* median (0.71) is smaller than the *RP* median (0.81). This means that the geometric inaccuracies are mainly caused by overlapping mismatches.

The local accuracy assessment is related to single objects and is based on object-specific *OGA* and *TGA* values. In Fig. 4, the sorted *OGA* values of all 57 tested objects are visualized. The related positions of the corresponding *TGA* values indicate the degree of under- or over-sizing. The red-emphasized vertical lines exemplify the accuracy assessment results of the example objects "A" and "B" (see Fig. 2). Accordingly, the red- and yellow-framed objects display classified (*T*) and reference objects (*F*). The corresponding accuracy metrics are listed in Tab. 4. While the *OGA*, *RA* and *RP* values are on the same level, the *TGA* values are different. In example "A", *T* is under-sized or smaller than *F* which is indicated by a negative difference of *OGA* and *TGA* ($0.67 - 0.96 = -0.29$). The opposite case of over-sizing shows example "B" where the classified object is characterized by a positive difference ($0.71 - 0.52 = 0.19$). The consideration of all 57 objects reveals that 33 objects (58 %) are under-sized.

The absolute differences of global *OGA* medians and 25th percentiles as well as of local *OGA* and *TGA* values reflect the global and local variation of geometric accuracy. User-defined tolerance levels could define which degree of variation is still acceptable. In Fig. 4, three examples of tolerance levels are shown exemplarily.

5 Discussion and Conclusion

European initiatives to harmonize geodata and the emergence of object-based image analysis (OBIA) techniques in remote sensing result in increased demands regarding the quality assessment of thematic classification results. Thus, there is an urgent need for the definition of commonly accepted protocols for the accuracy assessment of two-dimensional classified objects. Such protocols should include 1.) the definition of additional geometric accuracy metrics, 2.) their coupling with existing thematic accuracy metrics and 3.) adapted sampling strategies:

1. In this study, metrics for the geometric validation of single objects and land use classes are presented considering both requirements concerning ISO standards and OBIA. The metrics' calculation corresponds to a hierarchical comparison of reference and classified objects regarding differences of objects' areas and positions (gravity centres). The resulting basic metrics were combined to aggregated metrics which enable the identification of geometric inaccuracies of both single objects and entire thematic geodata sets. The used metrics have been derived in the same or a similar manner by other authors (section 1). This means that different metric's variants for the characterization of specific object properties already exist.
2. The introduced geometric accuracy metrics have been exemplarily calculated for classified and thematically verified land use changes. They have been detected within a DeCOVER 2 process chain which aims at the spatial and thematic updating of existing land use datasets. Quality assurance is an integral part of the classification workflow whereas geometric accuracy assessment is considered as an in-depth assessment of the thematic validation. A similar approach of coupling thematic and geometric metrics has been used by Zhan et al. (2005) and Persello & Bruzzone (2010). They have applied thematic and geometric accuracy indices for the global quality assessment of classification results and for the optimization of classification errors. Hernando et al. (2012) have gone one step further and coupled categories of thematic coincidence levels with geometric overlapping accuracy metrics in a so called "Object Fate Analysis" (OFA) matrix which characterizes the thematic and spatial agreement between classified and reference objects.
3. The coupling of thematic and geometric accuracy assessment affects the sampling strategy of reference objects. In this study, the sampling was restricted to the change detection category "true positive". The full integration of geometric accuracy assessment into a validation process of thematic classifications would entail a higher sampling effort. Thus, Radoux et al. (2011) introduced an object-based and statistically sound sampling strategy which noticeably reduces the sampling effort.

Acknowledgements

The project is funded by the German Ministry of Economics and Technology, managed by the German Aerospace Center (contract no.: FKZ 50EE0915). We would like to thank Anthony Gidudu (Department of Geomatics and Land Management, Makerere University, Uganda) and Lydia Nakayenze (Fraunhofer Institute for Cell Therapy and Immunology, Department of Vascular Biology) for helpful comments and linguistic advice.

References

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. – ISPRS Journal of Photogrammetry and Remote Sensing **58** (3–4): 239–258.

Blaschke, T., 2010: Object based image analysis for remote sensing. – ISPRS Journal of Photogrammetry and Remote Sensing **65** (1): 2–16.

Boschetti, L., Flasse, S.P. & Brivio, P.A., 2004: Analysis of the conflict between omission and commission in low spatial resolution dichotomic thematic products: The Pareto Boundary. – Remote Sensing of Environment **91** (3–4): 280–292.

Buck, O., 2010: DeCOVER 2 – The German GMES extension to support land cover data systems:

Status and outlook. – LACOSTE-FRANCIS, H. (ed.): ESA Living Planet Symposium. – ESA Special Publication **SP-686**, Bergen, Norway, on CD-ROM.

BUCK, O., PETER, B. & BÜKER, C., 2011: Zwei-skaliger Ansatz zu Aktualisierung von landwirtschaftlichen Referenzkulissen (LPIS). – PFG – Photogrammetrie, Fernerkundung, Geoinformation **2011/5:** 339–348.

CHEN, G., HAY, G.J., CARVALHO, L.M.T. & WULDER, M.A., 2012: Object-based change detection. – International Journal of Remote Sensing **33** (14): 4434–4457.

CLINTON, N., HOLT, A. & GONG, P., 2010: Accuracy assessment measures for object-based image segmentation goodness. – Photogrammetric Engineering & Remote Sensing **76** (3): 289–299.

FOODY, G., 2002: Status of land cover classification accuracy assessment. – Remote Sensing of Environment **80** (1): 185–201.

HERNANDO, A., TIEDE, D., ALBRECHT, F. & LANG, S., 2012: Spatial and thematic assessment of object-based forest stand delineation using an OFA-matrix. – International Journal of Applied Earth Observation and Geoinformation **19:** 214–225.

HUSSAIN, M., CHEN, D., CHENG, A., WEI, H. & STANLEY, D., 2013: Change detection from remotely sensed images: From pixel-based to object-based approaches. – ISPRS Journal of Photogrammetry and Remote Sensing **80:** 91–106.

INAN, H.I., SAGRIS, V., DEVOS, W., MILENOV, P., VAN OOSTEROM, P. & ZEVENBERGEN, J., 2010: Data model for the collaboration between land administration systems and agricultural land parcel identification systems. – Journal of Environmental Management **91:** 2440–2454.

LIU, C., FRAZIER, P. & KUMAR, L., 2007: Comparative assessment of the measures of thematic classification accuracy. – Remote Sensing of Environment **107** (4): 606–616.

LUCIEER, A. & STEIN, A., 2002: Existential uncertainty of spatial objects segmented from satellite sensor imagery. – IEEE Transactions on Geoscience and Remote Sensing **40** (11): 2518–2521.

MÖLLER, M., LYMBURNER, L. & VOLK, M., 2007: The comparison index: A tool for assessing the accuracy of image segmentation. – International Journal of Applied Earth Observation and Geoinformation **9** (3): 311–321.

MONTAGHI, A., LARSEN, R. & GREVE, M.H., 2013: Accuracy assessment measures for image segmentation goodness of the Land Parcel Identification System (LPIS) in Denmark. – Remote Sensing Letters **4** (10): 946–955.

OLOFSSON, P., FOODY, G.M., STEHMAN, S.V. & WOODCOCK, C.E., 2013: Making better use of accuracy data in land change studies: estimating accuracy and area and quantifying uncertainty using stratified estimation. – Remote Sensing of Environment **129:** 122–131.

PERSELLO, C. & BRUZZONE, L., 2010: A novel protocol for accuracy assessment in classification of very high resolution images. – IEEE Transactions on Geoscience and Remote Sensing **48** (3): 1232–1244.

RADOUX, J., BOGAERT, P., FASBENDER, D. & DEFOURNY, P., 2011: Thematic accuracy assessment of geographic object-based image classification. – International Journal of Geographical Information Science **25** (6): 895–911.

R CORE TEAM, 2012: R: A Language and Environment for Statistical Computing. – R Foundation for Statistical Computing, Vienna, Austria.

SANDAU, R., BRIESS, K. & D'ERRICO, M., 2010: Small satellites for global coverage: Potential and limits. – ISPRS Journal of Photogrammetry and Remote Sensing **65** (6): 492–504.

SEBARI, I. & HE, D.C., 2013: Automatic Fuzzy Object-Based Analysis of VHSR Images for Urban Objects Extraction. – ISPRS Journal of Photogrammetry and Remote Sensing **79:** 171–184.

STEHMAN, S., 1997: Selecting and interpreting measures of thematic classification accuracy. – Remote Sensing of Environment **62** (1): 77–89.

TYC, G., TULIP, J., SCHULTEN, D., KRISCHKE, M. & OXFORT, M., 2005: The RapidEye mission design. – Acta Astronautica **56** (1–2): 213–219.

TRIMBLE, 2012: eCognition Developer 8.8 reference book. – Trimble Germany GmbH, München.

VAN OORT, P.A.J., 2007: Interpreting the change detection error matrix. – Remote Sensing of Environment **108** (1): 1–8.

WANG, Z., JENSEN, J.T. & IM, J., 2010: An automatic region-based image segmentation algorithm for remote sensing applications. – Environmental Modelling & Software **25:** 1149–1165.

ZHAN, Q., MOLENAAR, M., TEMPFLI, K. & SHI, W., 2005: Quality assessment for geo-spatial objects derived from remotely sensed data. – International Journal of Remote Sensing **26** (14): 2953–2974.

Addresses of the Authors:

Dr. MARKUS MÖLLER & Prof. Dr. CORNELIA GLÄSSER, Martin-Luther-Universität Halle-Wittenberg, Institut für Geowissenschaften und Geographie, Fachgebiet Geofernerkundung und Kartographie, Von-Seckendorff-Platz 4, 06120 Halle (Saale), Tel.: +49-345-552-6026, Fax: +49-345-552-7168, e-mail: {markus.moeller}{cornelia.glaesser}@geo.uni-halle.de

Dr. Jens Birger, UMGEODAT – Umwelt- und GeodatenManagement GbR, Mansfelder Str. 56, 06108 Halle (Saale), Tel. +49-345-6141-8481, e-mail: jens.birger@umgeodat.de

Manuskript eingereicht: September 2013
Angenommen: November 2013