IEEE International Geoscience and Remote Sensing Symposium (IGARSS), Quebec City, Canada, 2014. This is a final submitted version. CLICK for the final version.

QUALITY ASSESSMENT OF BUILDING EXTRACTION FROM REMOTE SENSING IMAGERY

Janja Avbelj and Rupert Müller

German Aerospace Center, Earth Observation Center, Oberpfaffenhofen, 82234 Wessling, Germany

ABSTRACT

An automatic quality assessment of extracted buildings from remote sensing imagery is needed to evaluate extraction algorithms, or to support change detection. In this paper, four commonly used measures are compared to the newly proposed metric for comparison of polygons and line segments (PoLiS). The extracted polygons are compared to the reference polygons and the quality measures are computed for each pair. The symmetric measures, i.e. quality rate and PoLiS, estimate overall dissimilarity between polygons, whereas i.e. the root mean square error (RMSE) of the distances between the polygon vertices, completeness, and correctness, are not symmetric and should be therefore used for applications like change detection. The variability of the measures is assessed according to the area of the reference buildings. The variability is higher for the category of larger buildings, where the building polygon complexity is larger.

***Index Terms*—** Building extraction, Polygon comparison, Quality assessment, Shape similarity

1. INTRODUCTION

Information about buildings is needed by governmental and private organisations for urban planning, 3D visualization predicting micro climate, updating cadaster data, and real estate databases to name a few. Remote sensing imagery and methods enable semi- automatic extraction and modeling of buildings, and thus for some applications terrestrial data collection is not needed any more. However, the performance and quality of the building extraction methods must be evaluated with regard to ground truth (reference) data. In the recent years, the lack of standard evaluation techniques for building extraction has been addressed by many authors, and new indices or evaluation systems are being proposed [1–6].

The quality of extraction can be evaluated on a per pixel, per object or per scene level basis [5]. Per scene assessments quantify an overall quality of extraction. However, all buildings are not extracted with the same quality, so the per object level evaluation is also needed. For instance, given two images acquired at different times, from which the buildings are reliably extracted. Then a per object quality measure can support change detection, e.g. demolition or illegal building (part) construction. In per object evaluation either a quality measure to each extracted object is assigned, or the object is labelled as correctly or not-correctly extracted e.g. [6]. In this contribution we focus on the former per object evaluation of 2D building polygons (vector data) and compare four commonly used quality measures to the recently proposed alternative approach for comparison of polygons [7].

E-mail: (janja.avbelj, rupert.mueller)@dlr.de see http://www.dlr.de/eoc/

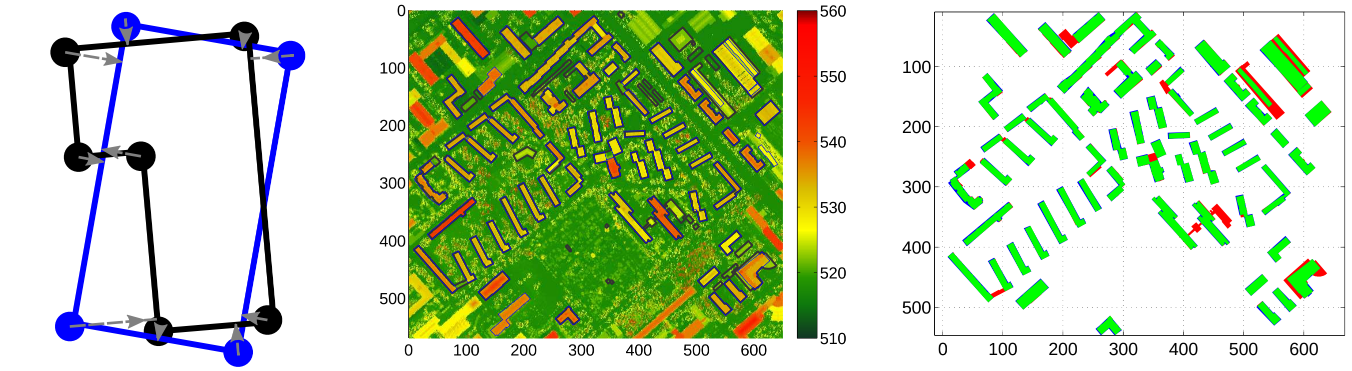
2. MEASURES FOR PER OBJECT BUILDING EXTRACTION EVALUATION

The four measures are defined in this Sec., these are completeness, correctness, quality rate, and root mean square error of the distances between the polygon vertices. An extracted polygon A and the reference polygon B have vertices $\mathbf{a}_i \in A$, $i = 1, \ldots, q+1$ and $\mathbf{b}_j \in B$, $j = 1, \ldots, r+1$, respectively. They are both closed polygons, the first and the last vertex of a polygon coincide, e.g. $\mathbf{a}_1 = \mathbf{a}_q + 1$. The term polygon is used for any closed, piecewise linear, and not self-intersecting polygon, which can have holes. The Euclidean distance between $\mathbf{a}_i$-th and $\mathbf{b}_j$-th vertex of the polygons is denoted $\|\mathbf{a}_i - \mathbf{b}_j\|$, $\mathbf{a}_i, \mathbf{b}_j \in \mathbb{R}^2$. A boundary ∂A of the closed polygon A consists of $q+1$ vertices $\mathbf{a}_j$, q edges, and points that lie on the boundary. Any point of a polygon, which has defined coordinates, is refereed to as a vertex, even if it is not a corner point of the polygon, but lies on the polygons' boundary.

2.1. Matched rates: completeness, correctness, and quality rate

The matched rates are based on the areas and the intersection area between the extracted and the reference polygon. The areas of the extracted polygon A and reference polygon B are denoted $ar(A)$ and $ar(B)$, respectively. The following detected areas are used for computation of the matched rates e.g. [5] as defined in Tab. 1
- true positive $ar(TP) = A \cap B$ (Fig. 1c, green),
- false negative $ar(FN) = ar(B) - A \cap B$ (Fig. 1c, red), and
- false positive $ar(FP) = ar(A) - A \cap B$ (Fig. 1c, blue).

IEEE International Geoscience and Remote Sensing Symposium (IGARSS), Quebec City, Canada, 2014. This is a final submitted version. CLICK for the final version.



(a) PoLiS principle computation (b) LiDAR DSM with building polygons (c) Comparison of the polygons

Fig. 1: The PoLiS metric computation principle (a) [7]. The normalized sum of the minimal Euclidean distances (gray lines) between the extracted A (blue) and the reference polygon B (black) and vice versa is computed. The grey arrows point into direction of the computation of the distance. Comparison of extracted building polygons (b, blue lines) from LiDAR DSM (b, the elevation is colour coded and is in [m]) and reference building polygons (c, black lines). On Fig. (c) are the TP (c, green), FN (c, red), and FP (c, blue) detected areas.

completeness	$comp =$	$\frac{ar(TP)}{ar(TP)+ar(FN)}$
correctness	$corr =$	$\frac{ar(TP)}{ar(TP)+ar(FP)}$
quality rate	$qual =$	$\frac{ar(TP)}{ar(TP)+ar(FP)+ar(FN)}$
		$\frac{comp\ corr}{comp+corr-comp\ corr}$

Table 1: Matched rates

Completeness is also named producer's accuracy, detection or true positive rate [1, 2], whereas correctness is also refereed to as user's accuracy [2] (Tab. 1). The quality rate is a symmetric rate and can be computed as a combination of $comp$ and $corr$. The codomain of the defined matched rates in Tab. 1 is on an interval $[0, 1]$, where 0 is low and 1 is high value of the matched rate, which are dimensionless quantities.

2.2. RMSE of the distances between polygons' vertices

The RMSE of the Euclidean distance between the vertices of the extracted (A) and the reference (B) polygon denoted $RMSE_{A,B}$ is computed by

$$RMSE_{A,B} = q^{-\frac{1}{2}} \sqrt{\sum_{j=1}^{q} (\min_{\mathbf{b} \in B} \|\mathbf{a}_j - \mathbf{b}\|)^2}, \qquad (1)$$

where $\mathbf{b} \in B$ is the closest vertex to the vertex $\mathbf{a}_i \in A$. The $RMSE_{B,A}$ is computed analogically to the computation in eq. (1). Note that in general $RMSE_{A,B} \neq RMSE_{B,A}$. [8] takes the distance to the nearest point of the polygon B ($\mathbf{b} \in \partial B$) and not the nearest vertex of the polygon B ($\mathbf{b} \in B$) for the computation of the $RMSE_{A,B}$. Moreover, they exclude the distances $\|\mathbf{a}_j - \mathbf{b}\|$, $\mathbf{a}_j \in A$, $\mathbf{b} \in B$ (or $\mathbf{b} \in \partial B$) exceeding a predefined threshold.

3. POLIS, A METRIC FOR COMPARISON OF POINTS AND LINE SEGMENTS

The PoLiS measure proposed by [7] is an alternative approach for per object building extraction assessment, which quantifies similarity between two polygons. It is a metric in mathematical sense, i.e. is a positive definite and symmetric function that satisfies a triangle inequality. It accounts for vertices and edges of the polygons, similarly like the RMSE defined in [8], but uses no threshold. PoLiS metric between two polygons changes approximately linearly with respect to small translation, rotation, and scale changes between extracted and reference building polygon. This metric is insensitive to the additional points on polygons‘ edges. However, it underestimates the actual dissimilarity if one of the polygons has much larger number of vertices than the other. The PoLiS metric $p(A, B)$ between polygons A and B (Fig. 1a) is defined by

$$p(A, B) = \frac{1}{2q} \sum_{i=1}^{q} \min_{\mathbf{b} \in \partial B} \|\mathbf{a}_i - \mathbf{b}\| + \frac{1}{2r} \sum_{j=1}^{r} \min_{\mathbf{a} \in \partial A} \|\mathbf{b}_j - \mathbf{a}\|.$$

PoLiS metric quantifies average overall dissimilarity per point, and has the same units as the vertices of the polygons.

4. COMPARISON OF THE MEASURES

4.1. Data description

The building polygons (Fig 1b, blue) are extracted from the LiDAR digital surface model (DSM) with 1 m spatial resolution (Fig 1b), which is resampled from a LiDAR point cloud with an average density of 1.69 points/m^2. The method for building polygon extraction is described in [9]. The reference building polygons were provided from the Municipal of Munich and are highly accurate cadaster data (Fig 1b, black). Only the extracted polygons, which have a corresponding reference polygon, are considered for the comparison of the measures.

The polygons on the boundary of the LiDAR DSM (Fig 1b) were removed because they are only partially extracted. The correspondence between 57 extracted and 73 reference building polygons is determined. First, the adjacent extracted polygons (these are polygons sharing at least one point) are merged. Second, every reference polygon is assigned to at least one extracted polygon, if the overlap is at least 10% according to the extracted or the reference polygon This threshold is proposed by [2]. For this example (Fig 1b, c), every extracted polygon correspond to either none or one reference polygon. If the reference polygon is assigned than one, the measures are computed for each extracted polygon separately and added as a weighted average according to the areas of the extracted corresponding polygons.

4.2. Results

For every merged extracted polygon the five measures (Sec. 2 and 3) are computed and the results are presented in Fig. 2. The values of the measures can not be compared directly to each other, because they do not all have the same units and do not compare the same quantities. Therefore, the colour bars are scaled from the best (Fig. 2, dark green) to the worst extracted polygon (Fig. 2, red) for every measure. The matched rates are presented as $1-$ the matched rate (Fig. 2a–c), thus the low values indicate that the polygon was quantified as well extracted according to that measure. The RMSE is computed twice, from the extracted to the reference polygons $RMSE_{A,B}$ and vice versa $RMSE_{B,A}$ (Fig. 2d–e).

Fig. 3a, b show the mean values and their standard deviations for all measures grouped in six categories grouped according to the area [m^2] of the reference building. The variability of all measures, except for correctness is higher for the categories containing larger buildings. The large variability (Fig. 3, error bars) of the measures for the category of larger buildings indicates a poor extraction of some larger buildings, due to their boundary complexity.

5. DISCUSSION

The per object assessment of the building extraction, which uses the number of e.g. correctly extracted buildings, is needed to asses how many buildings were not or were falsely detected. With this contribution we showed that even correctly detected buildings are detected with various quality. The choice for the per object quality measure(s) depends on the application, and must be jointly considered with the measures on per scene and per object count levels.

Quality rate and PoLiS metric are both symmetric measures and estimate overall dissimilarity between two polygons, whereas the RMSE, completeness, and correctness are not symmetric and should be therefore used for applications like change detection. For instance, $comp$ and $corr$ do not account for $ar(FP)$ and $ar(FN)$, respectively. Thus, if a larger part of a building is not detected, the 1-correctness is still low (Fig. 3b, building in the upper right corner). Similarly this is true for the $RMSE_{B,A}$. If $RMSE$ computes the distances to the polygons' edges, it provides a better numerical value to quantify the quality of the extraction method, but is still a non-symmetric measure. The PoLiS metric is symmetric and accounts directly for the polygons and their vertices, in comparison to the other measures. For building extraction evaluation, such symmetric measures like PoLiS metric should be used for majority of applications.

Acknowledgments

The authors would like to thank the city of Munich, Department of Environment and Health for providing the reference building footprints.

6. REFERENCES

[1] W. Song and T. Haithcoat, "Development of comprehensive accuracy assessment indexes for building footprint extraction," *IEEE Trans. Geosci. Remote Sens.*, vol. 43, no. 2, pp. 402–404, 2005.

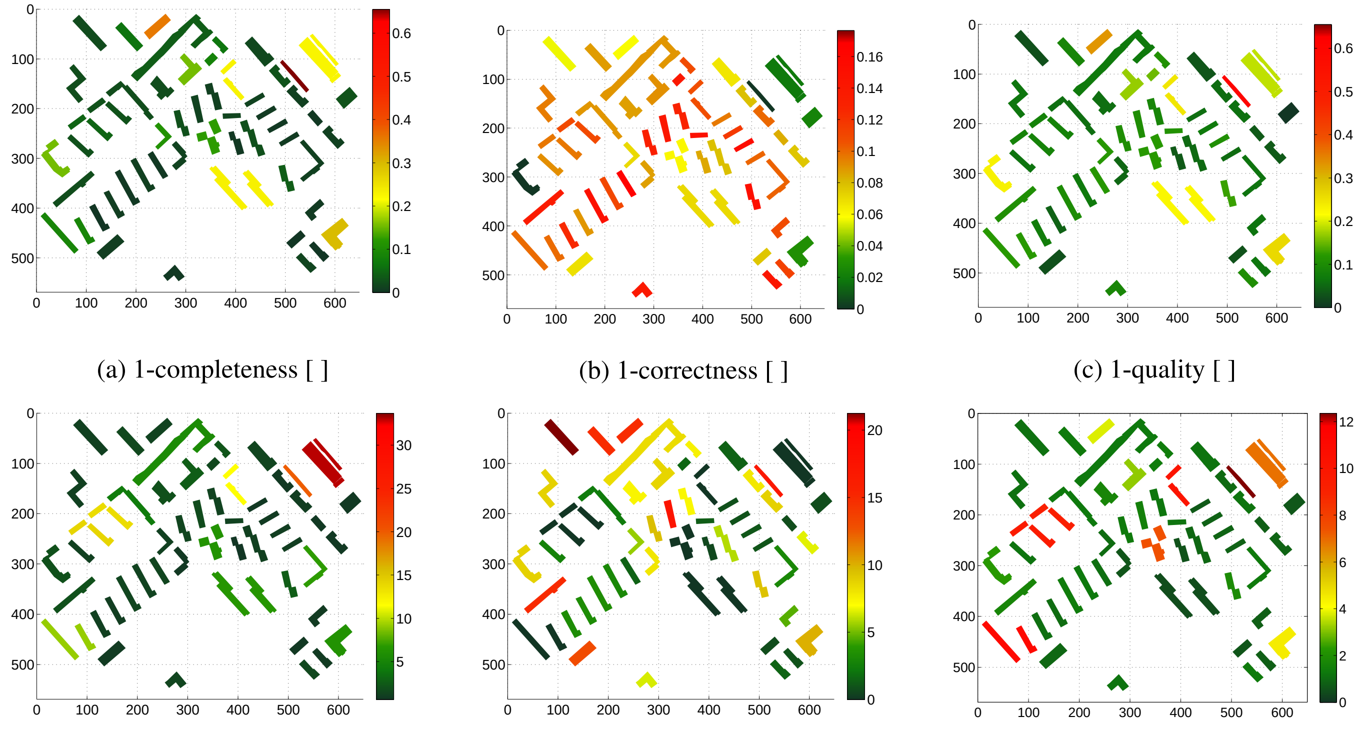
[2] M. Rutzinger, F. Rottensteiner, and N. Pfeifer, "A comparison of evaluation techniques for building extraction from airborne laser scanning," *IEEE J. Sel. Topics Appl. Earth Observ.*, vol. 2, no. 1, pp. 11–20, 2009.

[3] F. Rottensteiner, J. Trinder, S. Clode, and K. Kubik, "Building detection by fusion of airborne laser scanner data and multi-spectral images: Performance evaluation and sensitivity analysis," *ISPRS J. Photogrammetry and Remote Sensing*, vol. 62, no. 2, pp. 135–149, Jun. 2007.

[4] J. Shan and S. D. Lee, "Quality of building extraction from IKONOS imagery," *Journal of surveying engineering*, vol. 131, no. 1, pp. 27–32, 2005.

[5] C. Zeng, J. Wang, and B. Lehrbass, "An evaluation system for building footprint extraction from remotely sensed data," *IEEE J. Sel. Topics Appl. Earth Observ.*, vol. 6, no. 3, pp. 1640–1652, 2013.

[6] S. Freire, T. Santos, A. Navarro, F. Soares, J. D. Silva, N. Afonso, A. Fonseca, and J. Tenedrio, "Introducing mapping standards in the quality assessment of buildings



(a) 1-completeness [ ] (b) 1-correctness [ ] (c) 1-quality [ ]

(d) $RMSE_{A,B}$ [m] (e) $RMSE_{B,A}$ [m] (f) PoLiS [m]

Fig. 2: Evaluation of the building extraction by six measures. Their values (a)-(f) can be compared relative to each other. Colour bars for each measure is scaled from the best extracted building footprint (dark green) to the the worst (red) for easier visual interpretation.

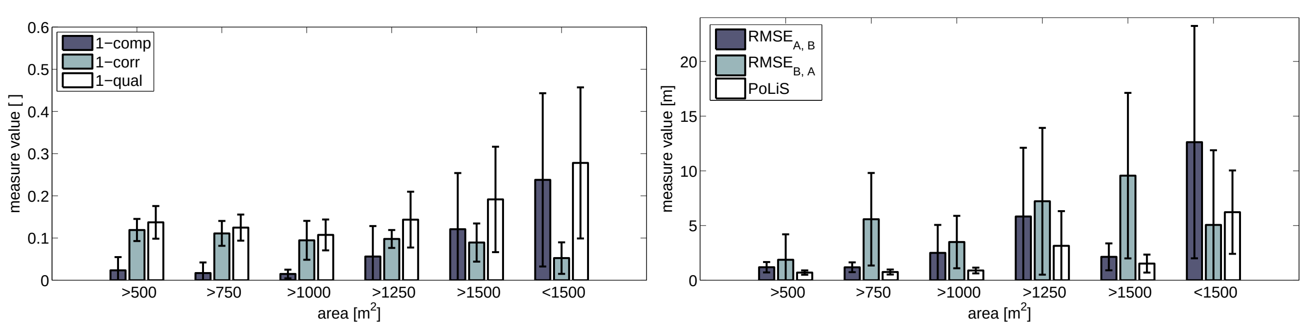


Fig. 3: Comparison of indices for building polygon extraction assessment. For each category and for each measure, the height of the bar represent the mean value of the measure, and the error bar represents its variability (standard deviation).

extracted from very high resolution satellite imagery," vol. 90, pp. 1–9.

[7] J. Avbelj, R. Müller, and R. Bamler, "A metric for polygon comparison and building extraction evaluation," *IEEE Geosci. Remote Sens. Lett.*, in press.

[8] F. Rottensteiner, G. Sohn, M. Gerke, J. D. Wegner, U. Breitkopf, and J. Jung, "Results of the ISPRS benchmark on urban object detection and 3D building reconstruction," *ISPRS J. Photogrammetry and Remote Sensing*, 2014.

[9] J. Avbelj, D. Iwaszczuk, R. Müller, P. Reinartz, and U. Stilla, "Coregistration refinement of hyperspectral images and DSM: An object-based approach using spectral information," *ISPRS J. Photogrammetry and Remote Sensing*, in press.