A Robust Performance Evaluation Metric for Extracted Building Boundaries From Remote Sensing Data

Emon Kumar Dey and Mohammad Awrangjeb, *Senior Member, IEEE*

***Abstract*—Various methods for automatic building extraction from remote sensing data including light detection and ranging (LiDAR) data have been proposed over the last two decades but a standard metric for evaluation of the extracted building boundary has not been found yet. An extracted building boundary from LiDAR data usually has a zigzag pattern with missing detail, which makes it hard to compare the boundary with its reference. The existing metrics do not consider the significant point (e.g., corner) correspondences, therefore, cannot identify individual extralap and underlap areas in the extracted boundary. This article proposes an evaluation metric for the extracted boundary based on a newly proposed robust corner correspondence algorithm that finds one-to-one true corner correspondences between the reference and extracted boundaries. Assuming a building has a rectilinear shape, corners and lines are first detected for the extracted boundary. Then, corner correspondences are obtained between the extracted and reference boundaries. Each corner has two corresponding lines on its two sides that ideally are perpendicular to each other. The corner correspondences are finally ranked based on their distance, angle, and parallelism of corresponding lines. The metric is defined as the average minimum distance d_{avg} from the extracted boundary points to their corresponding reference lines. Extralap and underlap areas are identified by comparing the point distances with d_{avg}. In experiments, the proposed metric performs more realistic than the existing metrics and finds the individual extralap and underlap areas effectively.**

***Index Terms*—Building boundary extraction, evaluation, object-based shape similarity, performance, polygon comparison.**

I. Introduction

BUILDING outlines are a standard part of recent topographic databases. For many applications such as 3-D city modeling, urban planning, and disaster management, the mapping of building outlines is important. Therefore, studies have been conducted and numerous approaches have been developed in fully automated or semiautomated manners using airborne light detection and ranging (LiDAR) data and/or remote sensing images to extract building boundaries [1]–[5].

A building boundary or outline can be defined as a closed polygon that represents the outer shape of a building roof. In order to evaluate the results of a building outline extraction method, extracted boundaries are compared with the corresponding reference data. The existing evaluation methods for extracted 2-D and 3-D building boundaries can be mainly categorized into two types, i.e., object-based [6]–[8] and area-based [8]–[10] methods. They use evaluation metrics to estimate the performance of an involved building extraction method. In object-based evaluation either a boundary is labeled as correctly or wrongly extracted or a quality measure is assigned to each extracted boundary [11]. It offers a quick assessment. The area-based approach corresponds to the horizontal accuracy of the extracted building footprint. Different metrics such as, completeness (recall), correctness (precision), and/or quality values are calculated using the traditional formulas that are based on the total area of overlapping building parts and the area of corresponding reference buildings [12]. If the image data are used as an input, then a rasterization process must be performed on the extracted building area [8]. But for the case of LiDAR data, rasterization of the vector data [13] may lead to a severe misalignment error and the estimated performance can distort considerably when the average building size is small [6].

The existing metrics [12], [14] for evaluating the performance of building boundary extraction techniques can be categorized based on matched rates (e.g., completeness, correctness, and quality rates) [6], [15], shape similarity (e.g., turning angle, Fourier descriptors, and dominant angle rotation) [16], [17], positional accuracy [e.g., the root mean square error (RMSE), the normalized median absolute deviation (NMAD), and the mean absolute error (MAE)] [18]–[22]. The matched rates are mainly used by the area-based, also known as the pixel-based, approach [12]. Boundary level shape similarity metrics have been used by many authors to evaluate the extracted buildings [16], [23], [24]. Different topological relations such as corner positional difference, perimeter difference, and overlapping areas between the reference and extracted buildings are used to define a metric in this case [25], [26]. The RMSE is a frequently used metric to measure the positional accuracy [7], [16], [24], [26]. It is based on the distance of each corner of an extracted building polygon to a corresponding reference corner. The NMAD and the MAE are used to investigate the accuracy of the reconstructed 3-D buildings from the satellite imagery [21], [22]. Yet, there is no standard metric for evaluation of the extracted building boundaries in the remote sensing community [27].

In a strict mathematical terminology, a metric or a distance function (d) should follow three basic properties for each pair of

Manuscript received April 12, 2020; revised June 11, 2020 and June 24, 2020; accepted June 26, 2020. Date of publication July 1, 2020; date of current version July 22, 2020. *(Corresponding author: Emon Kumar Dey.)*

The authors are with the School of Information and Communication Technology, Griffith University, Nathan, QLD 4111, Australia (e-mail: emonkumar.dey@griffithuni.edu.au; m.awrangjeb@griffith.edu.au).

Digital Object Identifier 10.1109/JSTARS.2020.3006258

elements in a set ρ, such as, nonnegativity ($d(x, y) \geq 0$), symmetry ($d(x, y) = d(y, x)$), and triangular inequality ($d(x, y) \leq d(x, z) + d(z, y)$) for all $x, y, z \in \rho$ [28]. For different shape matching applications and extracted building boundary evaluation, the Hausdorff [29] and the Chamfer [28] distances, which allow different number of points in the reference and extracted shapes, are used [30]–[32]. These distance metrics obviously fulfill the nonnegativity and the triangular inequality properties, but not the symmetry property because the distance measured from the extracted boundary to the reference boundary is not the same as that of from the reference boundary to the extracted boundary [11], [12].

To fulfill the symmetry property, Avbelj *et al.* [28] proposed the polygon and line segments (PoLiS) metric, where they applied a symmetrization technique by simply averaging the two distances (from the extracted to the reference and vice versa). The PoLiS is straightforward and easy to implement [33]. It considers only the shortest distance from a corner of one boundary to any point in the other boundary. However, this may mislead the actual result sometimes, particularly, in the case of segmentation errors (extralap and underlap areas) in an extracted boundary. Moreover, for a complex building with a higher number of extracted corner points than the number of reference corner, or vice-versa, influences the actual distance value considerably [12].

Extralap and underlap errors [10] in the extracted building boundary are a common issue (see Fig. 1). Due to dissimilar sources and methodologies involved in generation of the reference and extracted boundaries, in practice, an extracted boundary has always some misalignment (e.g., Areas D and E in Fig. 1) with its reference boundary. Actual segmentation errors (e.g., Areas A, B, and C in Fig. 1) happen when the involved building extraction method fails to handle the various issues (e.g., resolution, point density, shadow, and occlusion) with the input data. The existing evaluation methods [6], [34], [35] do not differentiate between the misalignment and segmentation errors. Therefore, they also do not individually identify these errors.

This article proposes a new performance evaluation metric that is based on robust corner correspondences between the extracted and reference boundaries. The newly proposed robust corner correspondence (RCC) algorithm finds one-to-one true corner correspondences between the reference and extracted boundaries. The obtained corner correspondences are ranked based on distance, angle, and parallelism of lines. The proposed metric, coined as the RCC metric, is then defined as the average minimum distance d_{avg} from the extracted boundary points to their corresponding reference lines. The RCC metric fulfills all three conditions of a mathematical metric. During performance evaluation, extralap and underlap errors are separately identified by comparing the point to line distances with d_{avg}. While long distances indicate a possible segmentation error, short distances represent a misalignment error.

The particular contributions of the article are as follows.

1) The RCC algorithm finds one to one true corner correspondences between the extracted and reference building boundaries. The algorithm works in the presence of noise (e.g., zigzag pattern) and extralap and underlap errors.

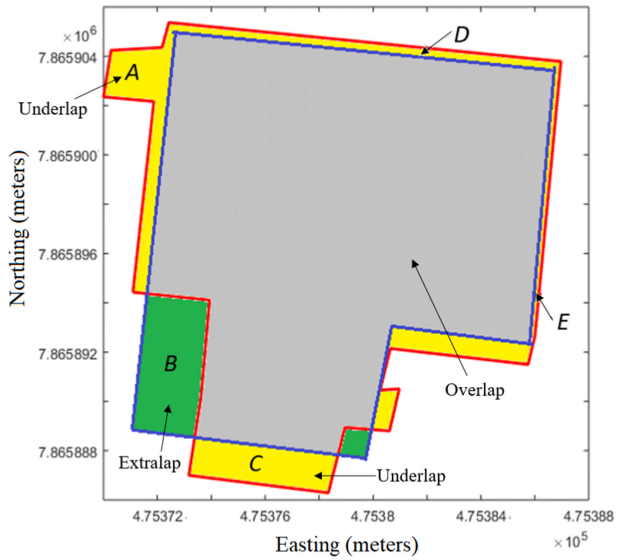

Fig. 1. Reference (red) and extracted (blue) building polygons. Gray, yellow, and green areas represent true positive, false negative, and false positive areas, respectively.

2) The newly proposed RCC metric offers a realistic measurement of distances between the extracted and reference building boundaries. Since the existing metrics do not use corner to corner correspondences, they may sometimes find unrealistic measurements.
3) During evaluation, we can identify the extralap or underlap errors based on the proposed RCC metric. An extensive experimentation has been carried out on datasets from three geographic locations.

Note that like the exiting distance metrics (e.g., PoLiS, RMSE), the proposed RCC metric finds the distance between a reference building and its corresponding extracted building. Each building correspondence, i.e., reference-extracted pair, is chosen based on the maximum overlap between the reference and extracted buildings [10].

The rest of this article is organized as follows. In Section II, we review some existing metrics and their challenges. The proposed metric is presented and discussed in Section III. Section IV represents the experimental results, and finally, Section V concludes this article.

II. Review of Existing Metrics

The extracted building boundary B_{e} from a building extraction method can be represented as an ordered set of n points $a_i \in S_{\text{e}}$, $1 \leq i \leq n$, where any two consecutive points a_i and a_{i+1} present two neighboring points on the boundary. Thus, the representation is like a polygon consisting of n points (vertices) and n lines (vector format) [28]. The reference boundary B_{r} also usually comes in the vector format consisting of points and lines, $b_j \in S_{\text{r}}$, where $1 \leq j \leq m$.

In practice, while B_{e} contains roof corner points (or points close to corners) as well as the points between two consecutive corners, B_{r} contains only corner points. However, for the sake of discussion in this article, we assume both B_{e} and B_{r} contain

roof corners (or points close to corners) and the points between successive corners. In B_{r}, if the points in between corners are not available, they can be simply inserted in an equal distance depending on the input LiDAR point density. In B_{e}, corners can be detected and inserted following the method presented in [36]. Let the number of corners in B_{r} be p and that in B_{e} be q. Also, let the sets of corners in B_{e} and B_{r} be $a_k \in \zeta_{\mathrm{e}} \subset S_{\mathrm{e}}$ and $b_l \in \zeta_{\mathrm{r}} \subset S_{\mathrm{r}}$, respectively, where $1 \leq k \leq q$ and $1 \leq l \leq p$.

In this section, we discuss the challenges of some existing metrics that are used for evaluation of the extracted building boundary.

Recall R_{c}, precision P_{r} (also known as completeness C_{m} and correctness C_{r}, respectively [37]) and quality Q_{l} defined in (1)–(3) are frequently used area-based metrics for evaluating the extracted building boundary or individual roof parts of buildings for generating 3-D models [37]–[40]. These are based on the matched rates between the extracted and corresponding reference polygon areas. Four different parameters are defined for these three metrics. These are true positive (TP, common area between B_{e} and B_{r}), true negative (TN, area which is neither in B_{e} nor in B_{r}), false positive (FP, area in B_{e} but not in B_{r}), and false negative (FN, area not in B_{e} but in B_{r}) [16], [41]. Fig. 1 shows an example of these parameters.

$$R_{\mathrm{c}} = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FN}} \tag{1}$$

$$P_{\mathrm{r}} = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FP}} \tag{2}$$

$$Q_{\mathrm{l}} = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FP} + \mathrm{FN}}. \tag{3}$$

Although R_{c}, P_{r}, and Q_{l} are frequently used by many authors, most of the times these three metrics require the rasterization process that may introduce errors and should be avoided [6], [12]. R_{c} and P_{r} do not fulfill the symmetry property of a mathematical metric and they are more appropriate for the applications like change detection [11]. Moreover, R_{c} (C_{m}) does not consider about the FP and P_{r} (C_{r}) does not consider the FN rates that sometimes give a false estimation when a large part of a vegetation is extracted as building part or a large part of a building remains undetected [11].

Metrics such as Matthew's correlation coefficient (MCC) and F1-score (F1), defined by (4) and (5), respectively, as follows based on the point coverage of extracted boundaries, are used by some authors to evaluate the reconstructed 3-D buildings [42], [43]. Individual extracted roof parts are evaluated with the corresponding reference data in this case. Orthogonal distances from the points of the extracted plane to the corresponding reference plane and the minimum distance of the detected corners are also used by some authors to evaluate the extracted 3-D roof planes [10], [44].

$$\mathrm{MCC} = \frac{\mathrm{TP} \times \mathrm{TN} - \mathrm{FP} \times \mathrm{FN}}{\sqrt{(\mathrm{TP} + \mathrm{FP})(\mathrm{TP} + \mathrm{FN})(\mathrm{TN} + \mathrm{FP})(\mathrm{TN} + \mathrm{FN})}} \tag{4}$$

$$\mathrm{F1} = \frac{2 \times \mathrm{TP}}{2 \times \mathrm{TP} + \mathrm{FN} + \mathrm{FP}}. \tag{5}$$

Li *et al.* [39] use three different metrics to evaluate the quality of the extracted 3-D roof planes of individual buildings. These are boundary precision (B_{pr}), boundary recall (B_{rc}), and average of precision-recall (F_{m}). Equations (6)–(8) define these metrics, respectively, where B_{e} and B_{r} denote the boundary points of the extracted and corresponding reference planes, respectively, and $|\,|$ indicates the number of points in a set. The least-squares 3-D surface matching (LS3D) method is applied by Akca *et al.* [45] to assess the quality of 3-D building data. The metric they used is based on the Euclidean distance between each extracted plane to its reference plane.

$$B_{\mathrm{pr}} = \frac{|B_{\mathrm{e}} \cap B_{\mathrm{r}}|}{|B_{\mathrm{e}}|} \tag{6}$$

$$B_{\mathrm{rc}} = \frac{|B_{\mathrm{e}} \cap B_{\mathrm{r}}|}{|B_{\mathrm{r}}|} \tag{7}$$

$$F_{\mathrm{m}} = \frac{2}{\frac{1}{B_{\mathrm{rc}}} + \frac{1}{B_{\mathrm{pr}}}}. \tag{8}$$

The Hausdorff and the Chamfer distances are frequently used by many authors to quantify the object-based similarities between two shapes [30]. These metrics do not require to establish the point correspondences [30], [46]. The Hausdorff distance $h(S_{\mathrm{e}}, S_{\mathrm{r}})$ between point sets S_{e} and S_{r} is a max-min distance and the Chamfer distance $c(S_{\mathrm{e}}, S_{\mathrm{r}})$ is a sum of distance between each point in S_{e} and its closest point in S_{r} (see (9) and (10), where $\|.\|$ represents the Euclidean distance [47]). Both h and c do not follow the symmetry property of a mathematical metric [i.e., $h(S_{\mathrm{e}}, S_{\mathrm{r}}) \neq h(S_{\mathrm{r}}, S_{\mathrm{e}})$ and $c(S_{\mathrm{e}}, S_{\mathrm{r}}) \neq c(S_{\mathrm{r}}, S_{\mathrm{e}})$] and they are very sensitive to outliers. Thus, the evaluation result using h or c is found to be unrealistic for a given extracted building boundary [28].

$$h(S_{\mathrm{e}}, S_{\mathrm{r}}) = \max_{a_i \in S_{\mathrm{e}}} \min_{b_j \in S_{\mathrm{r}}} \|a_i - b_j\| \tag{9}$$

and

$$c(S_{\mathrm{e}}, S_{\mathrm{r}}) = \sum_{a_i \in S_{\mathrm{e}}}^{p} \left(\min_{b_j \in S_{\mathrm{r}}} \|a_i - b_j\| \right). \tag{10}$$

The RMSE λ is another frequently used object-based metric to measure the positional accuracy of the extracted boundary with respect to its reference boundary [26]. For each point $a_i \in S_{\mathrm{e}}$, it finds the nearest point $b_j \in S_{\mathrm{r}}$. Using the Euclidian distance between the points, the RMSE is then calculated using (11) [48]

$$\lambda(S_{\mathrm{e}}, S_{\mathrm{r}}) = \sqrt{\frac{1}{n} \sum\nolimits_{i=1}^{n} \left(\min_{a_i \in S_{\mathrm{e}}, b_j \in S_{\mathrm{r}}} \|a_i - b_j\| \right)^2}. \tag{11}$$

Some authors calculate the RMSE from the points of the reference building to the extracted building. However, the RMSE distance also does not follow the symmetry property of a mathematical metric, that is, $\lambda(S_{\mathrm{e}}, S_{\mathrm{r}}) \neq \lambda(S_{\mathrm{r}}, S_{\mathrm{e}})$ [49]. Some authors instead of using all points along the boundaries, use the corners. So, they regularize the extracted building boundary and detect corners on the regularized boundary. Then, they calculate the RMSE using the corners from the extracted and reference boundaries. Sometimes, they use a threshold to exclude some distances,

which reduce the objectivity. Consequently, it brings several complications because of not finding the correspondences between the corner points of the reference and extracted boundaries, even though the extraction method is effective enough [12].

The NMAD is considered as a robust metric to evaluate the extracted building shapes in the presence of outliers with a non-normal distribution [22], [50]. The MAE is another frequently used metric, similar to the Chamfer distance used by several authors to evaluate the extracted 3-D buildings [21]. The NMAD ν and MAE $\hat{\mu}$ are defined in (12) and (13), respectively, where $\triangle h$ is the set of all minimum distances from the extracted points (S_{e}) to the reference points (S_{r}) and $c_\nu = 1.4826$ is a constant to normalize the distance [22]. However, none of these satisfies the properties of a mathematical metric discussed earlier.

$$\nu(S_{\mathrm{e}}, S_{\mathrm{r}}) = c_\nu \times \mathrm{median}_{i=1,\ldots,n} (\triangle h_i - \mathrm{median}_{j=1\ldots,n}(\triangle h_j)) \tag{12}$$

$$\hat{\mu}(S_{\mathrm{e}}, S_{\mathrm{r}}) = \frac{1}{q}\sum_{i=1}^{q} \| \triangle h_i \|. \tag{13}$$

The PoLiS is an object-based metric $\wp$ that does not use any threshold [28]. The distance between each corner of the extracted shape to any nearest point in the reference outline [see (15)] and vice versa [see (16)] is calculated first. The average distance is then considered as the final PoLiS distance [see (14)]. The result using $\wp$ changes almost linearly with respect to small changes that occur in rotation, translation, or scaling between B_{e} and B_{r}.

$$\wp = \frac{\wp(S_{\mathrm{e}}, S_{\mathrm{r}}) + \wp(S_{\mathrm{r}}, S_{\mathrm{e}})}{2} \tag{14}$$

where

$$\wp(S_{\mathrm{e}}, S_{\mathrm{r}}) = \frac{1}{q}\sum_{k=1}^{q} \left(\min_{a_k \in \zeta_{\mathrm{e}}} \|a_k - b_j\| \right)^2 \tag{15}$$

and

$$\wp(S_{\mathrm{r}}, S_{\mathrm{e}}) = \frac{1}{p}\sum_{l=1}^{p} \left(\min_{b_l \in \zeta_{\mathrm{r}}} \|b_l - a_i\| \right)^2. \tag{16}$$

Since, $\wp(S_{\mathrm{e}}, S_{\mathrm{r}}) \neq \wp(S_{\mathrm{r}}, S_{\mathrm{e}})$, $\wp(S_{\mathrm{e}}, S_{\mathrm{r}})$, or $\wp(S_{\mathrm{r}}, S_{\mathrm{e}})$ alone does not follow the symmetry property of a mathematical metric, the average in (14) is applied to symmetrize the final PoLiS distance [28].

The PoLiS or the RMSE consider the minimum distance that sometimes results in a wrong distance measure, particularly, when segmentation error happens. For example, Fig. 2 shows an extracted (B_{e}, black dots) and its corresponding reference (B_{r}, red lines) building boundaries. We assume that some part of the building is not properly extracted (extralap or underlap error) due to vegetation or missing LiDAR data. For the corner O on B_{e}, the nearest point on B_{r} is A. So, by the definition, the RMSE or the PoLiS will consider OA (dashed orange line) as the estimated distance for O. But practically, if we consider Q and R on B_{r} as corresponding corners of W and Z on B_{e}, respectively, then OC should be considered as the estimated distance for O. Therefore, both the RMSE and the PoLiS estimate wrong distance in this case. The same happens for U as well.

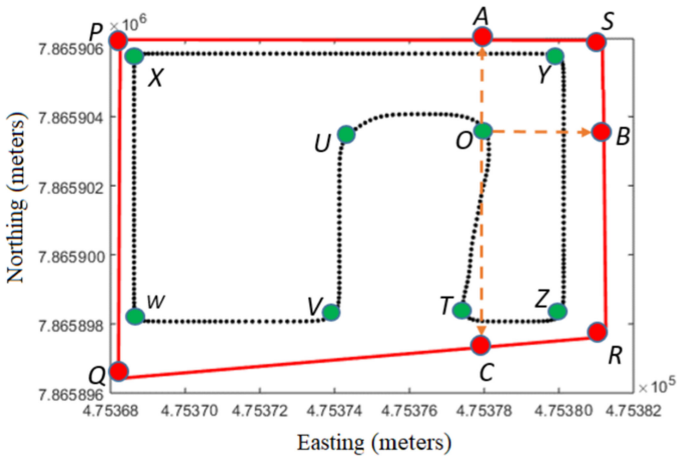

Fig. 2. Reference (red line) and extracted (black dots) building polygons. Filled circles represent some representative points or corners on the polygons.

To solve this issue, for each of the reference corners, we first find a true corner correspondence from the extracted boundary. Then, we estimate the distance appropriately by checking the corresponding boundary segments between any two consecutive true corners. For example, we can find the matching corner pairs as (P, X), (Q, W), (R, Z), and (S, Y) between ζ_{e} and ζ_{r}. Then, the two segments QR and $WVUOTZ$ between (Q, W) and (R, Z) are checked to estimate the appropriate distances for V, U, O, and T to QR. By applying an appropriate threshold, we can also correctly locate the extralap and underlap errors.

III. Proposed Method

From the aforementioned discussion, we can see that the existing metrics are mainly based on the minimum distance and do not consider the corner correspondences at all. This may mislead the distance estimation. Moreover, they do not locate the extralap and underlap errors. In this section, we first propose a robust algorithm for finding the true corner correspondences between ζ_{e} and ζ_{r}. Then, we define the new metric (RCC) that provides the average minimum distance d_{avg} from B_{e} to ζ_{r}. The use of B_{e} instead of ζ_{e} in the estimation of d_{avg} avoids any changes introduced to the extracted boundary by the involved boundary regularization technique. Finally, extralap or underlap areas are identified based on an appropriate threshold, which can be estimated based on either d_{avg} or the input LiDAR point density.

A. Robust Corner Correspondence (RCC)

As shown in Fig. 3, let any two consecutive sides (lines) of ζ_{r} be AP and PB and those of ζ_{e} be SQ and QR, where A, P, B, S, Q, and R are corners. A corner angle is shown between consecutive lines, i.e., $\angle APB$ and $\angle SQR$ are reference and extracted angles, respectively. In a rectilinear building, each such angle is ideally a right angle. If P and Q form a pair (P, Q) of corners between ζ_{r} and ζ_{e}, then ideally $PA \parallel QS$ and $PB \parallel QR$, which mean there are two pairs of corresponding parallel lines for (P, Q). However, in practice, a reference or extracted

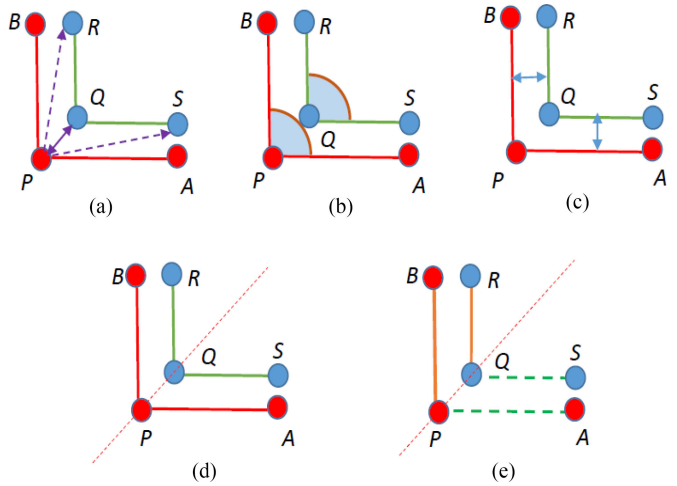

Fig. 3. Five properties to rank corner correspondences: (a) corner distance, (b) angle difference, (c) distance between corresponding lines, (d) position of adjacent corner pair, and (e) number of parallel line pairs. A, P, and B are three consecutive reference corners and S, Q, and R are three consecutive extracted corners.

corner may not show an exact right angle. Even if they show, their orientation may get changed. Therefore, an angle threshold θ_{th} is applied to decide parallelism [36]. Yet, parallel lines may not be found for a true corner correspondence, particularly, due to segmentation error or absence of data. Therefore, five different properties including distances and angle difference described as follows are considered to rank the corner pairs between ζ_{r} and ζ_{e}. For each property, a matrix M_f of size $p \times q$ is constructed, where p and q are number of reference and extracted corners, respectively, and $1 \leq f \leq 5$.

1) *Distance between corners:* For each reference corner, its distance to each of the extracted corners is calculated. The matrix M_1 is then constructed where $M_1(k,l)$ indicates the distance from kth extracted to lth reference corners. If (P,Q) is a true corner correspondence, then $\|PQ\|$ is expected to be the smallest among all the distances [see Fig. 3(a)] in the corresponding column in M_1.
2) *Angle difference between corners:* False corners may not have angles close to the right angle. So, for each reference angle, its absolute difference with each of the extracted angles is calculated. The matrix M_2 is then constructed where $M_2(k,l)$ indicates the absolute angle difference between kth extracted angle and lth reference angle. If $\angle SQR$ is the corresponding extracted angle for $\angle APB$ [see Fig. 3(b)], then $|\angle APB - \angle SQR|$ is expected to be the smallest difference among all the differences in the corresponding column in M_2.
3) *Average distance between corresponding lines:* For each pair of the reference and extracted corners, two distances are calculated between the corresponding parallel lines [see Fig. 3(c)]. The matrix M_3 is then constructed where $M_3(k,l)$ indicates the average parallel line distance from the kth extracted corner to lth reference corner. If (P,Q) is a true corner correspondence and $PA \parallel QS$ and $PB \parallel QR$, then their average distance is expected to be the smallest among all the average differences in the corresponding column in M_3.
4) *Position of adjacent corner pairs:* Referring to Fig. 3(d), for the corner pair (P,Q), it has two adjacent corner pairs (A,S), which is corresponding to parallel lines PA and QS, and (B,R), which is corresponding to parallel lines PB and QR. If (P,Q) is a true corner correspondence, an intuitive observation is that the two corners in each of the adjacent corner pairs reside on the same side of the line PQ. In Fig. 2 for the true corner correspondence (R,Z), corners S and Y of the pair (S,Y) reside on the same side of the line RZ. Moreover, corners Q and T of the pair (Q,T) reside on the other same side of RZ. For the false corner correspondence (R,T), though the pair (S,O) satisfies this observation, the other pair (Q,Z) does not with respect to line RT. For the false correspondence (S,T), none of the pairs (R,O) and (P,Z) satisfy this with respect to line ST. The matrix M_4 is constructed such that $M_4(k,l) = 0.25$ when both pairs satisfy this, $M_4(k,l) = 2N$ otherwise where $N > 1$.
5) *Number of parallel line pairs:* For each pair of the reference and extracted corners, the number of parallel line pairs is recorded [see Fig. 3(e)]. The matrix M_5 is constructed such that $M_5(k,l) = 0.25$ when both pairs of lines are parallel, $M_5(k,l) = 0.5$ when only one pair of lines is parallel, and $M_5(k,l) = N$ when no pairs of lines are parallel.

Two different thresholds are used to find corresponding parallel lines. A flexible threshold $\theta_{\text{th}} = \frac{\pi}{4}$ is used for all matrices except for M_4, for which $\theta_{\text{th}} = \frac{\pi}{8}$ is set [36]. The flexible value allows to find parallel lines even for false candidates, but the tight value offers credits mostly for the true correspondences.

Since the aforementioned first three properties are based on the distance and angle, the values of each of the matrices (M_1, M_2, or M_3) are normalized by using its maximum value

$$M_f(k,l) = \frac{M_f(k,l)}{\max(M_f)} \tag{17}$$

where $f = 1, 2, 3$. Finally, the following formula is defined to rank the pairs of extracted and reference corners:

$$\Re(k,l) = \prod_{f=1}^{5} M_f(k,l). \tag{18}$$

The values in $\Re$ are sorted in ascending order, so the corner pair (the kth extracted and the lth reference corners) with the smallest value is at the top of the rank.

The position property (M_4) shows high capability in differentiating the true and false corner correspondences between ζ_{r} and ζ_{e}. Any corner pair that does not satisfy this property is decided to be an "unsuitable" candidate and remains at the bottom of the rank or removed. For example, as shown in Fig. 2, for the reference corner P, only two (X and U) out of eight extracted corners satisfy the property, where (P,X) is a true correspondence and (P,U) is a false correspondence. The remaining six

pairs are "unsuitable" candidates. Consequently, only 8 out of a total of 32 correspondences will have $M_4(k,l) = 0.25$ and the rest ("unsuitable") will have $M_4(k,l) = 2N$. In contrast, the parallel line property (M_5) has relatively low capability as due to extralap or underlap error, some true corner correspondences may not satisfy this property. Thus, $M_5(k,l) = N$ is set (as compared to $M_4(k,l) = 2N$ when the position property is not satisfied for a correspondence which is more likely false) so that a true correspondence that does not meet the parallel property is at a higher rank than an "unsuitable" correspondence that does not meet the position property. For this purpose, the value of N is chosen to be a large number (say, 10^6).

The product of the normalized values of the first three properties in (18) is within the range [0,1]. Thus, when a given correspondence satisfies the fourth, the fifth, or both properties, it becomes at a higher rank. On the contrary, if it does not satisfy both the fourth and the fifth properties, it can come at the bottom of the rank.

For a given reference corner, the aforementioned ranking based on the numerical value only may put a false correspondence at a higher rank than its true correspondence. So, the final corner correspondences are decided based on the "adjacent correspondence consistency" property, where to decide a true corner correspondence (P,Q), the adjacent corners of P and Q are also checked if they are true correspondences too. For example, from Fig. 2, the reference corner P has two "suitable" candidate pairs (P,X) and (P,U). If (P,U) has a smaller ranking value than (P,X), (P,U) will be chosen as the true corner correspondence, which is a wrong decision. To avoid this, their adjacent corner pairs are checked. For (P,U), the adjacent pair (S,O) can be ranked high but (Q,V) is not because it is an "unsuitable" candidate. So, (P,U) is eventually decided as a false candidate. In contrast, for (P,X), both of the adjacent pairs (S,Y) and (Q,W) are "suitable" candidates and eventually found to be true correspondences by checking their adjacent corners. Thus, (P,X) is decided to be a true candidate. The recursive procedure, therefore, finalizes all four true correspondences (P,X), (Q,W), (S,Y), and (R,Z). Two stop conditions are considered for the recursion. If for P, no true candidate pair is found, its adjacent corners are decided based on their other adjacent corners. After the "adjacent correspondence consistency" check, if there are more than two "suitable" candidates found for P, the one with the lower numerical value (i.e., $\Re(k,l)$) is decided to be the true candidate.

Fig. 4 shows the true corner correspondences within red dashed ellipses for another representative building.

B. RCC Metric

Now, in order to quantify the overall dissimilarity between B_e and ζ_r, we find the average minimum distance d_{avg} from the LiDAR points on B_e to lines of ζ_r.

Because of the nature of remote sensing data, Type-I (false positive) and Type-II (false negative) errors [51] are common in any building boundary extraction methods. The proposed RCC metric finds the best corner correspondences, thus there can be some remaining corners in the extracted or reference boundaries

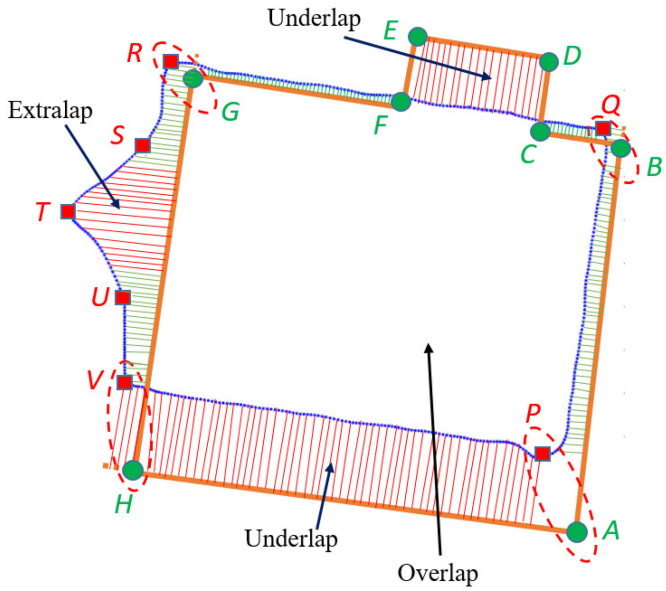

Fig. 4. True corner correspondences within red-dashed ellipses for another representative building. Green circles represent corners of the reference polygon ζ_r consisting of orange solid lines and red rectangles show corners of the extracted polygon B_e represented by blue dot (LiDAR) points. C, D, E, and F are missed true corners and S, T, and U are extracted false corners.

that do not have any matched correspondences. For example, in Fig. 4, the remaining corners in the extracted boundary lead to Type-I error and those in the reference boundary cause Type-II error. Therefore, there can be one of the following three situations between any two consecutive corner correspondences: a reference line corresponds to one extracted segment (e.g., AB with PQ or AH with PV); a reference line corresponds to two or more extracted segments (e.g., GH with $RSTUV$) because of the Type-I error; and an extracted segment corresponds to two or more reference lines (e.g., QR with $BCDEFG$) because of the Type-II error.

For the first two situations, between the two consecutive corner correspondences, we have only one reference line that corresponds to one or more extracted boundary segments. The perpendicular distances are calculated from the LiDAR points on the extracted segments to the corresponding reference line. For example, in Fig. 4, the extracted segment PQ and the reference line AB are within the true corner correspondences (A,P) and (B,Q). The green lines between PQ and AB are the estimated perpendicular distances. Similarly, within the true corner correspondences (G,R) and (H,V), there are four extracted segments RS, ST, TU, and UV for the reference line GH and the perpendicular distances, shown by green and red lines between $RSTUV$ and GH, are calculated.

For the third situation, between the two consecutive corner correspondences, we have only one extracted segment that corresponds to one or more reference lines. The perpendicular distances are again calculated from the LiDAR points on the extracted segment to its corresponding parallel reference lines. For example, in Fig. 4, the extracted segment QR has five corresponding reference lines BC, CD, DE, EF, and FG within the true corner correspondences (B,Q) and (G,R). In

this situation, the corresponding reference line for an extracted boundary point is selected based on the 2-D position (i.e., the start and end points) of the reference line and the extracted point. The perpendicular distances, shown by green and red lines between QR and $BCDEFG$, are calculated. The explanation for choosing two different color lines will be provided in Section III-C.

After calculation of all such distances, the RCC metric that calculates d_{avg} from the extracted to the reference boundaries is defined as follows:

$$d_{\text{avg}}(S_{\text{e}}, S_{\text{r}}) = \frac{1}{n}\sum_{t=1}^{n} d_{\text{p}} \tag{19}$$

where n is the total number of LiDAR points along B_{e}, and d_{p} is the perpendicular distance from a LiDAR point to the corresponding reference line. Two consecutive pair of corner correspondences indicate the exact reference lines for which d_{p} is calculated for the extracted boundary points within the consecutive matched corners on the extracted boundary. Like the PoLiS distance in (14), the aforementioned RCC metric can be symmetrized by taking the average of $d_{\text{avg}}(S_{\text{e}}, S_{\text{r}})$ and $d_{\text{avg}}(S_{\text{r}}, S_{\text{e}})$.

C. Segmentation Errors

We identify the segmentation errors (extralap and underlap areas) automatically by using the average minimum distance of the perpendicular distances between two consecutive corner correspondences. The extralap and underlap areas can be distinguished based on the overlap and nonoverlap areas between the extracted polygon and its reference polygon [10], [27], [52]. We mainly follow the definition in Awrangjeb and Fraser [10]. The common area shared by both the reference and the extracted boundaries is called the overlap area. When the boundary of the reference building splits the extracted boundary into more than one part, then the parts of the extracted boundary that fall outside the reference polygon are considered as extralap areas. Similarly, when a reference polygon is split by the extracted boundary into more than one part, then the parts which fall outside the extracted polygon are known as underlap areas. Figs. 1 and 4 show examples of these three types of areas. So, an extralap happens when some extracted points are found outside B_{r} and an underlap happens when some extracted points are found inside B_{r}. However, many such areas (e.g., Areas D and E in Fig. 1) are not important since a misalignment between B_{e} and B_{r} always happens in practice. So, we follow the following three steps to find noticeable errors (e.g., Areas A, B, and C in Fig. 1) using (19) between two consecutive corner correspondences, where we have one of the three situations discussed in Section III-B.

First, between two consecutive true corner correspondences, there is only one reference line that corresponds to more than one extracted segment. The perpendicular distances from the LiDAR points on the extracted segments to the reference line are compared with their average minimum distance ($d_{\text{p}} > c.d_{\text{avg}}$, where $c > 1$). For example, in Fig. 4, for the reference line GH, there are four segments RS, ST, TU, and UV. The perpendicular distances (green and red lines) from these segments to GH are compared with their average distance. The red lines related to the points with high d_{p} are decided to be in an extralap area since these LiDAR points on the extracted segments are outside B_{r}. Green lines indicate the distances with no extralap or underlap areas. Similarly, in Fig. 2, the points within the extracted segments VU, UO, and OT form an underlap area with the reference line QR since these points are inside the reference $PQRS$.

Second, similarly we find the errors when between two consecutive true corner correspondences, there is only one extracted segment that corresponds to more than one reference line. For example, in Fig. 4, for the extracted segment QR, there are five reference lines between corners B and G. The perpendicular distances (green and red lines) from these segments to BC, DE, and FG are compared with their average distance. The red lines related to the points with high d_{p} are decided to be in an underlap area since these LiDAR points on the extracted segments are inside B_{r}.

Finally, we exclude the extracted points related to the extralap and underlap areas found previously, and the overall d_{avg} is recalculated using (19). Then, we find the errors (extralap or underlap) when between two consecutive true corner correspondences, there is only one extracted segment that corresponds to only one reference line. For example, in Fig. 4, for the extracted segment PV, there is only one reference line AH. The perpendicular distances (red lines) between PV and AH are compared with their average distance. These lines are related to the points with high d_{p}, therefore, are decided to be in an underlap or extralap areas depending on their positions inside or outside B_{r}. This last step is iteratively followed until no extralap or underlap is found between two consecutive corner correspondences where there is only one reference line corresponding to only one extracted segment.

To choose the value for c in real datasets, first, we visually identify 12 different extralap and underlap areas, shown in Fig. 5, as the ground truth from the buildings of three different test datasets. Fig. 6 plots the number of the identified segmented areas against different values of c. We can see that for $c = 3$, all 12 true extralap and underlap areas are identified correctly. For low values of c, we observe that many false segmented (extralap or overlap) areas are found, whereas for high values of c, many true extralap and underlap areas remain unidentified. Therefore, $c = 3$ is chosen in the algorithm. In our performance study, presented in the following section, $c = 3$ has been used for all synthetic and real test datasets.

IV. Performance Evaluation

To evaluate the proposed RCC metric, we first use some synthetic building polygons with their corresponding reference boundaries. Thereafter, we evaluate the extracted building boundaries from three real datasets with different LiDAR point densities [10]. We compare the proposed RCC metric with the existing metrics. Using our method, we also find the individual extralap and underlap areas, if exist, between the extracted and reference building polygons. Finally, we present the time complexity of the proposed RCC metric. We used the MATLAB

Fig. 5. Some extralap and underlap areas visually identified from three different datasets. Red polygons indicate reference and blue polygons represent extracted boundaries. Green solid circles show extralap and yellow stars represent the underlap areas.

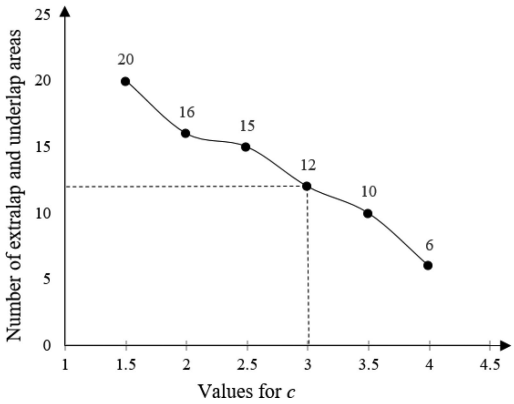

Fig. 6. Selection of a value for the parameter c. The number of identified segments plotted against different values of c.

platform (R2016a) to implement our method and conducted the experiments on a computer with Intel Core i7 2.6-GHz CPU and 16-GB RAM.

A. Datasets

We use Aitkenvale (AV) and Hervey Bay (HB) areas from the Australian benchmark datasets and Vaihingen (VH) area from the ISPRS benchmark datasets [10]. The AV datasets contain two different areas. The first area (AV1) covers 66 × 52 m^2 with a high point density (40 points per m^2) and contains five different buildings. The second area (AV2) contains 61 different buildings with point density 29 points per m^2. AV2 covers 214 × 159 m^2 area. The HB area has a medium point density (12 points per

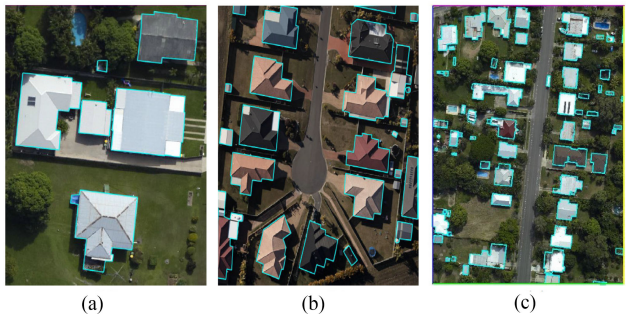

Fig. 7. Australian datasets with reference buildings of Aitkenvale (AV) and Hervey Bay (HB) areas: (a) AV1, (b) HB, and (c) AV2.

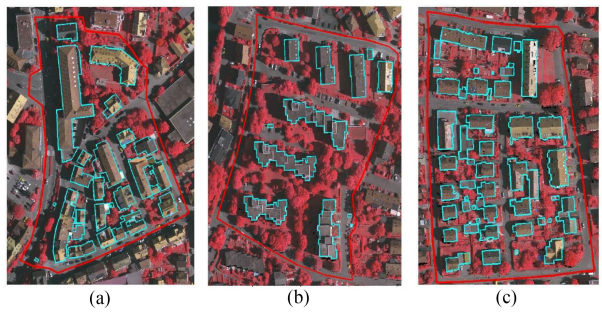

Fig. 8. ISPRS benchmark datasets with reference buildings of Vaihingen (VH) area: (a) VH1, (b) VH2, and (c) VH3.

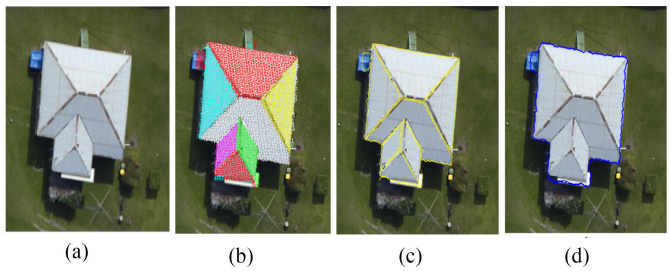

Fig. 9. Individual roof plane and building boundary extraction results using the method proposed by Dey *et al.* [53]: (a) Sample building. (b) Segmented LiDAR points of individual roof planes. (c) Extracted boundaries of individual roof planes (yellow). (d) Extracted building boundary (blue).

m^2) and covers 108 × 104 m^2 area with 28 different buildings. The ISPRS benchmark datasets have three different sites from Vaihingen (VH) area of Germany. We evaluated the buildings from all of these sites. The first site (VH1) consists of 37 different buildings with point density 3.5 points per m^2. The second site (VH2) has 14 large high rising buildings with 3.9 points per m^2 and the third site (VH3) contains 56 buildings with point density 3.9 points per m^2. Fig. 7 shows the three sites from the Australian benchmark datasets with the corresponding reference polygons and Fig. 8 shows the three sites of the ISPRS benchmark datasets with corresponding reference buildings by cyan polygons. The buildings with area less than 5 m^2 were not considered for evaluation in our experiments. We follow the method proposed by Dey *et al.* [53] for building boundary extraction. Fig. 9 shows the steps of extracting individual roof planes and the boundary of

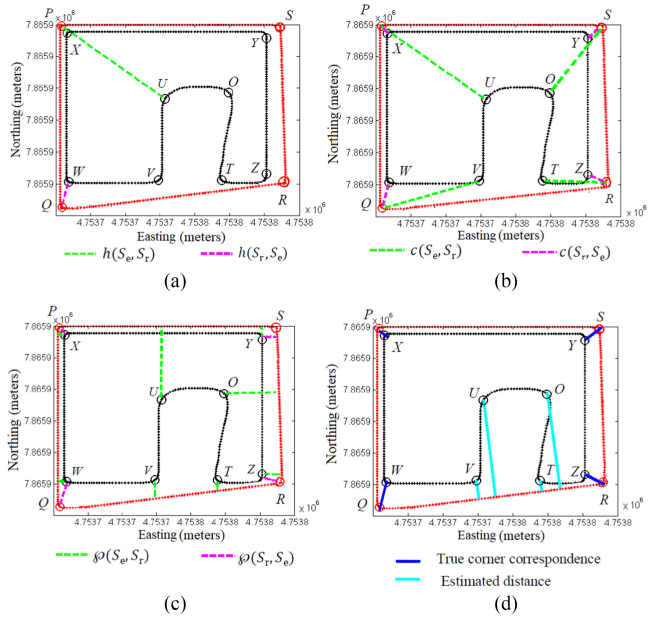

Fig. 10. Examples of finding distances using different metrics on a building with an underlap area $VUOT$. Note the RMSE, the NMAD, and the Chamfer find the same distance correspondences, and so, shown in the same figure. (a) Hausdroff (h). (b) Chamfer (c) or RMSE(λ) or NMAD (v). (c) PoliS ($\wp$). (d) RCC.

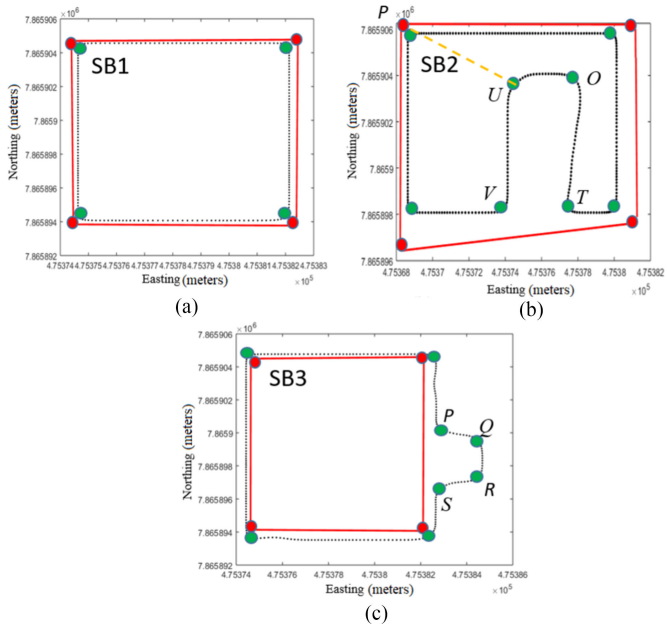

Fig. 11. Synthetic data: reference (red lines) and extracted (black dots) building polygons, where corners are shown by red and green solid circles, respectively. (a) Overlap area. (b) Underlap area (*VUOT*). (c) Extralap area (*PQRS*).

a sample building from the AV1 dataset using the method. The reference boundary polygons of buildings and individual roof planes have been extracted by monoscopic image measurement using the Barista software [19] (see Figs. 7, 8, and 13). Corners on an extracted building boundary have been detected by the method proposed in [54].

We also apply the proposed method to three simple synthetic building shapes with no extralap or underlap errors (SB1), with an underlap error (SB2), and an extralap error (SB3) generated from real point cloud data (see Fig. 11).

TABLE I
COMPARISON OF THE CALCULATED DISTANCES FOR SYNTHETIC BUILDING POLYGONS (SB1, SB2, AND SB3 IN FIG. 11)

Metric	SB1	SB2	SB3	Average
RCC	0.25	1.62	0.58	**0.81**
PoLiS	0.26	1.45	0.55	**0.75**
Hausdorff	0.32	5.48	1.68	**2.50**
Chamfer	0.25	1.41	0.59	**0.75**
RMSE	0.25	1.91	0.71	**0.99**
NMAD	0.23	1.42	0.54	**0.74**

All the distances are calculated in meters.

B. Experimental Results

First, we present and compare the estimated distances by different metrics when they are applied to evaluate the 2-D building polygons as well as the 3-D roof boundaries (both distances are estimated in 2-D). Second, we provide a sensitivity analysis when the number of points is varied in the building boundary or small geometric transformations are applied to the building boundary, and third, we present some visual results illustrating the identification of the individual extralap and underlap errors by the proposed RCC method.

1) Distance Estimation: For every pair of the reference and its corresponding extracted boundaries, we compute the distance using the proposed RCC metric and compare the result with the PoLiS [28], the Hausdorff, the Chamfer [28], [29], the RMSE [7], and the NMAD [20] distances. Fig. 10 shows some examples of the estimated distances using different metrics on the building shown in Fig. 2. The extracted building boundary has an underlap area $VUOT$, which means V, U, O, and T are false corners. For each metric, the distance is measured twice, (S_e, S_r) (from the extracted to the reference, green dashed lines) and (S_r, S_e) (from the reference to the extracted, magenta dashed lines). As can be seen in Fig. 10(a)–(d), the existing metrics mostly find wrong distances (green lines (S_e, S_r)) for the false corners. For instance, for the false corner U, they obtain the distance to either a noncorresponding reference corner (P using $h(S_e, S_r)$, $c(S_e, S_r)$, $\lambda(S_e, S_r)$, and $\nu(S_e, S_r)$) or a point (using $\wp(S_e, S_r)$) on the reference boundary. Clearly, these do not represent a realistic distance estimation using the existing metrics. In other word, the existing metrics do not work well if there is any segmentation error. In contrast, the proposed RCC metric is based on the true corner correspondences that are shown in Fig. 10(d) (blue lines). It finds the distances from the points on an extracted segment to its corresponding reference line. Therefore, these distances (see cyan lines from V, U, O, and T to QR) represent more realistic estimation.

For a fair comparison, we estimate the symmetrized distance for all metrics and present in this article. Table I shows the comparison of different metrics using three synthetic building polygons. We can see that for SB1, all metrics show almost the same result except the Hausdorff distance. This is because SB1 is

TABLE II
COMPARISON OF THE CALCULATED DISTANCES FOR DIFFERENT SITES OF AUSTRALIAN (AV1, AV2, AND HB) AND ISPRS (VH1, VH2, AND VH3) BENCHMARK DATASETS

Area	RCC	PoLiS	Hausdorff	Chamfer	RMSE	NAMD
AV1	0.22	0.21	0.53	0.21	0.36	0.26
AV2	0.58	0.53	1.41	0.54	0.63	0.44
HB	0.48	0.47	1.41	0.51	0.60	0.37
VH1	0.42	0.39	1.77	0.41	0.46	0.41
VH2	0.37	0.35	1.71	0.39	0.60	0.44
VH3	0.39	0.33	1.81	0.34	0.53	0.48
Average	**0.41**	**0.38**	**1.44**	**0.40**	**0.53**	**0.40**

All the distances are calculated in meters.

a very simple building polygon with no-segmentation error and the corresponding reference looks also almost the same as the extracted polygon. When the extracted and reference building polygons are almost the same, ideally the distance between the polygons should be close to zero, and so, each metric should calculate the similar evaluation result. However, the Hausdorff distance provides a different result because according to (9), it estimates the maximum among all the minimum distances from all the extracted points (S_e) to all the reference points (S_r). This is better explained and visualized using building SB2, where there is an underlap area as shown in Fig. 11. The Hausdorff distance finds the distance from U to P (marked by yellow dashed line) as the maximum among all the minimum distances from S_e to S_r. The same phenomenon is observed with SB3 where there is an extralap area as shown in Fig. 11(c). Therefore, the Hausdorff distance estimates the highest distance in all three synthetic buildings. The proposed RCC metric shows slightly larger than the PoLiS, the NMAD, and the Chamfer distances but slightly smaller than the RMSE distance. It is expected as the proposed metric considers the true corner correspondences as explained in Section III. We also observe that the Chamfer distance is very close to the PoLiS and the NMAD distances for all of the three polygons. It is because of the symmetric nature of the Chamfer, the NMAD, the PoLiS, and the proposed RCC metrics and this was clearly explained in Avbelj *et al.* [28].

Table II compares the results on the real datasets. Each value in this table represents the average result of all the buildings in a test datasets, except the last row that shows the average on all test datasets. Again, we can observe that the average distance by the RCC metric is close to the results of the PoLiS, the Chamfer, and the NMAD metrics.

Fig. 12 shows d_{avg} using the RCC for some selected buildings from three different sites of the test datasets with their corresponding extracted (cyan) and reference (blue) boundaries. The color bar is scaled to show from the best estimated distance (green) to the worst (red). We can see that while the three extracted buildings in Fig. 12(a), (b), and (e) are well estimated, the two extracted buildings in Fig. 12(d) and (f) are not. For the building in Fig. 12(b), although there are some underlap areas because of probable missing LiDAR data, the extracted boundary mostly coincides with the reference boundary. In contrast, the extralap areas in Fig. 12(f) because of similar height vegetation has contributed to the high RCC distance estimation.

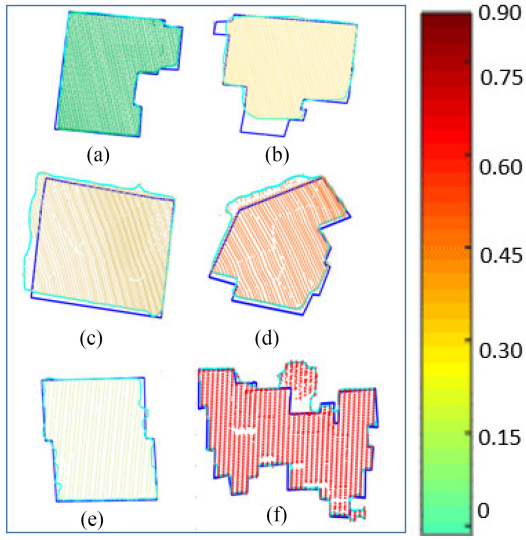

Fig. 12. Extracted and reference building polygons from: (a) and (b) Aitkenvale; (c) and (d) Hervey Bay; and (e) and (f): Vaihingen areas. Cyan polygons represent the extracted buildings and blue polygons indicate the reference buildings. The color bar indicates the distance estimation d_{avg} by the RCC metric and is scaled from the largest (red) to the smallest (green). The extracted LiDAR points within each extracted boundary are shown in the color that comes from the color bar according to the estimated RCC value using (19).

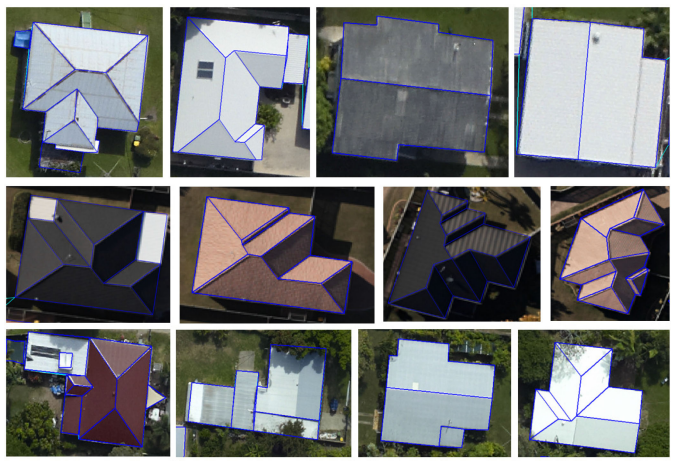

Fig. 13. Reference roof planes (marked with blue polygons) of selected buildings from three sites of Australian benchmark datasets. First row shows individual roof planes from AV1 area, and second and third rows show the roof planes of HB and AV2 areas, respectively.

The 3-D model reconstruction using a data-driven approach requires individual roof planes to be extracted properly [38], [55], [56]. The proposed RCC metric is also applicable to evaluate the 2-D shapes of the individual extracted roof planes of a building. For this purpose, we tested our method on individual roof planes of 12 selected buildings from the AV1, HB, and AV2 datasets. Fig. 13 shows these selected buildings with individual reference roof plane boundaries in blue polygons. We used the method proposed by Dey *et al.* [53] to extract the individual roof planes and apply the RCC metric to evaluate the extracted planes.

TABLE III
COMPARISON OF THE CALCULATED DISTANCES FOR EXTRACTED ROOF PLANES IN AUSTRALIAN BENCHMARK DATASETS

Metric	AV1	HB	AV2	Average
RCC	0.42	0.51	0.55	**0.49**
PoLiS	0.37	0.48	0.58	**0.47**
Hausdorff	0.92	1.32	1.09	**1.11**
Chamfer	0.38	0.52	0.55	**0.48**
RMSE	0.46	0.56	0.62	**0.55**
NMAD	0.29	0.37	0.48	**0.38**

All the distances are calculated in meters.

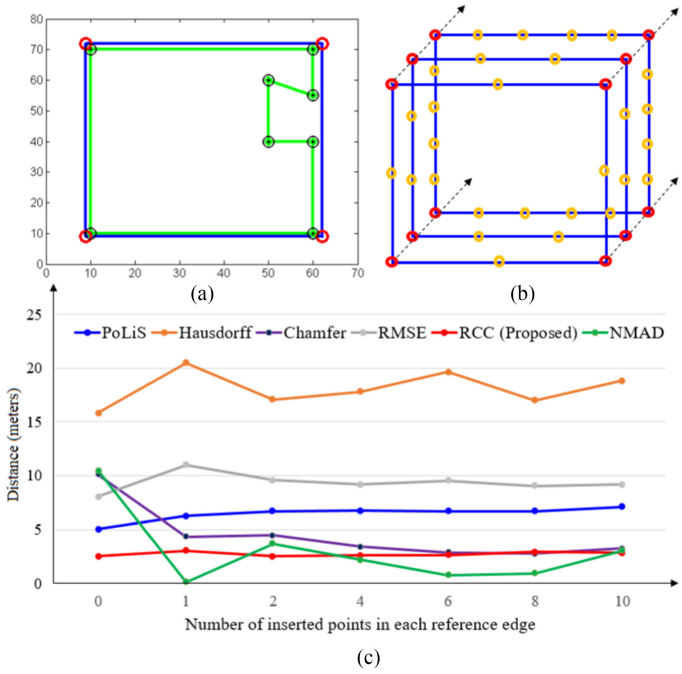

Fig. 14. Sensitivity of different metrics with respect to the change of the number of points in the reference boundary: (a) Green polygon represents the extracted boundary with eight points (black circles) and blue polygon represents reference polygon with four points (red circles). (b) First three examples of added points (orange circles) to the reference polygon. (c) Estimated distance by different metrics when the number of points in the reference boundary changes.

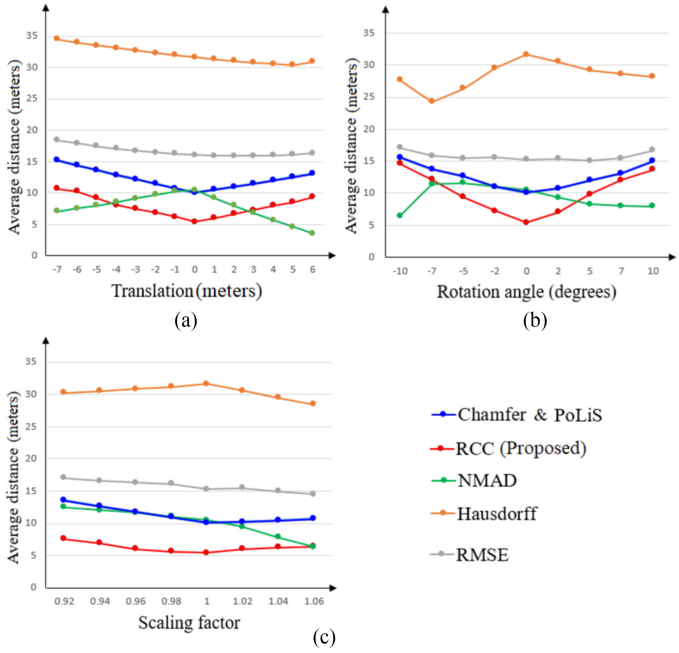

Fig. 15. Average distance using different metrics when different geometric transformations are applied to the reference polygon in Fig. 14(a): (a) translations (horizontal directions), (b) rotations, and (c) scaling.

Table III shows the average distances using different metrics including the proposed RCC metric. Like the building footprint extraction performance, the RCC, the PoLiS, and the Chamfer offer almost the same distance. The NMAD provides a shorter distance than these three. As usual, the Hausdorff distance is the highest among all the tested metrics.

2) Sensitivity Analysis: The proposed RCC metric is not sensitive to the number of points in the boundaries. It shows almost the same distance from the extracted polygon to the corresponding reference polygon irrespective of the number of points in the boundaries, which is expected. This situation can be explained using Fig. 14. We initially consider an extracted polygon (green) with eight points and corresponding reference (blue) building polygon with four points [see Fig. 14(a)]. To observe the performance, we manually add different number of points with equal distances in the reference polygon [orange, Fig. 14(b)] and each time calculate the distance using different metrics. Fig. 14(c) shows the calculated distances using different state-of-the-art metrics. The performance of the RCC metric (the linearity of the red line) indicates that it is not much sensitive to the number of additional points in the reference boundary, while the other metrics show different results with respect to the number of points in the boundary.

Like the PoLiS, in the case of small georeferencing errors, the performance of the RCC metric can also be approximated by a linear function. This means with the change of small geometric transformations, the estimated distance changes linearly in each situation, e.g., translate horizontally in the right direction, rotate clockwise, and scale up. Fig. 15 shows the three types of geometric transformations for the reference building polygon of Fig. 14(a), while the extracted building polygon is kept unchanged. During distance estimation, the metrics are not symmetrized, i.e., only the distance from the extracted points to the reference shape are considered. It can be observed that while the Chamfer, the PoLiS, and the RCC metrics mostly maintain the linear property, the other metrics show a parabolic nature.

3) Segmentation Errors: Fig. 16(a) shows some automatically detected extralap and underlap building areas (marked with green lines) by the proposed RCC method. The reference and extracted boundaries are shown in blue and cyan colors, respectively. Fig. 16(b) explains the detection process for an underlap area of a complex building from the Vaihingen area of the ISPRS datasets. The underlap area indicated in the left side of Fig. 16(b) is automatically marked with green lines, as shown in the right side of Fig. 16(b), by the proposed RCC method. For this building, the two consecutive true corner correspondences are (P, A) and (Q, B). Initially d_{avg} is calculated using (19)

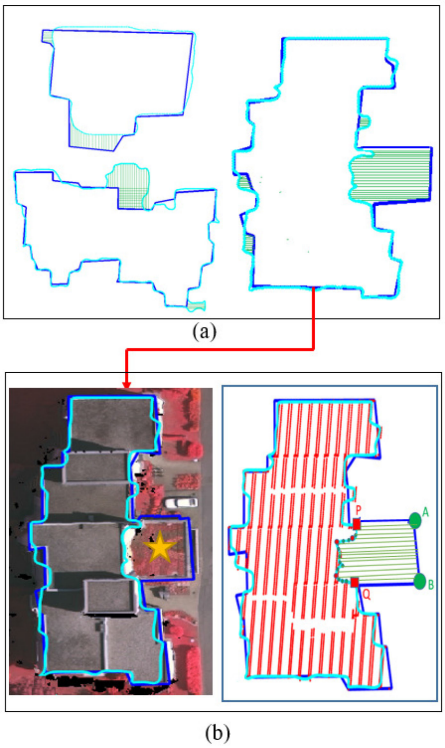

Fig. 16. Automatically obtained extralap and underlap areas: (a) Some sample buildings from different test datasets with reference (blue) and extracted (cyan) boundaries. Green lines represent perpendicular distances of the extralap and underlap areas. (b) Illustration for detection of an underlap area for a building from (a). Extracted LiDAR points are shown in red within the extracted (cyan) and reference (blue) boundaries. Green lines represent the perpendicular distances for the underlap area.

as 0.75 m. But the perpendicular distances (green lines) from the LiDAR points between corners P and Q to the reference line AB are in between 5 and 6 m. So, by using the technique presented in Section III-C, we find these points and consider this area as an underlap area since the LiDAR points are within the reference polygon.

C. Time Complexity

There are three main steps in the proposed RCC method: finding the corner correspondences (see Section III-A), estimating the distance (see Section III-B), and identifying the segmentation errors (see Section III-C). For a pair of an extracted building and its reference building, the extracted and reference boundaries (B_e and B_r) and the corners (ζ_{e} and ζ_{r}) are provided as inputs. In the first step, five M_f matrices are constructed for the five different properties we considered. The size of each of these matrices is $p \times q$, where $p = |\zeta_{\mathrm{r}}|$ and $q = |\zeta_{\mathrm{e}}|$. So, the construction of each M_f takes $O(pq)$ time. Their normalization and multiplication in (17) and (18) also take $O(pq)$ time. In the second step, the estimation of the RCC distance in (19) takes $O(n)$ time, where $n = |B_e|$. Similarly, the third step takes $O(n)$ time to compare the perpendicular distances with their average. Consequently, the total time complexity of the proposed RCC method is $O(pq + n)$. Since, in practice, $n \gg p$ and $n \gg q$, particularly, for large buildings and when the LiDAR point density is high, the best-case time complexity is $\Omega(n)$. However, when all points on B_e and B_r are corners, i.e., $n = q$ and $m = p$, the worst-case time complexity is $O(mn) \approx O(n^2)$, where $m = |B_r|$ and $m \approx n$.

V. Conclusion

To assess the quality of the extracted building boundaries from a dataset in a more realistic way is the main contribution of this article. For this purpose, we propose the RCC metric to estimate the average dissimilarity for an extracted building polygon compared to the corresponding reference building. It is a combined measure of the positional accuracy and shape similarity. We have tested its performance using several real datasets from three geographic areas and synthetic polygonal shapes where we obtained the expected result. Experimentally, it is shown that the proposed method provides the similar distance as the existing Chamfer and PoLiS distances. However, while the existing metrics offer unstable measurements in the presence of noise, segmentation errors, and variable number of points in the boundary, the RCC metric offers a quite stable performance.

Moreover, unlike the existing metrics, the proposed metric can be used to find individual underlap and extralap areas automatically. Some area-based metrics such as the area omission and commission errors Awrangjeb and Fraser [10] can estimate the overall extralap and underlap errors, respectively. So, they do not differentiate between the acceptable misalignment between the extracted and reference boundaries and the noticeable segmentation errors. In contrast, the proposed method makes differentiation between these two types of segmentation issues and indicates individual underlap and extralap areas.

We used the proposed metric to evaluate the shapes of the 2-D extracted building and plane boundaries. However, a further investigation can be done to extend the proposed RCC method to apply it to evaluate the 3-D building shapes. In this case, to calculate the corner correspondences between the reference and extracted polyhedrons, the five properties in Section III-A need to be modified to find true 3-D corner correspondences. After obtaining the corner correspondences, the orthogonal distances from the extracted plane points to the reference planes would be estimated. Moreover, we have so far considered the residential buildings that are mainly in simple and complex rectilinear shapes. We have not considered the nonrectilinear complex buildings, e.g., round-shaped buildings with nonplanar roof parts or building with atrium. Our future research could include a full investigation to evaluate such buildings and planes.

References

[1] S. Gilani, M. Awrangjeb, and G. Lu, "Segmentation of airborne point cloud data for automatic building roof extraction," *GISci. Remote Sens.*, vol. 55, no. 1, pp. 63–89, 2018.

[2] S. Shirowzhan, S. M. Sepasgozar, H. Li, J. Trinder, and P. Tang, "Comparative analysis of machine learning and point-based algorithms for detecting 3D changes in buildings over time using bi-temporal LiDAR data," *Autom. Construction*, vol. 105, 2019, Art. no. 102841.

[3] D. Chen, L. Zhang, P. T. Mathiopoulos, and X. Huang, "A methodology for automated segmentation and reconstruction of urban 3-D buildings from ALS point clouds," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 7, no. 10, pp. 4199–4217, Oct. 2014.

[4] M. Feng, T. Zhang, S. Li, G. Jin, and Y. Xia, "An improved minimum bounding rectangle algorithm for regularized building boundary extraction from aerial LiDAR point clouds with partial occlusions," *Int. J. Remote Sens.*, vol. 41, no. 1, pp. 300–319, 2020.

[5] L. Hao, Y. Zhang, and Z. Cao, "Active cues collection and integration for building extraction with high-resolution color remote sensing imagery," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 12, no. 8, pp. 2675–2694, Aug. 2019.

[6] 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. Remote Sens.*, vol. 2, no. 1, pp. 11–20, Mar. 2009.

[7] I. F. Shaker, A. Abd-Elrahman, A. K. Abdel-Gawad, and M. A. Sherief, "Building extraction from high resolution space images in high density residential areas in the great Cairo region," *Remote Sens.*, vol. 3, no. 4, pp. 781–791, 2011.

[8] A. Erener, "Classification method, spectral diversity, band combination and accuracy assessment evaluation for urban feature detection," *Int. J. Appl. Earth Observ. Geoinf.*, vol. 21, pp. 397–408, 2013.

[9] F. Rottensteiner, J. Trinder, S. Clode, and K. Kubik, "Using the Dempster–Shafer method for the fusion of LiDAR data and multi-spectral images for building detection," *Inf. Fusion*, vol. 6, no. 4, pp. 283–300, 2005.

[10] M. Awrangjeb and C. S. Fraser, "An automatic and threshold-free performance evaluation system for building extraction techniques from airborne LiDAR data," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 7, no. 10, pp. 4184–4198, Oct. 2014.

[11] J. Avbelj and R. Müller, "Quality assessment of building extraction from remote sensing imagery," in *Proc. IEEE Geosci. Remote Sens. Symp.*, 2014, pp. 3184–3187.

[12] M. Potučková and P. Hofman, "Comparison of quality measures for building outline extraction," *Photogrammetric Rec.*, vol. 31, no. 154, pp. 193–209, 2016.

[13] 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 Remote Sens.*, vol. 62, no. 2, pp. 135–149, 2007.

[14] M. Awrangjeb and C. S. Fraser, "Automatic segmentation of raw LiDAR data for extraction of building roofs," *Remote Sens.*, vol. 6, no. 5, pp. 3716–3751, 2014.

[15] D. S. Lee, J. Shan, and J. S. Bethel, "Class-guided building extraction from IKONOS imagery," *Photogrammetric Eng. Remote Sens.*, vol. 69, no. 2, pp. 143–150, 2003.

[16] 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. Remote Sens.*, vol. 6, no. 3, pp. 1640–1652, Jun. 2013.

[17] T. H. Nguyen, S. Daniel, D. Gueriot, C. Sintes, and J.-M. L. Caillec, "Unsupervised automatic building extraction using active contour model on unregistered optical imagery and airborne LiDAR data," 2019, *arXiv:1907.06206*.

[18] M. Awrangjeb, M. Ravanbakhsh, and C. S. Fraser, "Automatic detection of residential buildings using LiDAR data and multispectral imagery," *ISPRS J. Photogrammetry Remote Sens.*, vol. 65, no. 5, pp. 457–467, 2010.

[19] M. Awrangjeb, C. Zhang, and C. S. Fraser, "Automatic extraction of building roofs using LiDAR data and multispectral imagery," *ISPRS J. Photogrammetry Remote Sens.*, vol. 83, pp. 1–18, 2013.

[20] X. Gong, L. Shen, and T. Lu, "Refining training samples using median absolute deviation for supervised classification of remote sensing images," *J. Indian Soc. Remote Sens.*, vol. 47, no. 4, pp. 647–659, 2019.

[21] K. Bittner, M. Körner, and P. Reinartz, "DSM building shape refinement from combined remote sensing images based on WNET-CGANS," in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, 2019, pp. 783–786.

[22] T. Partovi, F. Fraundorfer, R. Bahmanyar, H. Huang, and P. Reinartz, "Automatic 3-D building model reconstruction from very high resolution stereo satellite imagery," *Remote Sens.*, vol. 11, no. 14, 2019, Art. no. 1660.

[23] O. Henricsson and E. Baltsavias, "3-D building reconstruction with ARUBA: A qualitative and quantitative evaluation," in *Automatic Extraction of Man-Made Objects From Aerial and Space Images (II)*. Berlin, Germany: Springer, 1997, pp. 65–76.

[24] Y. A. Mousa, P. Helmholz, D. Belton, and D. Bulatov, "Building detection and regularisation using DSM and imagery information," *Photogrammetric Rec.*, vol. 34, no. 165, pp. 85–107, 2019.

[25] J. L. Dungan, "Focusing on feature-based differences in map comparison," *J. Geographical Syst.*, vol. 8, no. 2, pp. 131–143, 2006.

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

[27] S. Jozdani and D. Chen, "On the versatility of popular and recently proposed supervised evaluation metrics for segmentation quality of remotely sensed images: An experimental case study of building extraction," *ISPRS J. Photogrammetry Remote Sens.*, vol. 160, pp. 275–290, 2020.

[28] J. Avbelj, R. Müller, and R. Bamler, "A metric for polygon comparison and building extraction evaluation," *IEEE Geosci. Remote Sens. Lett.*, vol. 12, no. 1, pp. 170–174, Jan. 2015.

[29] J. Zhang, R. Collins, and Y. Liu, "Representation and matching of articulated shapes," in *Proc. IEEE Comput. Soc. Conf. Comput. Vision Pattern Recognit.*, 2004, vol. 2, p. II.

[30] K. Liu, H. Ma, L. Zhang, Z. Cai, and H. Ma, "Strip adjustment of airborne LiDAR data in urban scenes using planar features by the minimum Hausdorff distance," *Sensors*, vol. 19, no. 23, 2019, Art. no. 5131.

[31] L. Xie *et al.*, "Hierarchical regularization of building boundaries in noisy aerial laser scanning and photogrammetric point clouds," *Remote Sens.*, vol. 10, no. 12, 2018, Art. no. 1996.

[32] H. Oriot and A. Michel, "Building extraction from stereoscopic aerial images," *Appl. Opt.*, vol. 43, no. 2, pp. 218–226, 2004.

[33] R. Dos Santos, M. Galo, and A. Carrilho, "Building boundary extraction from LiDAR data using a local estimated parameter for alpha shape algorithm," *Int. Arch. Photogrammetry, Remote Sens. Spatial Inf. Sci.*, vol. 42, no. 1, pp. 127–132, 2018.

[34] Y. Wei, Z. Zhao, and J. Song, "Urban building extraction from high-resolution satellite panchromatic image using clustering and edge detection," in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, 2004, vol. 3, pp. 2008–2010.

[35] B. Xiong, S. O. Elberink, and G. Vosselman, "A graph edit dictionary for correcting errors in roof topology graphs reconstructed from point clouds," *ISPRS J. Photogrammetry Remote Sens.*, vol. 93, pp. 227–242, 2014.

[36] M. Awrangjeb, "Using point cloud data to identify, trace, and regularize the outlines of buildings," *Int. J. Remote Sens.*, vol. 37, no. 3, pp. 551–579, 2016.

[37] Y. Chen, L. Tang, X. Yang, M. Bilal, and Q. Li, "Object-based multi-modal convolution neural networks for building extraction using panchromatic and multispectral imagery," *Neurocomputing*, vol. 386, pp. 136–146, 2019.

[38] M. Awrangjeb, S. A. N. Gilani, and F. U. Siddiqui, "An effective data-driven method for 3-D building roof reconstruction and robust change detection," *Remote Sens.*, vol. 10, no. 10, 2018, Art. no. 1512.

[39] L. Li, J. Yao, J. Tu, X. Liu, Y. Li, and L. Guo, "Roof plane segmentation from airborne LiDAR data using hierarchical clustering and boundary relabeling," *Remote Sens.*, vol. 12, no. 9, 2020, Art. no. 1363.

[40] H. Yang, P. Wu, X. Yao, Y. Wu, B. Wang, and Y. Xu, "Building extraction in very high resolution imagery by dense-attention networks," *Remote Sens.*, vol. 10, no. 11, 2018, Art. no. 1768.

[41] D. M. McKeown, S. D. Cochran, S. J. Ford, J. C. McGlone, J. A. Shufelt, and D. A. Yocum, "Fusion of HYDICE hyperspectral data with panchromatic imagery for cartographic feature extraction," *IEEE Trans. Geosci. Remote Sens.*, vol. 37, no. 3, pp. 1261–1277, May 1999.

[42] J. Sreevalsan-Nair, A. Jindal, and B. Kumari, "Contour extraction in buildings in airborne LiDAR point clouds using multiscale local geometric descriptors and visual analytics," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 11, no. 7, pp. 2320–2335, Jul. 2018.

[43] F. Dornaika, A. Moujahid, Y. El Merabet, and Y. Ruichek, "Building detection from orthophotos using a machine learning approach: An empirical study on image segmentation and descriptors," *Expert Syst. Appl.*, vol. 58, pp. 130–142, 2016.

[44] S. O. Elberink and G. Vosselman, "Quality analysis on 3D building models reconstructed from airborne laser scanning data," *ISPRS J. Photogrammetry Remote Sens.*, vol. 66, no. 2, pp. 157–165, 2011.

[45] D. Akca, M. Freeman, I. Sargent, and A. Gruen, "Quality assessment of 3D building data," *Photogrammetric Rec.*, vol. 25, no. 132, pp. 339–355, 2010.

[46] B. He, Z. Lin, and Y. F. Li, "An automatic registration algorithm for the scattered point clouds based on the curvature feature," *Opt. Laser Technol.*, vol. 46, pp. 53–60, 2013.

[47] D. P. Huttenlocher and W. J. Rucklidge, "A multi-resolution technique for comparing images using the Hausdorff distance," Cornell Univ., Ithaca, NY, USA, Tech. Rep. TR-92-1321, 1992.

[48] D. H. Lee, K. M. Lee, and S. U. Lee, "Fusion of LiDAR and imagery for reliable building extraction," *Photogrammetric Eng. Remote Sens.*, vol. 74, no. 2, pp. 215–225, 2008.

[49] 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 Remote Sens.*, vol. 93, pp. 256–271, 2014.
[50] J. Höhle and M. Höhle, "Accuracy assessment of digital elevation models by means of robust statistical methods," *ISPRS J. Photogrammetry Remote Sens.*, vol. 64, no. 4, pp. 398–406, 2009.
[51] K. M. Masoud, C. Persello, and V. A. Tolpekin, "Delineation of agricultural field boundaries from Sentinel-2 images using a novel super-resolution contour detector based on fully convolutional networks," *Remote Sens.*, vol. 12, no. 1, 2020, Art. no. 59.
[52] Y. Liu *et al.*, "Discrepancy measures for selecting optimal combination of parameter values in object-based image analysis," *ISPRS J. Photogrammetry Remote Sens.*, vol. 68, pp. 144–156, 2012.
[53] E. K. Dey, M. Awrangjeb, and B. Stantic, "Outlier detection and robust plane fitting for building roof extraction from LiDAR data," *Int. J. Remote Sens.*, vol. 41, no. 16, pp. 6325–6354, 2020.
[54] M. Awrangjeb, G. Lu, C. S. Fraser, and M. Ravanbakhsh, "A fast corner detector based on the chord-to-point distance accumulation technique," in *Proc. Digit. Image Comput.: Techn. Appl.*, Dec. 2009, pp. 519–525.
[55] S. Canaz Sevgen and F. Karsli, "An improved RANSAC algorithm for extracting roof planes from airborne LiDAR data," *Photogrammetric Rec.*, vol. 35, no. 169, pp. 40–57, 2020.
[56] F. Tarsha Kurdi and M. Awrangjeb, "Automatic evaluation and improvement of roof segments for modelling missing details using LiDAR data," *Int. J. Remote Sens.*, vol. 41, no. 12, pp. 4702–4725, 2020.

Emon Kumar Dey received the B.Sc. (Hons.) and M.S. degrees in computer science and engineering from the University of Dhaka, Dhaka, Bangladesh, in 2009 and 2011, respectively. He is currently working toward the Ph.D. degree with Griffith University, Nathan, Australia.

He is currently an Assistant Professor with the Institute of Information Technology (IIT), University of Dhaka. His research interests include automatic building extraction, image processing, pattern recognition, remote sensing data processing, and machine learning.

Mohammad Awrangjeb (Senior Member, IEEE) received the Ph.D. degree from Monash University, Clayton, VIC, Australia.

He is a Senior Lecturer with Griffith University, Nathan, Australia. Before joining Griffith University as a Lecturer, he worked as a (Senior) Research Fellow with Federation University, Monash University, and University of Melbourne. His research interest includes automatic feature extraction and matching, multimedia security and image processing.

Dr. Awrangjeb was the recipient of the Discovery Early Career Researcher Award by the Australian Research Council (2012–2015).