

An automatic and threshold-free performance evaluation system for building extraction techniques from airborne LIDAR data

Mohammad Awrangjeb and Clive S. Fraser

Abstract—Some performance evaluation systems for building extraction techniques are manual in the sense that only visual results are provided or human judgement is employed. Many evaluation systems that employ one or more thresholds to ascertain whether an extracted building or roof plane is correct are subjective and cannot be applied in general. There are only a small number of automatic and threshold-free evaluation systems, but these do not necessarily consider all special cases, for example, when over- and under-segmentation occurs during the extraction of roof planes. This paper proposes an automatic and threshold-free evaluation system that offers robust object-based evaluation of building extraction techniques. It makes one-to-one correspondences between extracted and reference entities using the maximum overlaps. Its application to the evaluation of a building extraction technique shows that it estimates different performance indicators including segmentation errors. Consequently, it can be employed for bias-free evaluation of other techniques whose outputs consist of polygonal entities.

Index Terms—Building, roof, extraction, detection, reconstruction, automatic, threshold-free, performance, evaluation.

I. INTRODUCTION

Building detection and reconstruction from remotely sensed data is important to the real estate industry, city planning, homeland security, disaster (flood or bushfire) management and many other applications. Building detection refers to the problem of identification of buildings in remotely sensed data such as aerial imagery and LIDAR (Light Detection And Ranging) data. Many building detection techniques also concentrate on accurate delineation of building boundaries. Building reconstruction implies the extraction of 3D building information, which includes corners, edges and planes of the building facades and roof. Digital reconstruction of the facades and roof then follows using the available information. The research reported in this paper concentrates on evaluation of building detection and roof plane extraction. A large number of building detection and 3D roof extraction techniques have been reported over the last few decades. Reviews of building detection and roof extraction techniques can be found in Awrangjeb et al. [1], Haala and Kada [2] and Wang [3].

The necessity for proper quality descriptions has been embedded in the fields of geodesy and photogrammetry for a

long time [4]. In fact, proper quality descriptions are essential for accurate quality assessment of the detected buildings and extracted roof planes. However, commonly accepted and widely used evaluation systems are lacking for both building detection and roof extraction techniques. Uniform and rigorous evaluation systems are hard to find, and there is an absence of standards [5].

An evaluation system may assess a building detection or roof extraction technique using pixel- [5], [6], [7] and/or object-based [5] metrics. While the latter quantifies the number of buildings and offers a quick assessment, the former (also known as area-based) is based on the number of pixels within the extracted buildings and provides a more rigorous evaluation [8]. The pixel-based evaluation indirectly corresponds to the horizontal accuracy of the detected building footprints. While Song and Haithcoat [8] and Shufelt [9] preferred pixel-based metrics over object-based metrics, Foody [10] pointed out problems associated with pixel-based metrics, as they may be severely affected by misalignment (due to sensor resolution, registration error etc.) between the reference and detected objects. The shape similarity metrics employed in [11] are similar to the area- or pixel-based indices in [1]. For example, area difference in [11] is same as the area omission and commission errors in [1] and overlap error in [11] is similar to branching and miss factors in [1]. Many evaluation systems also use geometric indices in order to assess the planimetric and height accuracy of the extracted objects. A review of different evaluation criteria can be found in [12].

There are a number of issues related to the current evaluation systems. Many evaluation schemes [13], [14] do not involve any objective criteria. Instead, visual results for a limited number of buildings are used for quality assessment. The use of 1 to 2 evaluation indices only has characterized many studies [15], [16], [17]. There are also manual evaluation systems [18] that are based on human judgement as to whether roofs, for example, are partially or fully extracted or not. These manual evaluation systems are very subjective and therefore the results may be biased. They are also expensive in time when applied to large data sets. Evaluation systems [5], [6], [7], [19], [20] that involve one or more thresholds for deciding whether an extracted building or roof plane is correct or not may be biased. The use of such thresholds does not work well in all cases, and this approach can be controversial since there is no unique way to select these thresholds [9]. Moreover, the estimated performance may be drastically affected when the areas of the extracted object boundaries change. There are

M. Awrangjeb is with the Gippsland School of Information Technology, Monash University, Churchill Vic 3842, Australia e-mail: mohammad.awrangjeb@monash.edu (see <http://users.monash.edu.au/mawrangj/>).

C. S. Fraser is with the Cooperative Research Centre for Spatial Information, Dept. of Infrastructure Engineering, University of Melbourne, Parkville Vic 3010, Australia e-mail: c.fraser@unimelb.edu.au

Manuscript received April dd, yyyy; revised December dd, yyyy.

many evaluation methods [18], [21], [22] which evaluate the reconstructed roof models based on the number of buildings, not on the number of planes. Such results do not reflect the complete scenario of the reconstructed roof planes. Some systems, including the ones described in [11], [23], [12], do not include sufficient details of how the correspondences are made between the reference and extracted plane sets. Some evaluation systems [5] exploit improper quality formula that may be mathematically undefined in certain situations. Last but not least, many of the systems reported to date for evaluation of roof models do not consider the area (number of pixels) of the reconstructed roof planes. Since over- and under-segmentations frequently occur in algorithms applied to LIDAR and/or image data, it is common not to obtain the exact plane boundary. An area- or pixel-based evaluation can reflect this scenario.

This paper proposes an automatic and threshold-free evaluation system that does not involve any human judgement or threshold setting. The approach considers both object- and pixel-based evaluations as well as geometric accuracy. In addition to completeness, correctness and quality metrics, the system employs detection and reference cross-lap rates, miss and branching factors in order to reflect the cases of over- and under-segmentations. The root-mean-square error (RMSE) is used to fully reflect the geometric accuracy in planimetry and height. The directional and displacement errors are also measured for more rigorous evaluation of the extracted planes.

The proposed evaluation system has been tested on an evaluation of the performance of a recently developed building extraction technique [24], [25]. Three areas from the ISPRS (International Society for Photogrammetry and Remote Sensing) benchmark data sets [26] have been employed in the evaluation. In order to show the robustness of the proposed system, the roof plane boundaries have been extracted in two different ways by the building extraction technique being evaluated so that the two differently extracted boundaries of a plane can differ by 7 to 9 m² in area. While compared with the threshold-based system [5], that has been widely adopted for the ISPRS benchmark data sets, the proposed system offers stable object-based evaluation results when the extracted plane area has been changed. In contrast, the object-based performance offered by Rutzinger et al. [5] has changed considerably.

The rest of the paper is organised as follows. Section II briefly presents the evaluation systems that have been reported in the literature. Section III summarises the building roof extraction technique [24] being evaluated in this research. Section IV describes the proposed evaluation system. Evaluation results and comparisons with Rutzinger et al. [5] are presented in Section V, and concluding remarks are provided in Section VI.

II. REVIEW

Performance evaluation systems for building detection and roof extraction techniques can be divided into two major groups: those using one or more overlap thresholds indicating the minimum degree of overlap (between a reference and

detected building) required to decide a true positive entity [6], [7], [5] and those not using any thresholds [9], [27]. In this section, a brief summary of some of the evaluation systems of both types is presented. Interested readers are referred to detailed review reports in [28], [10] and [5]. Note that the voxel-based systems [29], which mainly evaluate 3D complete building models, are not considered in this paper.

A. Threshold-based systems

In Rottensteiner et al. [6] and Rutzinger et al. [5], a correspondence was established between a detected building and a reference building if they overlapped each other either strongly, by more than 80% overlap, or partially, by 50% to 80% overlap. In order to establish one-to-one correspondences between the detected and reference building sets, a split-and-merge technique was proposed [5]. This threshold-based system was later applied to evaluate roof plane extraction techniques [26].

You and Lin [20] applied a tolerance circle of radius $K\sigma$ to decide whether the boundary of a planar segment was correctly reconstructed, where $K = 2$ was a robust weighting factor and $\sigma = 0.58\text{m}$ was the standard deviation between the registered LIDAR boundary and the boundary from topographic map data. Otherwise, it was decided that the boundary was incorrectly reconstructed and would need a manual refinement step. A similar threshold was used by Akca et al. [30]. In fact, the value of K has to be carefully chosen in these systems in order to control omission (Type I) and commission (Type II) errors [31]. For example, for high values of K , Type I errors decrease, while Type II errors increase. The system proposed in Cheng et al. [21] used both angle (4°) and similar distance (0.5 m) thresholds to decide whether an extracted boundary of a building roof was correct.

Oude Elberink and Vosselman [4] presented an approach to assessment of the geometric quality of 3D building models without the use of any reference measurements. The following three criteria were analysed: the orthogonal distance from the LIDAR points to their corresponding roof planes, the shortest distance from the corners of the 3D models to the nearest LIDAR points on the corresponding LIDAR segments, and the LIDAR segments that had been used or removed during reconstruction of the 3D models. Height and distance thresholds were applied to evaluate the reconstructed planes and roof models and experimental results showed that around 4-7% of the LIDAR segments were removed, which affected around 15-22% of the reconstructed building models. In the absence of a reference data set, it is not possible to verify the third criterion. In fact, the primary shortcoming of the evaluation system reported in Oude Elberink and Vosselman [4] is that without using a reference data set, an independent description of absolute accuracy is precluded. Satari et al. [32] used support vector machines (SVM) to verify the extracted planes. The use of SVM may be infeasible for automatic evaluation: first, it requires training data sets and second, the order of the polynomial kernel has to be set beforehand. In addition to following the system in [5] for evaluation of the segmentation quality and geometric errors, Xiong et al. [33]

proposed a single criterion for the evaluation of a complete building model without the use of reference building models. If a segmented LIDAR point was more than 0.3 m away from the reconstructed building model, it was considered as a deviated point. By using a connected component analysis, if a deviated region is smaller than 3 m² in area it was ignored as an error and the reconstructed model was accepted; otherwise, the model was corrected with the help of a manual graph editing tool.

B. Threshold-free systems

Without using a particular overlap threshold, Shufelt [9] showed building detection performance graphically as the overlapped area varied from 0-100%. Shan and Lee [27] presented results by histograms showing the frequency of buildings as functions of *underlap*, *overlap*, *extralapl*, *crosslap*, and *fitness*. The number of *false negative* buildings was indicated by the frequency at 100% *underlap* and the number of *false positive* buildings was indicated by the frequency both at *crosslap* 0 and 0% *fitness*.

Pfeifer et al. [34] used a threshold-free system in which it was checked whether the centres of the detected building footprints fell within the reference building footprints. All buildings successfully passing this test were assumed to be correctly classified. Rutzinger et al. [5] also presented a similar system based on the centroid-in-polygon approach, which can result in unequal numbers of true positive entities in the detected and reference building sets. Moreover, the original centroid of a polygon may reside outside the polygon and the algorithm looking for a pseudo centroid that resides inside the polygon may not converge.

Awrangjeb et al. [35] proposed an automatic and threshold-free system that uses the distance between the centres of two rectangular detected buildings which overlap each other. If a reference building is overlapped by more than one detected building, then the nearest detected building is chosen for the reference building. However, this system has the limitation that an actual building boundary cannot be represented by a rectangle. A similar system was later used in [36], where the centre-distance was replaced by the number of overlapping pixels. Thus, they were able to apply the evaluation system for planes having polygonal boundaries.

Jochem et al. [37] used the centre of each reference plane as a reference point and the reference points were checked to ascertain whether they resided within the detected planes. However, in the case of over segmentation, when a reference plane may correspond to more than one detected segment, the corresponding reference point can be within a segment that has smaller overlap with the reference plane than other segments. This can result in some inappropriate correspondences between the detected and reference planes.

C. Contributions

- An automatic and overlap threshold-free evaluation system is presented in this paper. The selection of one or more overlap thresholds by threshold-based systems is deemed too subjective. Although Rutzinger et al.

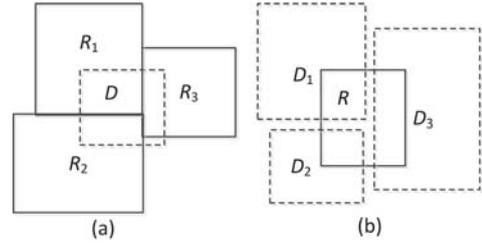


Fig. 1. Examples of estimating ‘overall coverage’ and possible (a) one-to-zero (1:0) and (b) zero-to-one (0:1) relations between extracted and reference entities using the threshold-based evaluation system [6]. Solid rectangles indicate reference entities and dashed rectangles indicate extracted entities.

[5] found that the selection of the overlap thresholds marginally affected the evaluation results, the reported research in this paper has shown that the evaluation results may be severely affected when the plane boundaries are extracted in two different ways by the building extraction technique (Section V-C).

- Rutzinger et al. [5] follow an estimation of overlap between extracted and reference entities from Rottensteiner et al. [6] that they term as ‘overall coverage’. As shown in Fig. 1a, Extracted Entity *D* has more than 50% overlap with Reference Entities *R*₁, *R*₂ and *R*₃. In this case, *D* is marked as *true positive* (TP), but all three reference entities are marked as *false negative* (FN) since none of them has more than 50% coverage with *D* (i.e., 1:0 relation). The opposite happens in Fig. 1b, where Reference Plane *R* is marked as a TP but all three extracted planes are marked as *false positive* (FP) (i.e., 0:1 relation).

Such an estimation using ‘overall coverage’ is acceptable for evaluation of building detection results, but is not suitable for evaluation of roof plane extraction results. Because while a building boundary has a 2D nature, a roof plane boundary does not, except for a horizontal plane. As shown in Fig. 2a, two neighbouring buildings are extracted as a single building and an estimation of ‘overall coverage’ would find that both reference buildings have been detected. Thus, a split operation is sufficient to obtain the two separate building boundaries (Fig. 2b). However, Fig. 2c shows that Planes *A* and *C* are extracted as a single Plane *A*. Here, the assessment based on ‘overall coverage’ that Reference Plane *C* (Fig. 2d) has been extracted is wrong. This is because due to 3D nature, pixels that are outside Reference Plane *A* but inside Extracted Plane *A* physically do not reside within Reference Plane *C*.

This paper does not consider ‘overall coverage’ in order to avoid the above erroneous estimation. Instead, it finds correspondences using the maximum overlap.

- Since planes on a building roof are physically connected to each other, many evaluation systems including the one in Rottensteiner et al. [6] that is based on ‘overall coverage’ find many one-to-many, many-to-one and many-to-many relationships when the overlap between any two reference and extracted planes exceeds the overlap thresh-

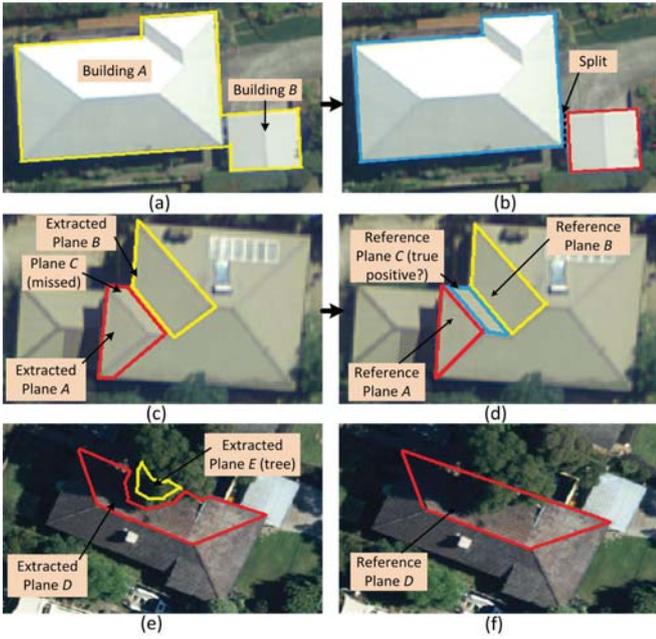


Fig. 2. Evaluation of building detection and plane extraction results by Rottensteiner et al. [6]: (a-b) Split operation for two neighbouring buildings to find two separate boundaries, (c-d) one-to-many relation from Extracted Plane A to Reference Planes A and C, and (e-f) many-to-one relation from Extracted Planes D and E to Reference Plane D.

old. For illustration, as shown in Fig. 2c, let a LIDAR-based roof plane extraction algorithm extract Planes A and B and extend Plane A towards Plane C (i.e., area commission error) to accommodate the error in LIDAR data. Consequently, Plane C is missed since it is small in size. Thus, Reference Plane C in Fig. 2d should be marked as FN by the evaluation system. However, Rottensteiner et al. [6] finds it as TP and establishes a one-to-many relation from Extracted Plane A to Reference Planes A and C. This wrong assessment increases the number of TP entities in the reference plane set (TPr). Again, as shown in Figs. 2e-f, let the roof plane extraction algorithm extract Planes D and E for Reference Plane D. Since, Extracted Plane E is on a tree, it should be marked as FP by the evaluation system. However, Rottensteiner et al. [6] finds it as TP and establishes a many-to-one relation from Extracted Planes D and E to Reference Plane D. This wrong assessment also increases the number of true positive entities in the extracted plane set (TPe). The many-to-many relations may increase TP entities in both extracted and reference plane sets. Thus, TPr and TPe used for computing completeness and correctness, respectively, are not of equal size and the estimated completeness, correctness and quality values are found much higher than their actual values. In this paper, only the one-to-one correspondences are considered based on the first and second largest overlaps between the extracted and reference planes. Thus, the number of TPs are same in the extracted and reference sets (Section IV-B).

- The modified quality formula (Eq. 1) used in [5] will be

undefined when both TPr and TPe are empty.

$$Quality = \frac{1}{1 + \frac{\|FN\|}{\|TPr\|} + \frac{\|FP\|}{\|TPe\|}}. \quad (1)$$

This will happen when no correspondences can be established between the extracted and reference plane sets. Moreover, Rutzinger et al. [5] showed examples using centroid-in-polygon tests where TPr is not empty, but TPe is empty, and vice versa (0:1 and 1:0 relations). The same may happen in calculation of TP, FP and FN entities based on ‘overall coverage’ [6] as shown in Fig. 1. Therefore, this paper only establishes one-to-one correspondences and suggests to use the original quality formula:

$$Quality = \frac{\|TP\|}{\|TP\| + \|FP\| + \|FN\|}. \quad (2)$$

- Many evaluation systems apply only a small number of evaluation metrics [15], [16], [17]. This paper presents a set of comprehensive evaluation metrics in three categories: object-based, pixel-based and geometric.

III. BUILDING ROOF EXTRACTION TECHNIQUE

Two different approaches have been followed for building extraction from remotely sensed data. Methods in the first approach extract building regions along with many other objects during land cover classification [38]. They first extract features (colour, texture, height) from the input data and then classify the features into different objects (building, grass, bare-earth, tree, road, water). For classification while some authors employed a single classifier, e.g., Dempster-Shafer [6], support vector machine [39], [40] and supervised maximum likelihood [38], others [41] fused the decisions from multiple classifiers for improved classification accuracy. Methods in the second approach explicitly extract buildings and remove other objects (tree, ground, bushes) from the input data. These methods apply segmentation algorithms on the input data and sequentially remove the unwanted objects (ground, bushes, tree) to extract buildings. For segmentation different tools have been used, e.g., edge detector [42], neural oscillator network [43] and hierarchical clustering [44]. For extraction of 3D building roofs, they then apply different region-growing algorithms [44], [36].

Based on the usage of the input data, there are three main categories of building extraction methods. The first category of methods [39], [41] fully relies on high resolution aerial imagery. Although they have shown promising results on some data sets, they generally do not perform well in densely built-up areas, partially due to shadows, occlusions and poor contrast. The second category of methods employs LIDAR data, and offers an improved level of automation when compared to image-based methods alone [44], [24]. Methods in the third category integrate aerial imagery and LIDAR data in order to exploit the complementary information from both data sources [36].

Fig. 3 shows an overview of a recently developed LIDAR-based building extraction procedure [24], that has been used to test the proposed evaluation system. The input data consists

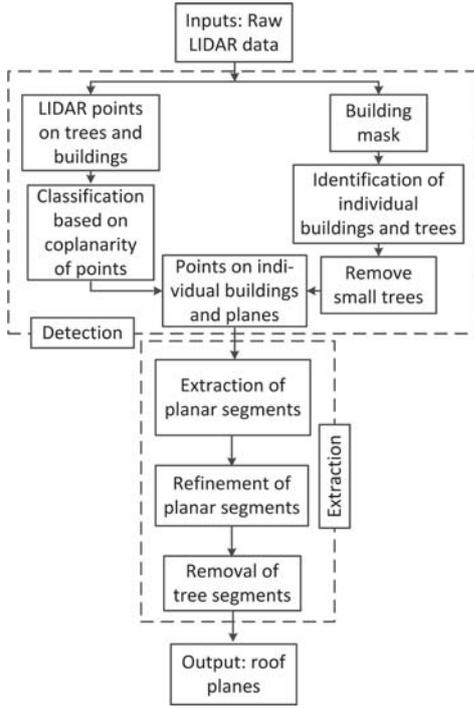


Fig. 3. The proposed roof plane extraction technique.

of a raw LIDAR point cloud. In the detection step (top dashed rectangle in Fig. 3), the LIDAR points are classified into two groups: *ground points*, such as ground, road furniture, cars and bushes that are below the threshold, and *non-ground points*, which represent elevated objects such as buildings and trees. The building mask, known as the ‘ground mask’, is generated using the ground points [1]. Individual buildings and trees are obtained as clusters of black pixels in the building mask and trees with low density canopies are removed. The coplanarity of each individual non-ground LIDAR point is ascertained based on its associated Delaunay neighbourhood (consisting of non-ground points). The planar segments are extracted from the non-ground LIDAR points on individual buildings and trees. The extracted LIDAR segments are then refined using a newly proposed rule-based procedure. The false planes on trees are removed using information such as area and neighbourhood, as well as any point spikes within the planar boundary. Finally, each building boundary is extracted as the boundary of a group of neighbouring roof planes.

Fig. 4 shows the extracted building boundaries and roof plane boundaries for Area 2 and Area 3, respectively, of the Vaihingen data set [45], that has been adopted as an ISPRS benchmark test data set [26] (see Section V).

IV. PROPOSED EVALUATION SYSTEM

The proposed evaluation system assumes that the roofs are represented by 3D polyhedral models. An individual roof consists of a set of planes. Therefore, a roof boundary or a plane can be represented as a 3D polygon having a set of corner points. If a roof boundary or plane has a curved side, that side can be reasonably approximated by a set of points.



Fig. 4. Building extraction by Awrangjeb and Fraser [24] on the Vaihingen data set: (a) building boundaries on Area 2 and (b) roof plane boundaries on Area 3.

The proposed evaluation system is shown as a block diagram in Fig. 5. Table I shows the symbols used in this section to explain the proposed evaluation system. Let R_d and R_r be the sets of extracted and reference roof boundaries, where each individual roof is represented by a boundary polygon. Furthermore, let P_d and P_r be the sets of extracted and reference roof planes, where each individual plane is also represented by a polygon. For each reference plane $p_r \in P_r$, there is a reference roof ID i_r that refers to the corresponding roof in R_r . So, all reference planes having the same i_r consist of a cluster of planes belonging to the same roof. Similarly, for each extracted plane $p_d \in P_d$, there is an extracted roof ID i_d that refers to the corresponding roof in R_d .

The 3D reference data sets R_r and P_r were created for the Vaihingen data set via image measurement using the Barista software [46]. The available reference label images¹ were used to find the reference data sets. Fig. 6 shows reference entities on Area 3 of the Vaihingen data set.

The proposed evaluation system in Fig. 5 works in two main steps: First, it is essential to establish pseudo one-to-one *roof correspondences* between the roof sets R_d and R_r (roof-level evaluation). *Pseudo one-to-one* correspondence means that each entity in one set has at most one correspondence in the other set (will be called ‘one-to-one correspondence’ or simply ‘correspondence’ from here). Once a roof correspondence (r_d, r_r) is found, where $r_d \in R_d$ and $r_r \in R_r$, then one-to-one *plane correspondences* between the planes of the corresponding extracted and reference clusters ($P_1 \subseteq P_d$ and $P_2 \subseteq P_r$, respectively) are established (plane-level evaluation). At each level, all necessary information (e.g., TP, FP and FN roofs, planes and pixels) for object- and pixel-based evaluations as well as for geometric evaluation is gathered for estimation of different evaluation indices in the final stage.

The dashed arrow between the roof- and plane-level evaluations in Fig. 5 indicates that the plane-level evaluation can be performed independently (i.e., without the roof-level evaluation) when R_r and R_d are not available. In this case,

¹Received along with the evaluation results when the roof extraction outputs from [24] had been submitted to ISPRS WG III/4: approach MON at <http://www2.isprs.org/commissions/comm3/wg4/tests.html>.

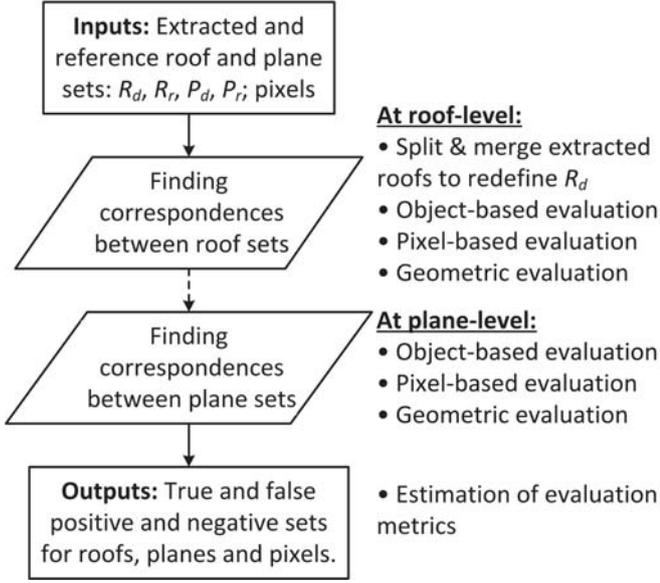


Fig. 5. Proposed evaluation system.

plane correspondences are established comparing each extracted plane to all reference planes. However, when R_r and R_d are available, establishing roof correspondences (r_d, r_r) before the plane-level evaluation limits the search space for each extracted plane, since only planes in P_1 are compared with planes in P_2 . Thus, the roof-level evaluation before the plane-level evaluation saves a significant computational time, as the number of planes is usually much larger than that of roofs.

During each evaluation step, each of the extracted and reference entities is marked as either TP, FP or FN. In addition, since reference buildings in R_r are expected to be physically separated from each other, the term *multiple detection* (MD) [35], which indicates that for an entity presented in the reference set there are two or more entities in the extracted set, has also been used. As shown in Fig. 7a, there may be two or more extracted entities (dashed rectangles) for the same reference entity (solid rectangle). In order to establish one-to-one correspondences it is important that only one of these extractions is considered as a TP and the rests are counted as MDs and removed from R_d . An index named *multiple detection rate* M_m indicates this error at roof-level evaluation. Since the planes on a building roof are physically connected to each other, M_m is not estimated at plane-level evaluation.

In addition, *detection cross-lap* and *reference cross-lap* [1] are used to estimate the under- and over-segmentation cases. Detection cross-lap is the number of extracted boundaries that overlap more than one reference boundary (see Fig. 7b) and reference cross-lap indicates the number of reference boundaries that are overlapped by more than one extracted boundary (see Fig. 7c)

The proposed evaluation system finds roof and plane correspondences based on maximum overlap areas between extracted and reference entities. A grid of pixels is generated at 0.25 m resolution. This provides approximately 16 pixels/m²

TABLE I
SYMBOLS USED IN SECTION IV.

Symbols	Explanations
R_r	Reference roof boundary set
R_d	Extracted roof boundary set
$r_r \in R_r$	A reference roof boundary in R_r
$r_d \in R_d$	An extracted roof boundary in R_d
P_r	Reference roof plane set
P_d	Extracted roof plane set
$p_r, p_k \in P_r$	A reference plane in P_r
$p_d, p_l \in P_d$	An extracted plane in P_d
i_r	Reference roof boundary ID
i_d	Extracted roof boundary ID
(r_d, r_r)	A true roof correspondence
$P_1 \subseteq P_d$	Extracted plane set within r_d
$P_2 \subseteq P_r$	Reference plane set within r_r
$S_r \subseteq P_d$	Extracted plane set within r_r
n_d	Number of planes in P_1
n_r	Number of planes in P_2
M_d	Metrics showing overlap amounts for P_1 (with P_2)
M_r	Metrics showing overlap amounts for P_2 (with P_1)
L_j	List of ext. planes that overlap p_j (reference)
L_i	List of ref. planes that are overlapped by p_i (extracted)
p_{i1}, p_{i2}, \dots	List of extracted planes
p_{j1}, p_{j2}, \dots	List of reference planes
(p_i, p_j)	A true plane correspondence
N_1	Number of extracted roof boundaries for r_r
N_2	Number of reference roof boundaries overlapped by r_d
$A_d(a_i)$	Set of pixels within r_d (within p_i)
$A_r(a_j)$	Set of pixels within r_r (within p_j)
TP (TPp)	True positive roof boundary (TPp: pixels in a TP)
FP (FPp)	False positive roof boundary (FPp: pixels in an FP)
FN (FNp)	False negative roof boundary (FNp: pixels in an FN)
TPP (TPPp)	True positive plane boundary (TPPp: pixels in a TPP)
FPP (FPPp)	False positive plane boundary (FPPp: pixels in an FPP)
FNP (FNPp)	False negative plane boundary (FNPp: pixels in an FNP)
d_{max}	Twice the max. point spacing in LIDAR data (optional)
MD	Number of roof boundaries that are multiply extracted
A, B, C, \dots	Extracted planes

and, therefore, is sufficient for roof planes as small as 1 m². All the pixels within each of the reference and extracted roofs and planes are obtained. For each pixel, the extracted and reference roof and plane IDs within which it resides are saved. Consequently, the overlap area between any two entities can be easily estimated.

A. Evaluation at Roof-Level

If a building is extracted multiple times, there will be more than one roof boundary for this building in R_d . Only one of these is kept and the others are removed. If two or more extracted boundaries $r_d \in R_d$ overlap each other and they only overlap the same reference boundary $r_r \in R_r$, then the r_d that has the largest overlap with r_r is kept and the others are removed from R_d . MD indicates the number of such extracted boundaries that are removed.

A topological clarification is now carried out on the remaining extracted roof boundaries in establishing one-to-one correspondences between R_d and R_r . A correspondence between r_d and r_r can be established using either point-in-polygon test or plane-in-polygon test. In point-in-polygon test, points or pixels inside the roof boundaries are used to find the overlap amount [5], [47]. Then the split-and-merge technique is applied for topological clarification [5]. This paper uses

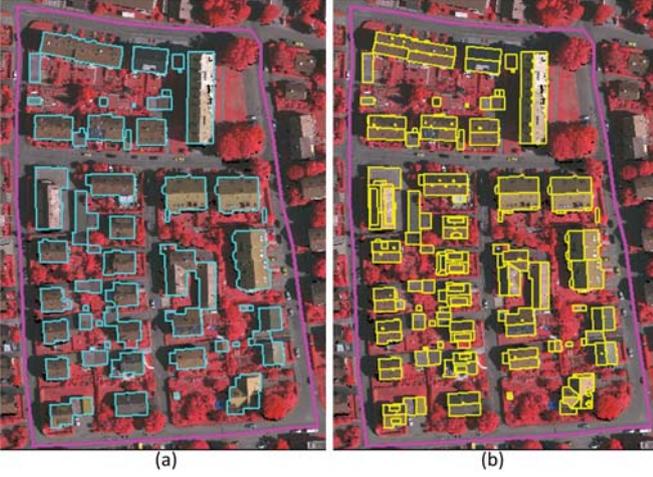


Fig. 6. Reference data sets on Area 3 of the Vaihingen data set: (a) roof boundaries and (b) plane boundaries.

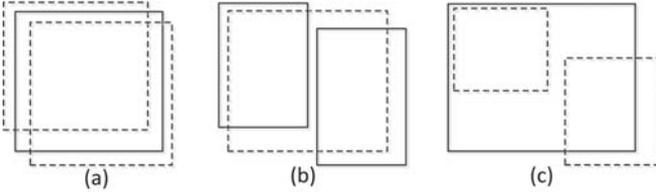


Fig. 7. Some erroneous situations in establishing pseudo one-to-one correspondences. Solid rectangles denote reference entities and dashed rectangles denote extracted entities: (a) multiple detection, (b) detection cross-lap (under-segmentation) and (c) reference cross-lap (over-segmentation).

plane-in-polygon test discussed below in order to find roof correspondences and applies a new split-and-merge technique for topological clarification. In the absence of extracted planes, the proposed evaluation system can use the point-in-polygon test presented in Awrangjeb and Fraser [47].

For the extracted roof boundaries, a split-and-merge technique can be followed to redefine their boundaries and to assign the extracted planes to the appropriate roof boundaries. The number of detections and reference cross-laps are also counted irrespective of the number of split and merge operations.

Ideally, an extracted plane boundary $p_d \in P_d$ overlaps only one reference roof boundary $r_r \in R_r$ and so p_d is simply decided to be in r_r . However, when two buildings are close to each other or when they are ‘connected’ by vegetation, p_d may overlap more than one reference boundary. In such a case, p_d is in the reference boundary that has the largest overlap with p_d . Let $S_r = \{p_d\}$ be the set of all the extracted planes that are in r_r . If all the planes in S_r have the same roof ID i_d referring to the same r_d , then a TP roof correspondence (r_r, r_d) is established. Figs. 8a and 8b show such an example where two extracted planes (yellow polygon) that form the same extracted roof (magenta polygon) are in the same reference roof (cyan polygon).

However, split and merge operations are required for the extracted roof boundaries when extracted planes from two or more different r_d are inside the same r_r (i.e., a reference

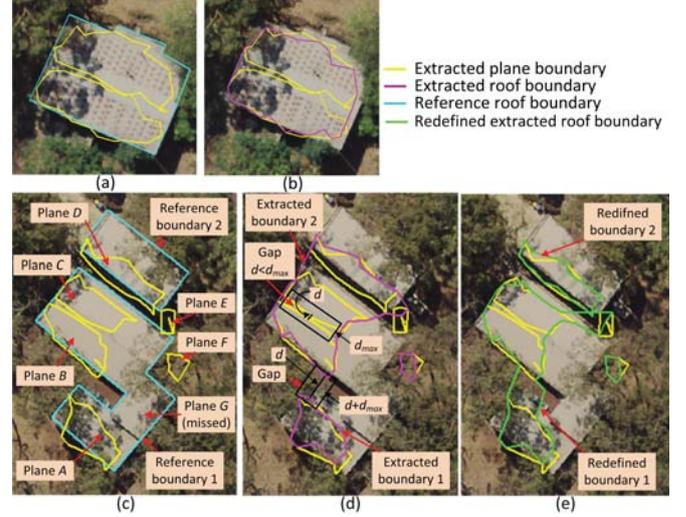


Fig. 8. Finding correspondences: (a-b) a simple case and (c-e) a complex case that requires split and merge operations.

cross-lap that requires one or more merge operations) and/or when extracted planes from the same r_d reside in two or more different r_r (i.e., a detection cross-lap that requires one or more split operations). As shown in Fig. 8c, Extracted Planes A, B and C reside within Reference Roof Boundary 1 (RRB1) and Plane D resides within Reference Roof Boundary 2 (RRB2). However, as shown in Fig. 8d, Awrangjeb and Fraser [24] finds only Plane A residing within Extracted Roof Boundary 1 (ERB1) since Plane A is far away (more than twice the maximum LIDAR point spacing, d_{max} [24]) from other planes, and Planes B and C falling into Extracted Roof Boundary 2 (ERB2) since Plane C is close (within d_{max}) to Plane D. Note, Plane E is a false plane which is found within ERB2 and RRB1 and Plane F forms a false extracted roof boundary.

The proposed evaluation system first excludes Planes B, C and E from ERB2 and then includes them into ERB1. These inclusion and exclusion operations simply change the respective plane IDs i_d to which ERB they belong. An actual split-and-merge operation happens when an extracted roof boundary whose entities have been reassigned by the above inclusion and/or exclusion operations is redefined. For example, the evaluation system redefines separate boundaries for ERB1 and ERB2, shown in Fig. 8e (green polygons), as follows. For the redefined ERB1, a binary mask is first formed using the pixels (or LIDAR points if available from the involved roof plane extraction technique) within Planes A, B, C and E. As shown in Fig. 8d, let d be the distance between two nearest points on two neighboring boundaries. If $d \leq d_{max}$ a rectangular gap of width d_{max} is filled. In Fig. 8d such a gap between Planes B and C is shown within a black rectangle. If $d > d_{max}$, a rectangular gap of width $d + d_{max}$ is filled. Such a gap is shown within another black rectangle in Fig. 8d between Planes A and B. Once a single shape is found within the mask for ERB1, the Canny edge around the shape is the boundary shown in Fig. 8e. The height at each edge point is the height of the nearest point on a plane boundary.

For the redefined ERB2, a new binary mask is formed using the pixels in Plane D and the Canny edge around the single shape is the redefined boundary.

In general, the merge operation is a rare case and only happens when some planes in between two or more planes on the same roof are missing or partially extracted leaving a large gap among the extracted planes. For example, as shown Fig. 8a, Planes B and C are partially extracted and Plane G is missed. Therefore, Plane A is found far away from Plane B that results in two extracted roof boundaries (ERB1 and ERB2). In contrast, the split operation happens when two or more neighbouring buildings are extracted as a single building. In a densely built-up area where buildings are close to each other or ‘connected’ by vegetation, the number of split operations will increase.

The required number of merge operations may be more than the number of reference cross-laps, because during a reference cross-lap a reference building may be extracted in as many as N_1 components (over-segmentation) and $N_1 - 1$ merge operations are thus required to form a single extracted boundary. Similarly, the required number of split operations may be more than the number of detection cross-laps, because during a detection cross-lap an extracted building may overlap N_2 reference boundaries (under-segmentation) and $N_2 - 1$ split operations are required to split them into N_2 extracted boundaries.

The above procedure continues until all the one-to-one correspondences are established between R_r and R_d . For object-based evaluation, both r_r and r_d are marked as TP for a true correspondence (r_r, r_d) . A reference boundary that does not have a correspondence with an extracted boundary is marked as an FN and an extracted boundary that does not have a correspondence with a reference boundary is marked as an FP. The numbers of split and merge operations, and detection and reference cross-laps are counted. For geometric evaluation, the root-mean-square error (RMSE) in position is estimated between the extracted and reference boundaries for each true correspondence (r_r, r_d) . For pixel- or area-based evaluation, for a roof correspondence (r_d, r_r) , let the set of pixels which reside inside r_d be A_d and the set of pixels which reside inside r_r be A_r . Pixels that reside within the overlapping area of r_d and r_r are true positive pixels (TPp = $A_d \cap A_r$), those which reside in r_d but not in r_r are false positive pixels (FPp = $A_d \setminus A_r$) and those which reside in r_r but not in r_d are false negative pixels (FNp = $A_r \setminus A_d$). Once pixels in all the TP roofs are processed, pixels in the roofs that were marked as FN and FP, are now directly added to the FNp and FPp sets, respectively.

Note that finding roof correspondences, as discussed above, is equivalent to finding building correspondences via methods reported in the literature [35]. This means that any building boundary extraction techniques that offer roof boundaries as polygons can be evaluated using the proposed roof-level evaluation. The split-and-merge technique presented in this paper uses a parameter d_{max} only in merge operations. This parameter simply indicates how many pixels can be added to bridge the two separately extracted planes on the same roof. This is not an overlap threshold and does not cause any

change to TP, FP and FN entities. Thus, it does not have any effect to object-based performance. This parameter can be considered optional by simply connecting the extracted planes via a line of pixels. It has been observed that the use of d_{max} or the line pixels have very negligible effect on the pixel-based performance, since the number of merge operations is limited in general. In the absence of extracted plane boundaries, the split-and-merge technique follows the procedure in Rutzinger et al. [5] using the generated pixels, but the correspondences can be established without using any overlap thresholds. If a reference entity r_r is overlapped by more than one extracted entity, then the extracted entity that has the largest overlap with r_r is chosen for r_r , and vice versa.

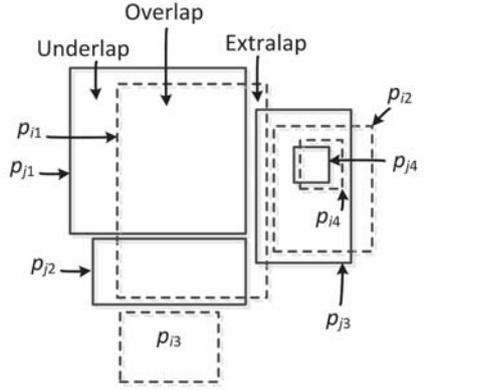
B. Evaluation at Plane-Level

Unlike evaluation at roof-level, no split and merge operations are applied at plane-level evaluation for topological clarification. Thus, for object-based evaluation at plane-level, all the planes in the corresponding plane clusters P_1 and P_2 , related to a roof correspondence (r_d, r_r) , are examined directly without redefining any plane boundaries. The correspondences between P_1 and P_2 are again made using maximum overlaps.

Let the number of planes in P_1 and P_2 be n_d and n_r . Using the pixel sets $A_d \in r_d$ and $A_r \in r_r$, two metrics M_d (size $n_d \times n_r$) and M_r (size $n_r \times n_d$) are estimated. Further, let the pixels that reside inside an extracted plane $p_i \in P_1$ be $a_i \in A_d$, where $1 \leq i \leq n_d$, and those that reside inside a reference plane $p_j \in P_2$ be $a_j \in A_r$, where $1 \leq j \leq n_r$. The array element $M_d(i, j)$ represents the percentage of pixels (from p_i) that also reside within p_j . Similarly, $M_r(j, i)$ represents the percentage of pixels (from p_j) that also reside in p_i . Through the use of M_d , a list of extracted planes $L_j = \{p_{i1}, p_{i2}, \dots\}$ that maximally overlap p_j can be obtained (from column j of M_d). Similarly, use of M_r allows a list of reference planes $L_i = \{p_{j1}, p_{j2}, \dots\}$ that are maximally overlapped by p_i to be obtained (from column i of M_r). Entities in L_i (and L_j) are sorted in descending order of the overlapping percentages with p_i (and p_j respectively).

For example, Fig. 9 shows extracted and reference plane lists for four reference planes and four extracted planes, respectively. From the list L_{i1} it is seen that the extracted plane p_{i1} shares 60% of its pixels with the reference plane p_{j1} , 20% with p_{j2} , 5% with p_{j3} , and the remaining 15% pixels do not make any overlaps with any reference planes. Similarly, from the list L_{j3} it is seen that p_{j3} shares 65% of its pixels with p_{i2} , 10% with p_{i1} , and the remaining 25% pixels do not make any overlaps with any extracted planes. The reference plane list L_{i3} for p_{i3} is empty since it does not overlap any reference planes. An exception happens when a plane resides inside another plane. For example, p_{j4} is inside p_{j2} or p_{i4} is inside p_{i2} . In this case, 12% of the pixels that p_{i2} shares with p_{j4} also shares with p_{j3} . Moreover, p_{j4} shares its 100% pixels with p_{i2} but also share its 70% pixels with p_{i4} .

In order to find a *plane correspondence* (p_i, p_j) , a check on whether p_i is the extracted plane that has caused the most overlap with p_j and vice versa is carried out. This means the extracted plane having the largest overlap among the planes



True correspondences: $(p_{i1}, p_{j1}), (p_{i2}, p_{j3}), (p_{i4}, p_{j4})$

False positive: p_{i3}

False negative: p_{j2}

Reference plane lists (in descending order of overlaps):

with extracted plane p_{i1} , $L_{i1}=\{p_{j1}=0.6, p_{j2}=0.2, p_{j3}=0.05\}$,

with extracted plane p_{i2} , $L_{i2}=\{p_{j3}=0.85, p_{j4}=0.12\}$

with extracted plane p_{i3} , $L_{i3}=\{\}$

and extracted plane p_{i4} , $L_{i4}=\{p_{j3}=1.0, p_{j4}=0.55\}$

Extracted plane lists (in descending order of overlaps):

with reference plane p_{j1} , $L_{j1}=\{p_{i1}=0.75\}$,

with reference plane p_{j2} , $L_{j2}=\{p_{i1}=0.82\}$

with reference plane p_{j3} , $L_{j3}=\{p_{i2}=0.65, p_{i1}=0.10\}$

and with reference plane p_{j4} , $L_{j4}=\{p_{i2}=1.0, p_{i4}=0.70\}$

Fig. 9. Overlapping plane lists. Solid rectangles denote reference planes and dashed rectangles denote extracted planes: i stands for index to extracted plane boundaries and j stands for index to reference plane boundaries.

in L_j is $p_i = p_{i1}$ and the reference plane having the largest overlap among the planes in L_i is $p_j = p_{j1}$ are checked. If the above condition is met, then (p_i, p_j) is an eligible plane correspondence and both p_i and p_j are marked as true positive planes (TPP). For example, from Fig. 9 L_{i1} indicates that p_{j1} has the largest overlap with p_{i1} and L_{j1} indicates that p_{i1} has the largest overlap with p_{j1} . Thus, (p_{i1}, p_{j1}) is a true plane correspondence. Similarly, (p_{i2}, p_{j3}) is another true plane correspondence.

After establishing all the eligible plane correspondences based on the above condition, a further condition based on the first or second most overlapping entities is applied to establish other plane correspondences. This check helps in finding correspondences for planes that reside inside other planes. For an extracted plane p_i that does not yet have a corresponding reference plane, both first (say, p_{j1}) and second (say, p_{j2}) reference entities, if any, in L_i are checked. If p_{j1} does not yet have a corresponding extracted plane, then the two largest overlapping entities $p_{i1}, p_{i2} \in L_{j1}$ are checked. If $p_{i1} = p_i$ or $p_{i2} = p_i$, then (p_i, p_{j1}) is a new plane correspondence and both p_i and p_{j1} are marked as TPP. If p_{j1} already has a correspondence, but p_{j2} does not, then the two largest overlapping entities $p_{i1}, p_{i2} \in L_{j2}$ are checked. If $p_{i1} = p_i$ or $p_{i2} = p_i$, then (p_i, p_{j2}) is a new plane correspondence and both p_i and p_{j2} are marked as TPP.

Fig. 9 shows an example. From L_{i4} it is evident that p_{j4} has the largest overlap with p_{j3} , which already has a

correspondence with p_{i2} . Therefore, the list L_{j4} of p_{j4} is checked since p_{i4} has the second largest overlap with p_{j4} . From L_{j4} it is evident that p_{j4} has the largest overlap with p_{i2} which already has a correspondence with p_{j3} . Moreover, p_{j4} has the second largest overlap with p_{i4} . Thus, a true correspondence (p_{i4}, p_{j4}) is found.

After checking all the extracted planes that do not yet have correspondences, a similar check can be executed for all reference planes that as yet do not have correspondences. Thereafter, if an extracted or a reference plane still does not have a correspondence, it is marked as a false positive plane (FPP) or a false negative plane (FNP). For example, as shown in Fig. 9, p_{i3} does not overlap any reference planes. Therefore, p_{i3} may be a tree and marked as an FPP. In addition, p_{j2} has the only overlap with p_{i1} which already has a true correspondence with p_{j1} . Therefore, p_{j2} is marked as an FNP.

A detection cross-lap may only happen with FNP and a reference cross-lap may only happen with FPP. Since two neighbouring reference planes on a building roof are connected to each other and a small reference plane may completely reside within a large reference plane, a true extracted plane may overlap two or more reference planes and a reference plane may overlap two or more true extracted planes. For example, as shown in Fig. 9, p_{i1} overlaps p_{j1} , p_{j2} and p_{j3} . The overlap between p_{i1} and p_{j1} establishes a true correspondence (p_{i1}, p_{j1}) . The overlap between p_{i1} and p_{j3} is to be anticipated, since p_{j1} and p_{j3} are neighbours and p_{j3} has a true correspondence with p_{i2} . However, the overlap between p_{i1} and p_{j2} is considered quite unexpected, since a correspondence could not be established for p_{j2} .

Thus, a detection cross-lap occurs for an extracted plane p_i if a reference plane $p_k \in L_i$, where $k = 1, 2, \dots, n_r$, still remains without a correspondence. This means that although p_k has an overlap with p_i a correspondence could not be established in the above procedure. Similarly, a reference cross-lap occurs for a reference plane p_j if a detected plane $p_l \in L_j$, where $l = 1, 2, \dots, n_d$, still remains without a correspondence, which implies although p_l has an overlap with p_j , but a correspondence could not be established. For example, a detection cross-lap occurs between p_{i1} and p_{j2} in Fig. 9.

Similar to the roof-level evaluation, pixels in TPP, FPP and FNP planes are used to form TPPp, FPPp and FNPp sets for pixel- or area-based evaluation and the RMSE is estimated for TPP for geometric evaluation. If the segmented LIDAR points are available for the extracted planes, then different height errors can be estimated by comparing the estimated heights from the plane equations to the actual heights. Appropriate metrics include the *mean absolute error* (MAE) and RMSE.

An important observation with the proposed evaluation system is that it does not apply the split-and-merge technique at plane-level evaluation. Since planes on a roof are connected to one another, if a small plane is not extracted separately but jointly extracted with a neighbouring large plane, the pixels that are outside the reference of the large plane are marked as FNPp, although they may be within the small reference plane. For example, in Fig. 9 the extracted plane p_{i1} covers the area of three reference planes p_{j1} , p_{j2} and p_{j3} . According to Shan and Lee [27], the common area for the correspondence

(p_{i1}, p_{j1}) is marked as overlap, the area that is outside p_{i1} but inside p_{j1} is marked as underlap and the area that is inside p_{i1} but outside p_{j1} is marked as extralapl. Consequently, although the pixels in an underlap area can be covered by another TP extracted plane, they are marked as FNPP. Similarly, although the pixels in an extralapl area can be within the area of a neighbouring reference plane, they are marked as FPPP. This clearly reflects the actual performance of a 3D plane extraction technique being evaluated. Because pixels within underlap and extralapl areas of an extracted planes may not physically reside within its neighbouring 3D reference planes. Thus, the pixel-based accuracy at plane-level may be much lower than that at roof-level.

C. Evaluation metrics

Within object-based evaluation, there is a determination of whether for a given entity in the reference set there is a corresponding entity present in the detection set. While at the roof-level the number of roofs is considered, at the plane-level the number of planes is considered. In pixel- or area-based evaluation, pixels within the TP, FP and FN roof boundaries are considered at roof-level evaluation and those within plane boundaries are considered as plane-level evaluation. Table II shows the evaluation metrics used in this study. In object-based evaluation, completeness C_m is also known as detection rate [8] or producer's accuracy [10], and correctness C_r is also known as user's accuracy [10]. In pixel-based evaluation, completeness C_{mp} is also known as matched overlay [8] and detection rate [48]. Ideally, completeness, correctness and quality values should be maximum at 100% and all other metric values in Table II should be zero.

All the geometric accuracy indices are estimated using the true positive entities only. For each one-to-one correspondence between detected and reference sets, the $RMSE_{xy,r}$ (with respect to reference) and $RMSE_{xy,d}$ (with respect to detection) are measured as the average distance between a pair of detected and reference entities. For $RMSE_{xy,r}$, for each corner of a reference entity the nearest LIDAR point is chosen from the boundary of the corresponding detected entity. For $RMSE_{xy,d}$, for each point on a detected entity the minimum perpendicular distance to the corresponding reference entity sides is chosen. The use of two $RMSE_{xy}$ indices are necessary when the detected and reference entities do not have the same number of corner points. In addition, the plane fitting error, also known as shaping error [18], is estimated at the plane-level using the difference in point height values between the estimated (from its corresponding plane equation) and LIDAR heights. The average height difference for all the LIDAR points within a given plane is considered as the fitting error of that plane. In the evaluation, for a test area the largest negative error (Δ_- , for points below the planes), the largest positive error (Δ_+ , for points above the planes), mean of largest negative errors over all planes ($\overline{\Delta_m}$), mean of largest positive errors over all planes ($\overline{\Delta_M}$), mean absolute error (MAE_z) and RMSE ($RMSE_z$) have been used to express the plane fitting error [50].

Moreover, plane distance (d_p), normal displacement (d_n) and directional accuracy (Δ_θ) between the corresponding

extracted and reference planes (p_i, p_j) are estimated using two plane (unit) normal vectors. The directional accuracy Δ_θ is simply measured as the angle between the normal vectors. Since, p_i and p_j may not be exactly parallel $\|\Delta_\theta\| \geq 0$ and thus the plane distance d_p is measured as the average perpendicular distance between p_i and p_j . For finding distance from p_i to p_j , and vice versa, the pairs of boundary points for (p_i, p_j) , obtained above during estimation of $RMSE_{xy}$, and their mean points are used. Since the reference entities do not come with height information, for each reference corner point the height of the nearest LIDAR point is used as reference height as in [23]. The perpendicular distances from points in p_i to p_j are estimated using the normal vector of p_j , and vice versa. The normal displacement d_n is simply the magnitude of the difference vector of the two unit normal vectors.

The majority of the evaluation metrics (e.g., completeness, correctness and quality) presented above are popular and they express the performance of a building extraction technique in different applications. The other metrics (e.g, cross-laps, multiple detection and numbers of split and merge operations) are chosen to show the segmentation errors. The geometric metrics express the 2D and 3D accuracy of the extracted building roofs and planes. Note that Zeng et al. [12] proposed a single overall metric through combining individual metrics. A single metric can make the comparison of different building extraction techniques straightforward. However, the derivation of such a single metric is subjective because it requires setting of weights to different metrics. Nevertheless, these weights depend on applications. In addition, a single evaluation metric fails to detail the performance of a building extraction technique.

V. EVALUATION RESULTS

The test data set employed is the Vaihingen (VH) data set [45] that has been adopted as an ISPRS benchmark test data set [26]. There are three test sites in this data set. Area 1 has a point density of 3.5 points/m² and is characterised by dense development consisting of historic buildings having complex shapes. The point density in Area 2 is 3.9 points/m² and this area is characterised by a few high rise residential buildings surrounded by trees. Area 3 is purely residential with detached houses and many surrounding trees and has a point density of 3.7 points/m². The number of buildings (larger than 2.5 m²) in each of these three areas is 37, 14 and 56, and the corresponding numbers of planes are 288, 69 and 235, respectively. Figs. 4 and 6 show some results and reference data for Areas 2 and 3 of the test data set.² Two performance studies for the roof-level and plane-level evaluations are separately presented.

A. Roof-level performance

Tables III and IV show the evaluation results at roof-level. In object-based evaluation (Table III), among the three scenes, the building extraction algorithm [24] performed better in

²All the test data sets along with the reference and extracted roofs and planes are available at <http://users.monash.edu.au/~mawrangj/RExtraction.html>.

TABLE II
EVALUATION METRICS USED AT ROOF- AND PLANE-LEVEL EVALUATIONS ('R' INDICATES A PARTICULAR METRIC HAS BEEN APPLIED AT ROOF-LEVEL AND 'P' INDICATES IT HAS BEEN APPLIED AT PLANE-LEVEL).

Object-based	Applied	Pixel-based	Applied	Geometric	Applied
Completeness C_m [5]	R, P	Completeness C_{mp} [5]	R, P	$RMSE_{xy}$ [49]	R,P
Correctness C_r [5]	R, P	Correctness C_{rp} [5]	R, P	Max negative error Δ_- [51]	P
Quality Q_l [5]	R, P	Quality Q_{lp} [5]	R, P	Max positive error Δ_+ [51]	P
Multiple detection M_m [1]	R	Area omission error A_{oe} [8]	R, P	Mean of Δ_- errors $\overline{\Delta_-}$ [51]	P
Detection cross-lap C_d [1]	R, P	Area commission error A_{ce} [8]	R, P	Mean of Δ_+ errors $\overline{\Delta_+}$ [51]	P
Reference cross-lap C_r [1]	R, P	Branching factor B_f [48]	R, P	Mean absolute error MAE_z	P
Split operations N_s	R	Miss factor M_f [48]	R, P	$RMSE_z$ [15]	P
Merge operations N_m	R			Plane distance d_p	P
				Normal displacement d_n	P
				Directional accuracy Δ_θ	P

TABLE III
OBJECT-BASED EVALUATION RESULTS AT ROOF-LEVEL (C_m = COMPLETENESS, C_r = CORRECTNESS, Q_l = QUALITY ($C_{m,50}$, $C_{r,50}$ AND $Q_{l,50}$ ARE FOR BUILDINGS OVER 50 M²), M_m = MULTIPLE DETECTION RATE, C_d = DETECTION CROSS-LAP RATE AND C_r = REFERENCE CROSS-LAP RATE ARE IN PERCENTAGE; N_s = NUMBER OF SPLIT AND N_m = NUMBER OF MERGE OPERATIONS).

Areas	C_m	C_r	Q_l	$C_{m,50}$	$C_{r,50}$	$Q_{l,50}$	N_s	N_m	C_d	C_r	M_m
1	73.1	100	73.1	100	100	100	7	0	33.3	0	0
2	76.9	100	76.9	100	100	100	0	0	0	0	0
3	79.6	100	79.6	100	100	100	6	0	12.1	0	0
Avg	76.5	100	76.5	100	100	100	4.3	0	15.1	0	0

TABLE IV
PIXEL-BASED AND GEOMETRIC EVALUATION RESULTS AT ROOF-LEVEL (C_{mp} = COMPLETENESS, C_{rp} = CORRECTNESS, Q_{lp} = QUALITY, A_{oe} = AREA OMISSION ERROR, A_{ce} = AREA COMMISSION ERROR, B_f = BRANCHING FACTOR AND M_f = MISS FACTOR IN PERCENTAGE; $RMSE_{xy,r}$ AND $RMSE_{xy,d}$ IN METRE).

Areas	C_{mp}	C_{rp}	Q_{lp}	A_{oe}	A_{ce}	B_f	M_f	$RMSE_{xy,r}$	$RMSE_{xy,d}$
1	86.1	95.4	82.7	13.9	4.6	4.8	16.2	1.61	0.40
2	85	99.1	84.3	15	0.9	0.9	17.7	1.28	0.46
3	84.9	98.2	83.5	15.2	1.8	1.8	17.9	1.06	0.49
Average	85.3	97.6	83.5	14.7	2.4	2.5	17.3	1.32	0.45

Area 3 since this area mainly contains residential buildings, as shown in Fig. 6. In all three areas, some small garden shed, garages and carports were missed, thus the completeness and quality values were not at their maximum values. The correctness value is maximum in all three areas, which indicates that the algorithm did not extract any trees. However, the algorithm extracted all buildings larger than 50 m² in area. In Areas 1 and 3, the algorithm extracted some neighbouring buildings together and, therefore, a number of split operations have been required for topological clarification, and there is detection cross-lap rates of 33% and 12%, respectively. In Area 2, buildings are well separated from each other and thus each of the extracted buildings covered only one actual building. In all three areas, there were no multiple detections as the algorithm extracted only one roof boundary by accumulating all LIDAR points from each group of neighbouring planes. Moreover, there were no merge operations and so no reference cross-lap rate since no single buildings were extracted in two or more parts by the extraction algorithm.

In pixel-based evaluation (Table IV), high area omission error and miss factor values indicate that the algorithm missed some parts of each extracted building, specially along the building boundary. This is due to use of LIDAR data alone. As can be seen in Fig. 4a, the extracted roof boundaries did not exactly match the actual boundaries. It is also evident from

the low geometric accuracy values shown in Table IV. The $RMSE_{xy}$ values are about one to two times the maximum point-spacing in the raw LIDAR data. Since, the detected roof boundary is not regularised some reference corners may stay away from the nearest detected boundary points. But the points on the detected boundary stay close to the reference roof sides. Thus, the $RMSE_{xy,r}$ is higher than $RMSE_{xy,d}$. The low area commission error and branching factor in all three areas again prove that the algorithm extracted only a small number of non-building areas (trees etc).

B. Plane-level performance

Tables V to VII show the evaluation results at plane-level. In object-based evaluation (Table V), when all the planes are considered, the algorithm performed the best in Area 2. In Areas 1 and 3, there are many planes smaller than 10 m², which the algorithm [24] either missed or merged with neighbouring planes. Therefore, when only the planes that are larger than 10 m² in area are considered the algorithm performed the best in Area 3. The detection cross-lap rate is higher than the reference cross-lap rate. This indicates that the algorithm merged many small planes with the neighbouring large planes and the number of over-segmentation cases was lower than the number of under-segmentation cases.

TABLE V

OBJECT-BASED EVALUATION RESULTS AT PLANE-LEVEL (C_m = COMPLETENESS, C_r = CORRECTNESS, Q_l = QUALITY ($C_{m,10}$, $C_{r,10}$ AND $Q_{l,10}$ ARE FOR PLANES OVER 10 M^2), C_d = DETECTION CROSS-LAP RATE AND C_r = REFERENCE CROSS-LAP RATE ARE IN PERCENTAGE).

Areas	C_m	C_r	Q_l	$C_{m,10}$	$C_{r,10}$	$Q_{l,10}$	C_d	C_r
1	50.8	88.4	47.7	84.5	88.4	76.1	37.6	6
2	67.1	83.6	59.3	91.1	83.6	77.3	16.4	11.8
3	52.2	90.3	49.4	93.8	90.3	85.2	35.8	5.6
Avg	56.7	87.4	52.1	89.8	87.4	79.5	29.9	7.8

TABLE VI

PIXEL-BASED EVALUATION RESULTS AT PLANE-LEVEL (C_{mp} = COMPLETENESS, C_{rp} = CORRECTNESS, Q_{lp} = QUALITY, A_{oe} = AREA OMISSION ERROR, A_{ce} = AREA COMMISSION ERROR, B_f = BRANCHING FACTOR AND M_f = MISS FACTOR IN PERCENTAGE).

Areas	C_{mp}	C_{rp}	Q_{lp}	A_{oe}	A_{ce}	B_f	M_f
1	77.9	64.5	54.5	22.1	35.5	55.1	28.3
2	85.6	86.5	75.6	14.4	13.5	15.6	16.8
3	87.6	81.4	73	12.4	18.6	22.8	14.2
Average	83.7	77.5	67.7	16.3	22.5	31.2	19.8

In pixel-based evaluation (Table VI), as can be expected, the algorithm performed the best in Area 2, which has the least number of small planes among the three areas. While compared to Areas 2 and 3, the high values for omission and commission errors and branching and miss factors for Area 1 imply that the algorithm performed the worst in Area 1, which possesses a dense development consisting of historic buildings having complex shapes. When these errors are compared with those in Table IV, it is evident that the pixel-based performance at roof-level is better than that at plane-level. This is due to the policy that the proposed evaluation system does not involve the split-and-merge technique at plane-level evaluation. Consequently, although the pixels in the underlap and extralapl areas reside within the area of other plane boundaries, as shown in Fig. 9, they are marked as false (negative and positive, respectively) by the proposed system.

The planimetric accuracy of the extracted planes (see $RMSE_{xy}$ in Table VII, within 1.5 times the maximum point-spacing in the raw LIDAR data) is better than that of the extracted roof boundaries (see Table IV). This is because while the LIDAR points that reflect from the walls have been correctly excluded from the planes along the roof boundary, points in between roof planes have also been correctly assigned to the appropriate extracted planes. The maximum negative and positive errors are within 25 to 60 cm for all three scenes, whereas the two means of these values are within 9 to 16 cm which may be within the error bound in the input LIDAR height. This observation suggests that only a small number of points on the extracted planes showed large errors, while the majority of the points fit exactly with the extracted planes. Another observation, that the MAE and $RMSE_z$ values are only 2 to 4 cm for all three scenes, entails the fact that the roof extraction technique [24] correctly classifies most of the LIDAR points on building roofs to the appropriate extracted planes.

Table VII further shows that the directional accuracy is between 2 to 4 degrees, the plane distance is between 9 to

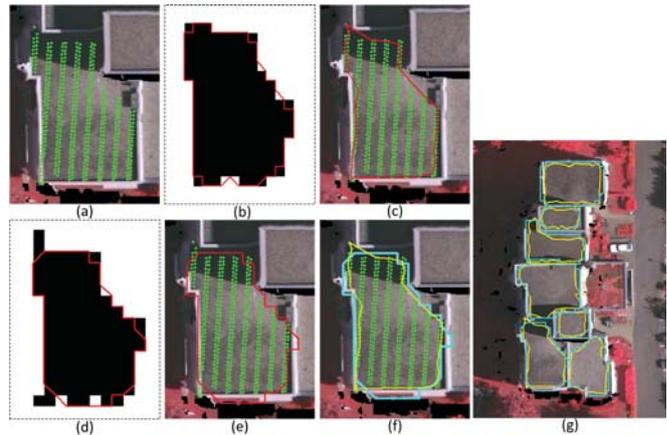


Fig. 10. Extracting a plane boundary in two different methods: (a) LIDAR points on the plane, (b) Canny edge around a mask of resolution 1 m, (c) boundary after assigning nearest LIDAR points to boundary in (b), (d) Canny edge around a mask of resolution 0.25 m, (e) boundary after assigning only height values from nearest LIDAR points to boundary in (d), (f) comparing two extracted boundaries, (g) two types of extracted boundaries on the whole building roof.

18 cm, and the normal displacement is 3 to 24 cm. Area 2 has shown the best directional, displacement and distance accuracies, as the majority of the extracted planes in this area are flat and large in size. Some small and sloppy planes, especially in Area 1, showed large directional and distance errors.

C. Robustness test

In order to test the robustness of the proposed evaluation method, boundaries around the plane segments are extracted in two different methods as shown in Fig. 10. LIDAR points on a plane segment are shown in Fig. 10a. Using the LIDAR points on the plane, the first method (adopted from Awrangjeb et al. [36]) initially forms a binary mask of resolution 1 m (same as the maximum LIDAR point-spacing) as illustrated in Fig. 10b. This method then extracts a Canny edge around the black shape in Fig. 10b and assigns the nearest LIDAR points to the edge points. Fig. 10c shows the extracted plane boundary.

The second method generates a slightly different mask of resolution 0.25 m, following the procedure in [42] (Fig. 10d). It then extracts a Canny edge around the black shape and assigns only the heights values from the nearest LIDAR points to the edge points. Fig. 10e shows the extracted plane boundary using the second method.

Figs. 10f and 10g compare the two sets of extracted plane boundaries on an individual plane and on the whole roof. In three areas of the test data set the area of plane boundaries varied between 7 and 9 m^2 . Hereafter, a boundary from the first set of boundaries will be named a ‘LIDAR boundary’ and one from the second will be named a ‘Canny boundary’. Note that the tracking of a Canny edge has been described in [52] and the results shown in Tables III to VII are based on the second method.

The proposed evaluation system, as well as the threshold-based evaluation system [5], previously adopted for the ISPRS

TABLE VII

GEOMETRIC EVALUATION RESULTS AT PLANE-LEVEL ($RMSE_{xy,r}$, $RMSE_{xy,d}$ = PLANIMETRIC ACCURACIES, Δ_- = LARGEST NEGATIVE ERROR, Δ_+ = LARGEST POSITIVE ERROR, $\overline{\Delta}_m$ = MEAN OF LARGEST NEGATIVE ERRORS, $\overline{\Delta}_M$ = MEAN OF LARGEST POSITIVE ERRORS, MAE_z = MEAN ABSOLUTE (HEIGHT) ERROR, $RMSE_z$ = HEIGHT ACCURACY, d_p = PLANE DISTANCE AND d_n = NORMAL DISPLACEMENT IN METRE; Δ_θ = DIRECTIONAL ACCURACY IN DEGREES).

Scenes	$RMSE_{xy,r}$	$RMSE_{xy,d}$	Δ_-	Δ_+	$\overline{\Delta}_m$	$\overline{\Delta}_M$	MAE_z	$RMSE_z$	d_p	d_n	Δ_θ
1	0.678	0.500	-0.585	0.329	-0.152	0.118	0.027	0.037	0.173	0.237	3.744
2	0.884	0.413	-0.449	0.251	-0.151	0.095	0.021	0.030	0.099	0.039	2.257
3	0.663	0.449	-0.414	0.466	-0.121	0.114	0.021	0.029	0.131	0.142	2.297
Average	0.742	1.362	-0.483	0.349	-0.141	0.109	0.023	0.032	0.134	0.139	2.766

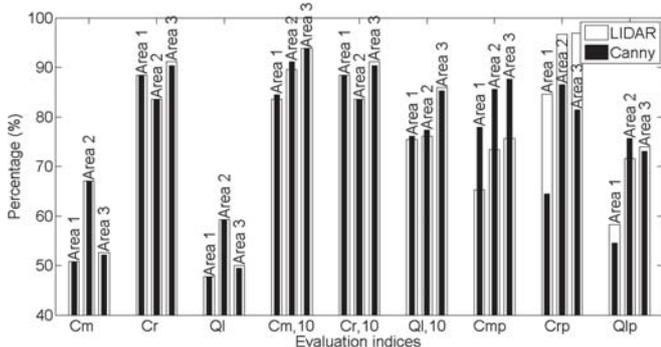


Fig. 11. Evaluation results by the proposed evaluation system on two sets of extracted plane boundaries: LIDAR and Canny (Object-based indices: C_m = completeness, C_r = correctness, Q_l = quality ($C_{m,10}$, $C_{r,10}$ and $Q_{l,10}$ are for planes over 10 m^2); pixel-based indices: C_{mp} = completeness, C_{rp} = correctness, Q_{lp} = quality in percentage).

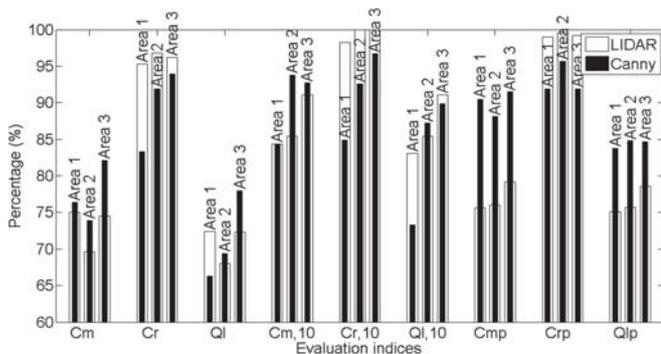


Fig. 12. Evaluation results by the threshold-based evaluation system [5] on two sets of extracted plane boundaries: LIDAR and Canny (Object-based indices: C_m = completeness, C_r = correctness, Q_l = quality ($C_{m,10}$, $C_{r,10}$ and $Q_{l,10}$ are for planes over 10 m^2); pixel-based indices: C_{mp} = completeness, C_{rp} = correctness, Q_{lp} = quality in percentage).

benchmark data sets [26], have been applied on both LIDAR and Canny boundaries from the Vaihingen data set. Note that the estimation of TP, FN and TP entities in both object- and pixel-based evaluations is different in [5] from that in the proposed evaluation system. Rottensteiner et al. [6] considers ‘overall coverage’, where the percentage of overlap for a given extracted entity is estimated by accumulating all its regions that are covered by one or more reference entities (see Fig. 1a) and the percentage of overlap for a given reference entity is estimated by accumulating all its regions that are covered by one or more extracted entities (see Fig. 1b). Since the roof planes have a 3D nature and may be connected to each

other, it is practical that the extracted planes overlap each other. Thus, the accumulation of multiple overlap regions to estimate the ‘overall coverage’ does not reflect the actual segmentation performance. For example, the accumulation of three overlap regions in Fig. 1a deems impractical unless all three reference planes are flat at the same height, which in fact makes them a single plane. Consequently, the evaluation based on the ‘overall coverage’ eventually increases the estimated completeness, correctness and quality values in both object- and pixel-based evaluations for the threshold-based evaluation system [5]. This paper does not accumulate overlap regions, but estimates the percentage of overlaps separately in order to establish correspondences (Fig. 9).

Figs. 11 and 12 show the evaluation results on two sets of boundaries from the proposed and threshold-based systems, respectively. It is evident that the object-based performance by the proposed system is almost the same for the two boundary types. Completeness, correctness and quality (both for all planes and for planes larger than 10 m^2) values are almost unchanged even when the areas of the extracted plane boundaries changed from 7 to 9 m^2 . This also shows that the use of parameter d_{max} has almost no effect although the area of each extracted plane has been changed. In contrast, the object-based performance of the threshold-based system [5] changes significantly with the change of area of extracted boundaries. This system marks a reference plane as TP if the plane has an accumulated ‘overall coverage’ of more than 50% with one or more extracted planes. Similarly, it marks an extracted plane as TP if the plane has an accumulated ‘overall coverage’ of more than 50% with one or more reference planes. Therefore, many of the true correspondences having 50 to 60% overlap using one type of boundary may become false correspondences when other type of boundary is used, and many of the false correspondences having 40 to 50% overlaps using one type of boundary may become true correspondences when other type of boundary is used. As a result, there may have been a dramatic change in the number of true and false correspondences that resulted in significant swings in its object-based performance. However, the pixel-based performance changes as expected for both evaluation systems, since the area has been changed between the two types of boundaries.

An important note from Figs. 11 and 12 is that since the proposed evaluation system does not accumulate overlap regions to estimate the ‘overall coverage’ at plane-level, its object-based performance is significantly lower when all planes are considered than when only the planes larger than 10 m^2 in area

are considered. A large number of small planes, which may have been missed or merged with the neighbouring planes, have caused this performance anomaly, specially in Areas 1 and 3. Nevertheless, since the threshold-based system [5] accumulates overlap regions to estimate the ‘overall coverage’ at plane-level (see Fig. 1), it shows higher performance than the proposed system, both when all planes are considered and when planes larger than 10 m² in area are considered. One such case has been exemplified in Fig. 2, where an originally missed reference plane in between two large planes has been considered as a TP by Rutzinger et al. [5].

VI. CONCLUSION

This paper has proposed a new overlap threshold-free evaluation system that can be used for automatic evaluation of building extraction techniques that extract building roof boundaries and/or individual roof planes. Since it does not require any human judgement, it can be used for unbiased performance evaluation on large data sets. The proposed evaluation system assesses the extracted objects in a more comprehensive manner than existing systems through use of a number of evaluation indices in three categories at both roof- and plane-levels. Moreover, since the roof- and plane-level evaluations can be carried out independently, the proposed system can be employed to evaluate other building detection and roof extraction techniques whose outputs consist of polygonal entities.

While the estimated performance of the proposed system has been compared with that of a threshold-based system [5], it has been shown that the proposed system offers more robust object-based results than Rutzinger et al. [5]. The increased performance by Rutzinger et al. [5] at plane-level evaluation is due to application of overlap region accumulation that artificially increases the percentage of overlap, and therefore does not reflect the actual performance of a building extraction technique being evaluated.

ACKNOWLEDGMENT

Dr. Awrangjeb is the recipient of the Discovery Early Career Researcher Award by the Australian Research Council (project number DE120101778). The Vaihingen data set was provided by the German Society for Photogrammetry, Remote Sensing and Geoinformation (DGPF) [45]: <http://www.ifp.uni-stuttgart.de/dgpf/DKEP-Allg.html>.

REFERENCES

- [1] M. Awrangjeb, M. Ravanbakhsh, and C. S. Fraser, “Automatic detection of residential buildings using LIDAR data and multispectral imagery,” *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 65, no. 5, pp. 457–467, 2010.
- [2] N. Haala and M. Kada, “An update on automatic 3D building reconstruction,” *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 65, no. 6, pp. 570–580, 2010.
- [3] R. Wang, “3D building modeling using images and LIDAR: a review,” *International Journal of Image and Data Fusion*, vol. 1, no. 4, pp. 273–292, 2013.
- [4] S. Oude Elberink and G. Vosselman, “Quality analysis on 3D building models reconstructed from airborne laser scanning data,” *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 66, no. 2, pp. 157–165, 2011.
- [5] M. Rutzinger, F. Rottensteiner, and N. Pfeifer, “A comparison of evaluation techniques for building extraction from airborne laser scanning,” *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 2, no. 1, pp. 11–20, 2009.
- [6] 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,” *Information Fusion*, vol. 6, no. 4, pp. 283–300, 2005.
- [7] D. Lee, K. Lee, and S. Lee, “Fusion of LIDAR and imagery for reliable building extraction,” *Photogrammetric Engineering & Remote Sensing*, vol. 74, no. 2, pp. 215–226, 2008.
- [8] W. Song and T. Haithcoat, “Development of comprehensive accuracy assessment indexes for building footprint extraction,” *IEEE Transactions on Geoscience and Remote Sensing*, vol. 43, no. 2, pp. 402–404, 2005.
- [9] J. Shufelt, “Performance evaluation and analysis of monocular building extraction from aerial imagery,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 21, no. 4, pp. 311–326, 1999.
- [10] G. Foody, “Status of land cover classification accuracy assessment,” *Remote Sensing of Environment*, vol. 80, no. 1, pp. 185–201, 2002.
- [11] O. Henricsson and E. Baltsavias, “3-D building reconstruction with ARUBA: A qualitative and quantitative evaluation,” in Gruen, A., Baltsavias, E.P., Henricsson, O. (Eds), *Automatic Extraction of Man-Made Objects from Aerial and Space Images (II)*, Monte Verit, Birkhuser Basel, 1997, pp. 65–76.
- [12] C. Zeng, J. Wang, and B. Lehrbass, “An evaluation system for building footprint extraction from remotely sensed data,” *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 6, no. 3, pp. 1640–1652, 2013.
- [13] J. Park, I. Y. Lee, Y. Choi, and Y. J. Lee, “Automatic extraction of large complex buildings using LIDAR data and digital maps,” *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, vol. XXXVI, no. 3/W4, pp. 148–154, 2006.
- [14] L. Yong and W. Huayi, “Adaptive building edge detection by combining LIDAR data and aerial images,” *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, vol. 37, no. part B1, pp. 197–202, 2008.
- [15] F. Rottensteiner, “Consistent estimation of building parameters considering geometric regularities by soft constraints,” *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, vol. 36, no. 3/W4, pp. 13–18, 2006.
- [16] N. Demir, D. Poli, and E. Baltsavias, “Extraction of buildings using images & LIDAR data and a combination of various methods,” *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, vol. 38, no. part 3/W4, pp. 71–76, 2009.
- [17] T. Vu, F. Yamazaki, and M. Matsuoka, “Multi-scale solution for building extraction from LIDAR and image data,” *International Journal of Applied Earth Observation and Geoinformation*, vol. 11, no. 4, pp. 281–289, 2009.
- [18] L. Chen, T. Teo, C. Hsieh, and J. Rau, “Reconstruction of building models with curvilinear boundaries from laser scanner and aerial imagery,” *Lecture Notes in Computer Science*, vol. 4319, pp. 24–33, 2006.
- [19] Q. Zhan, M. Molenaar, K. Tempfli, and W. Shi, “Quality assessment for geo-spatial objects derived from remotely sensed data,” *International Journal of Remote Sensing*, vol. 26, no. 14, pp. 2953–2974, 2005.
- [20] R. J. You and B. C. Lin, “A quality prediction method for building model reconstruction using LIDAR data and topographic maps,” *IEEE Transactions on Geoscience and Remote Sensing*, vol. 49, no. 9, pp. 3471–3480, 2011.
- [21] L. Cheng, J. Gong, M. Li, and Y. Liu, “3D building model reconstruction from multi-view aerial imagery and LIDAR data,” *Photogrammetric Engineering & Remote Sensing*, vol. 77, no. 2, pp. 125–139, 2011.
- [22] P. Dorninger and N. Pfeifer, “A comprehensive automated 3D approach for building extraction, reconstruction, and regularization from airborne laser scanning point clouds,” *Sensors*, vol. 8, no. 11, pp. 7323–7343, 2008.
- [23] B. Sirmacek, H. Taubenböck, P. Reinartz, and M. Ehlers, “Performance evaluation for 3-D city model generation of six different DSMs from air- and spaceborne sensors,” *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 5, no. 1, pp. 59–70, 2012.
- [24] M. Awrangjeb and C. S. Fraser, “Rule-based segmentation of LIDAR point cloud for automatic extraction of building roof planes,” *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, vol. II, no. 3/W1, pp. 1–6, 2013.
- [25] M. Awrangjeb and C. S. Fraser, “Automatic segmentation of raw LIDAR data for extraction of building roofs,” *Remote Sensing*, in press, 2014.
- [26] F. Rottensteiner, G. Sohn, J. Jung, M. Gerke, C. Baillard, S. Benitez, and U. Breitkopf, “The ISPRS benchmark on urban object classification

- and 3D building reconstruction,” *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, vol. 1-3, pp. 293–298, 2012.
- [27] J. Shan and S. Lee, “Quality of building extraction from IKONOS imagery,” *Journal of Surveying Engineering*, vol. 131, no. 1, pp. 27–32, 2005.
- [28] L. Ragia and S. Winter, “Contributions to a quality description of areal objects in spatial data sets,” *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 55, no. 3, pp. 201–213, 2000.
- [29] J. Meidow and H. F. Schuster, “Voxel-based quality evaluation of photogrammetric building acquisitions,” *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, vol. 36, no. part 3/W24, pp. 117–122, 2005.
- [30] D. Akca, M. Freeman, I. Sargent, and A. Gruen, “Quality assessment of 3D building data,” *The Photogrammetric Record*, vol. 25, no. 132, pp. 339–355, 2010.
- [31] J. X. Zhang and X. G. Lin, “Filtering airborne LIDAR data by embedding smoothness-constrained segmentation in progressive TIN densification,” *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 81, no. 1, pp. 44–59, 2013.
- [32] M. Satari, F. Samadzadegan, A. Azizi, and H. G. Maas, “A multi-resolution hybrid approach for building model reconstruction from LIDAR data,” *The Photogrammetric Record*, vol. 27, no. 139, pp. 330–359, 2012.
- [33] B. Xiong, S. Oude Elberink, and G. Vosselman, “A graph edit dictionary for correcting errors in roof topology graphs reconstructed from point clouds,” *ISPRS Journal of Photogrammetry and Remote Sensing*, <http://dx.doi.org/10.1016/j.isprsjprs.2014.01.007>, 2014.
- [34] N. Pfeifer, M. Rutzinger, F. Rottensteiner, W. Muecke, and M. Hollaus, “Extraction of building footprints from airborne laser scanning: comparison and validation techniques,” in *Proc. Urban Remote Sensing Joint Event*, Paris, France, 2007, pp. 1–9.
- [35] M. Awrangjeb, M. Ravanbakhsh, and C. S. Fraser, “Building detection from multispectral imagery and LIDAR data employing a threshold-free evaluation system,” *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, vol. 38, no. part 3A, pp. 49–55, 2010.
- [36] M. Awrangjeb, C. Zhang, and C. S. Fraser, “Automatic extraction of building roofs using LIDAR data and multispectral imagery,” *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 83, no. 9, pp. 1–18, 2013.
- [37] A. Jochem, B. Höfle, V. Wichmann, M. Rutzinger, and A. Zipf, “Area-wide roof plane segmentation in airborne LIDAR point clouds,” *Computers, Environment and Urban Systems*, vol. 36, no. 1, pp. 54–64, 2012.
- [38] X. K. Zhu and T. Toutin, “Land cover classification using airborne LIDAR products in beaupoirt, Québec, Canada,” *International Journal of Image and Data Fusion*, vol. 4, no. 3, pp. 252–271, 2013.
- [39] A. Turlapaty, B. Gokaraju, Q. Du, N. H. Younan, and J. V. Aanstoos, “A hybrid approach for building extraction from spaceborne multi-angular optical imagery,” *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 5, no. 1, pp. 89–100, 2012.
- [40] J. X. Zhang, X. G. Lin, and X. G. Ning, “SVM-based classification of segmented airborne LIDAR point clouds in urban areas,” *Remote Sensing*, vol. 5, no. 8, pp. 3749–3775, 2013.
- [41] C. Senaras, M. Ozay, and F. T. Y. Vural, “Building detection with decision fusion,” *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 6, no. 3, pp. 1295–1304, 2013.
- [42] M. Awrangjeb, C. Zhang, and C. S. Fraser, “Building detection in complex scenes thorough effective separation of buildings from trees,” *Photogrammetric Engineering & Remote Sensing*, vol. 78, no. 7, pp. 729–745, 2012.
- [43] C. Liu, B. Shi, X. Yang, N. Li, and H. Wu, “Automatic buildings extraction from LIDAR data in urban area by neural oscillator network of visual cortex,” *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 6, no. 4, pp. 2008–2019, 2013.
- [44] S. Sun and C. Salvaggio, “Aerial 3D building detection and modeling from airborne LIDAR point clouds,” *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 6, no. 3, pp. 1440–1449, 2013.
- [45] M. Cramer, “The DGPF test on digital aerial camera evaluation - overview and test design,” *Photogrammetrie Fernerkundung Geoinformation*, vol. 2, pp. 73–82, 2010.
- [46] Barista. (2011) The barista software (accessed 21 august, 2013). [Online]. Available: www.baristasoftware.com.au
- [47] M. Awrangjeb and C. S. Fraser, “Automatic and threshold-free evaluation of 3D building roof reconstruction techniques,” in *Proc. IEEE International Geoscience and Remote Sensing Symposium*, Melbourne, Australia, 2013, pp. 3970–3973.
- [48] D. Lee, J. Shan, and J. Bethel, “Class-guided building extraction from IKONOS imagery,” *Photogrammetric Engineering & Remote Sensing*, vol. 69, no. 2, pp. 143–150, 2003.
- [49] A. F. Habib, R. Zhai, and K. Changjae, “Generation of complex polyhedral building models by integrating stereo-aerial imagery and LIDAR data,” *Photogrammetric Engineering & Remote Sensing*, vol. 76, no. 5, pp. 609–623, 2010.
- [50] K. Khoshelham, Z. Li, and B. King, “A split-and-merge technique for automated reconstruction of roof planes,” *Photogrammetric Engineering & Remote Sensing*, vol. 71, no. 7, pp. 855–862, 2005.
- [51] A. Sampath and J. Shan, “Segmentation and reconstruction of polyhedral building roofs from aerial LIDAR point clouds,” *IEEE Transactions on Geoscience and Remote Sensing*, vol. 48, no. 3, pp. 1554–1567, 2010.
- [52] M. Awrangjeb and G. Lu, “An improved curvature scale-space corner detector and a robust corner matching technique for transformed image identification,” *IEEE Transactions on Image Processing*, vol. 17, no. 12, pp. 2425–2441, 2008.



Dr. Mohammad Awrangjeb is currently a Senior Research Fellow at Gippsland School of Information Technology of Monash University. Before joining Monash University, he worked as a Research Fellow in University of Melbourne in 2008–12. He obtained his Ph.D. in 2009 from Monash University, Australia. He worked as a lecturer in University of Asia-Pacific, Dhaka in 2002; as a Research Scholar in the National University of Singapore in 2002–04; and as a lecturer in American International University–Bangladesh in 2004–05. He is a recipient of Discovery Early Career Researcher Award by the Australian Research Council for the period 2012–15. Dr. Awrangjeb’s research interest includes automatic feature extraction and matching, multimedia security and image processing. He is a coauthor of more than 30 research articles in internationally renowned journals and conferences.



Prof. Clive S. Fraser is currently a Program Science Director in the Cooperative Research Centre for Spatial Information, and a Professorial Fellow in the Dept. of Infrastructure Engineering at the University of Melbourne. Prof. Frasers particular areas of research interest lie in digital close-range photogrammetry, including 3D forensic analysis and accident reconstruction, industrial measurement systems, and the metric exploitation of high-resolution satellite imagery. In recognition of his academic and professional work he has earned numerous international awards and has authored more than 300 scientific publications.