Quantitative Evaluation Strategies for Urban 3D Model Generation from Remote Sensing Data

Linh Truong-Hong[(1)] and Debra F. Laefer[(2)]

[(1)] Research scientist, Urban Modelling Group (UMG), School of Civil, Structural and Environmental Engineering (CSEE), University College Dublin (UCD), Dublin 4, Ireland. Email: *linh.truonghong@ucd.ie*, *corresponding author*

[(2)] Associate Professor, Head UMG, CSEE (and Earth Institute), UCD, Dublin 4, Ireland. Email: *debra.laefer@ucd.ie*

Abstract:

Over the last decade, several automatic approaches have been proposed to reconstruct 3D building models from aerial laser scanning (ALS) data. Typically, they have been benchmarked with datasets having densities of less than 25 points/m^2. However, these test datasets lack significant geometric points on vertical surfaces. With recent sensor improvements in airborne laser scanners and changes in flight path planning, the quality and density of ALS data have improved significantly. The paper presents quantitative evaluation strategies for building extraction and reconstruction when using these dense datasets. The evaluation strategies are to measure not only the capacity of a method to detect and reconstruct individual buildings but also the quality of the reconstructed building models in terms of shape similarity and positional accuracy.

Keywords: LiDAR data; Aerial Laser Scanning; Point Cloud; Building Detection; Building reconstruction; Evaluation Strategy

1. Introduction

Urban, three-dimensional (3D) model reconstruction is a rapidly growing topic for a large range of applications from environmental planning and computational simulation to disaster management and location-based services. Currently, the raw data used for 3D building reconstruction at a city-scale is obtained from various resources through a wide range of techniques including photogrammetry and high-density aerial laser scanning (ALS), and while 3D building reconstruction has been an active research topic for more than a decade, the field is still rapidly changing, as raw data improves and user expectations continue to grow, particularly with respect to façade details [1]. Currently, existing methods in 3D building model reconstruction from ALS data typically start by extracting data points from roofs and reconstructing roof models [2, 3]. A complete building model is then generated by

extruding roof outlines to a ground plane or extruding building footprints based on the building's height [4, 5]. These methods have the disadvantages of assuming (1) the availability of accurate a priori data and (2) the absence of overhanging roofs. Unfortunately neither assumption can be relied upon, as shown by Kaartinen et al. [6] and Rottensteiner et al. [7].

As part of the EuroSDR building extraction project, the quality, accuracy, feasibility, and economic aspects of semi-automatic and automatic building extraction techniques based on either ALS data or a combination of ALS data and aerial images were evaluated [6]. The project consisted of three test sites, each with typical ALS density of less than 20 points/m^2. The project concluded that only ALS data with allowable accuracy could be used for building extraction, although no threshold was specified [6]. In an ISPRS benchmark project to evaluate building detection accuracy, tree detection, and 3D building reconstruction methods were assessed. Data (aerial images and ALS point clouds) from two sites were used where ALS density was less than 10 points/m^2 [7]. The results showed that existing methods satisfactorily detected buildings larger than 50m^2, but struggled with Level of Detail (LoD) 2 [8] for smaller buildings and in urban areas, as a direct outgrown of the absence of façade data. Thus, a contest was devised as part of the IQMULUS project's workshop on Processing Large Geospatial Data held in Cardiff University on 8 July 2014 to test building detection and 3D building reconstruction algorithms using high-density data in an urban area. Specifically, the competition was based on dense ALS data (typically 225 points/m^2) for 1km^2 area in Dublin Ireland. The dataset was acquired by the Urban Modelling Group at the University College Dublin to maximize vertical surface data acquisition [9].

The contest included two tasks, from which participants could submit for either or both. The goal of Task A was to identify and extract the data points of each individual building in the study area and reconstruct the buildings' boundary lines. The evaluation focused on identifying the level of deviation in location, the level of shape similarity, and the positional accuracy of the detected buildings with respect to the ground truth (GT). In Task B, the participants were asked to reconstruct 3D building models from a subset of the Task A data. The evaluation was benchmarked not only on the achieved LoD of the submitted 3D building models but also on their final quality through geometric accuracy. The participants were required to submit results to the competition organizers for evaluation based on GT. This paper reports the data, task objectives, evaluation strategies, and the results achieved to date.

2. Data

The test area is approximately 1km^2 and consists of 205 blocks, each of which may contain in excess of a dozen buildings per block, as shown in Fig. 1. The area is comprised of mostly four-storey masonry (brick and stone) buildings [10]. The buildings are generally rectilinear with only a limited amount of ornamentation. The typical building is 11–15m in height, less than 5m in width and 6–10m in length [10]. They are mostly closely spaced or abutting each other, with some sharing an adjoining wall, commonly referred to as a "party wall".

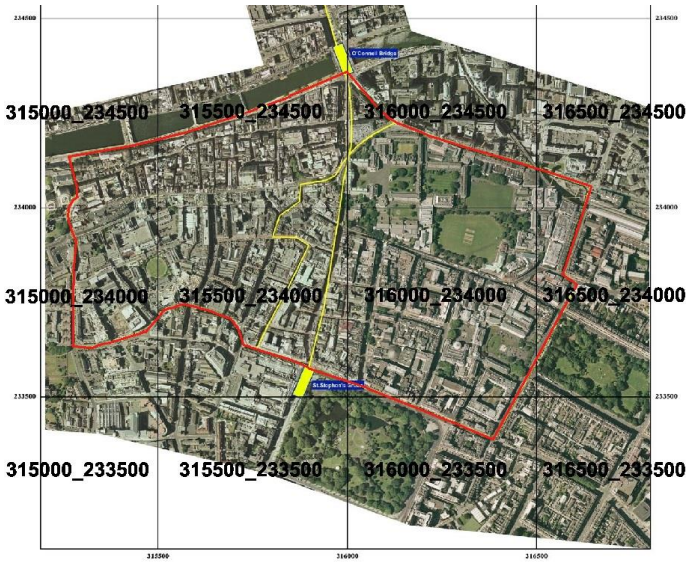


Fig. 1. Acquired ALS area in Dublin central and ALS tiles (contest area outlined in red)

The dataset was acquired by ALS using the FLI-MAP 2 system, which generated 1000 pulses for each scan line. The system operated at a scan angle of 60 degrees, with an angular spacing of 60/1000 degrees between pulses, which is roughly equal to one milli-radian. The quoted accuracy of the FLI-MAP 2 system is 8 cm in the horizontal plane and 5cm in the vertical direction, including both laser range and navigational errors [11]. Acquired points were provided in a global coordinate system with reference to the National Irish Grid (Irish Grid), relating to the use of a Global Navigation Satellite System (GNSS) to determine the aircraft position during scanning. The FLI-MAP 2 system is capable of recording up to four echoes for each emitted pulse and spectral data with intensity values. Unfortunately, the RGB values were not available due to an equipment malfunction during the data collection.

Since the study area is urban, besides constrained flight planning (e.g. flight path and altitude), three major factors impact the vertical façade data capture: building geometry,

street widths, and street layout. The flight path was designed to maximize the vertical data of the building walls [9]. The dominant directions of the flight tracks were chosen as north-east, north-west, south-east and south-west (Fig. 2). The flight attitude varied between ~380-480m (as low as possible with respect to approval by the Irish Aviation Authority), with an average elevation of ~400m. A total of 44 flight strips with 2,823 flight path points were collected during data acquisition. With a scan angle of 60 degrees, the distance between two adjacent flight paths projected onto the ground was approximately 70 m (Fig. 2). As a result, a total of 370,154 scan lines were acquired, with approximately 225 million data points, giving a typical point density of 225 points/m^2. Each emitted pulse was recorded with up to four echoes. The echo distribution is shown in Table 1. The vast majority of points were first echoes. Secondary echoes constituted only a small portion of the points, as the overwhelming majority of surfaces in the study area were solid (in the form of streets and buildings). For further information about this ALS data, participants are referred to the following reference [11]. The data set was organized into 9 tiles, each covering 500m x 500m (Fig. 1) and is publicly available for download [12].

Fig. 2. Designed flight path of data collection

Table 1. Echo distribution of acquired ALS points

Echo	**Count**	**Percentage**
1st	217,497,975	96.33%
2nd	7,902,595	3.50%
3rd	383,840	0.16%
4th	4,028	0.001784%
Total	225,788,438	100%

3. Tasks and evaluation

3.1 Task A

Task A was to identify and extract the data points for each building in the study area and reconstruct the buildings' boundary lines. In an urban region, buildings have little, if any, space between them (Fig. 3). This makes automated detection challenging, as one cannot rely upon a distinctive change in elevation around each building as a clear delineator between structures.

Fig. 3. Example of the challenge of extracting an individual building in urban areas [13]

Participants were given 3 months and asked to submit the results with segregated ALS datasets for each building. The building outlines were to be generated by the contest organizers for evaluation. The GT used to evaluate the submissions consisted of 2D footprints generated by the Ordnance Survey Ireland (OSI).

The organizer's method (based on an angle criterion and 2D cell grids) was used to extract points on the boundary of the ALS data of each building provided by participants, after these points were projected onto the ground. Subsequently, the boundary points were segmented to generate a set of boundary lines describing the building outline; details of this method are given in section 4.

The extracted building outlines were overlaid with GT for evaluation using the procedure proposed by Preifer et al [14]. In this, the evaluation identified the level of locational deviation, the level of shape similarity, and the positional accuracy of the detected buildings

with respect to GT. The location was determined based on a match rate between the detected building and the reference data. While this can be done using either a threshold-based system or a threshold-free one [15-17], the former requires an overlap threshold to determine the most important quality indicators, which may bias the results. The latter system does not require any overlap threshold, and the evaluation result performs with various levels of overlap (from 0% to 100%). While the threshold-free system provides a more stable, object-based performance [15], the approach cannot determine a building-by-building identification for reconstructing building outline, which was the objective of this task [18]. Consequently, a threshold-based system with an object-based evaluation was employed, which mirrors the strategy adopted for ISPRS benchmark project [7].

The overall building detection was evaluated in terms of completeness, correctness, and quality of the results. The concept of true positives, false negatives, and false positives (as proposed by Rutzinger et al. [17]) was be applied by using group classification as either *object* or *background*. A True Positive (TP) was defined as an entity, which was classified as the *object group* that also corresponded to an object in the reference model. A False Negative (FN) was defined as an entity classified as an *object group* in the reference model corresponding to a *background group*. A False Positive (FP) was defined as an entity classified as an *object group* that does not correspond to an object in the reference model. Therefore, evaluation quantities could be calculated using Eq.s 1-3.

$$\mathrm{Comp} = \frac{\|\mathrm{TP}\|}{\|\mathrm{TP}\| + \|\mathrm{FN}\|} \qquad \text{Eq. 1}$$

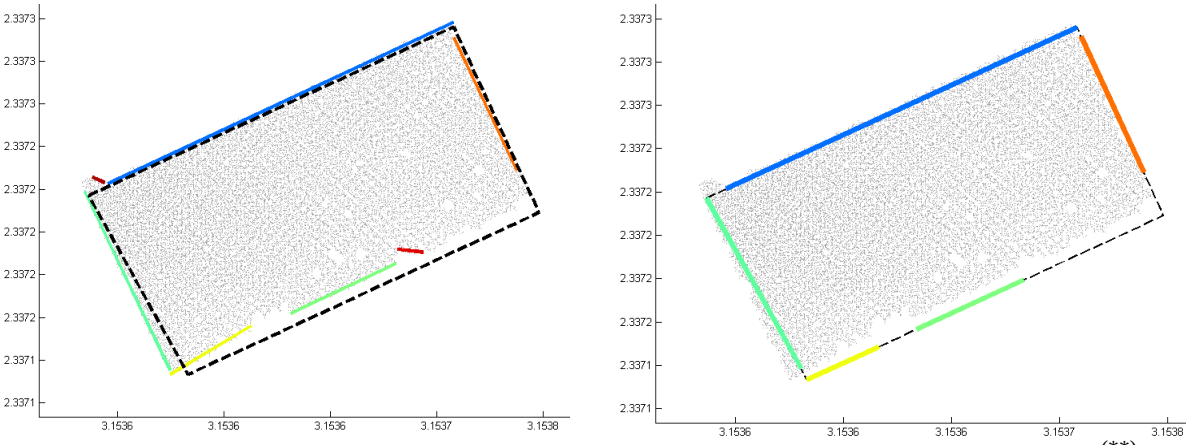
$$\mathrm{Corr} = \frac{\|\mathrm{TP}\|}{\|\mathrm{TP}\| + \|\mathrm{FP}\|} \qquad \text{Eq. 2}$$

$$\mathrm{Quality} = \frac{\|\mathrm{TP}\|}{\|\mathrm{TP}\| + \|\mathrm{FP}\| + \|\mathrm{FN}\|} \qquad \text{Eq. 3}$$

The extracted building was considered as a TP, if it overlapped the reference building by a minimum of 50%. The extracted building was classified as a FP, if the building was not in the reference data or the area of the extracted building overlapped less than 50% of the reference building area. Finally, the reference building was classified as FN, if the reference building did not match the extracted building to any extent.

Additionally, to measure quality of each extracted building, the matched line (L_{GT}) of the extracted boundary line (L_{EX}) was determined from the GT. From a set of boundary lines of

the extracted building, the L_{GT} was considered the matched line, if all of the following criteria were met: (1) the averaged distances between the two end points of L_{EX} to L_{GT} is the minimum of candidate lines and no larger than the distance threshold by 3m [7]; (2) the angle between L_{EX} and L_{GT} was less than the angle threshold, where the value of 30 degrees was adopted; (3) the overlap length ($L_{overlap}$) between L_{EX_proj} and L_{GT} was greater than zero, where L_{EX_proj} was a projection of L_{EX} onto L_{GT}. Fig.4 illustrates the result of determining the nearest lines of the extracted boundary lines and the overlap lengths.



a) Overlaid boundary lines from extracted and GT buildings(*)

b) Illustration of overlap lengths(**)

* Dashed lines are boundary lines (L_{GT}) of the ground truth building and solid lines are the boundary lines (L_{EX}) of the extracted building.

** Colour solid lines illustrate overlap lengths of L_{EX_proj} and L_{GT}

Figure 4. Illustration of determining the best matched ground truth line (L_{GT}) of the extracted boundary line (L_{EX}) and overlap length

Furthermore, to measure a shape similarity, area differences and an overlap perimeter between the extracted building and the GT were introduced. Notably, only the extracted buildings as TP were considered for evaluation. An area of each extracted building was computed from the boundary points extracted from the data points of the building by the proposed method. The quality indicators measuring the difference in areas involved summing the absolute, mean, and standard deviation differences (see Eq.s 4-6).

$$\sum \delta A = \sum_{i=1}^{n} \left(A_{GTi} - A_{EXi} \right) \qquad \text{Eq. 4}$$

$$\overline{A} = \frac{\sum_{i=1}^{n} \left(A_{GTi} - A_{EXi} \right)}{n} \qquad \text{Eq. 5}$$

$$s = \sqrt{\frac{\sum_{i=1}^{n}\left[(A_{GTi} - A_{EXi})^2 - \bar{A}\right]}{n-1}} \quad \text{Eq. 6}$$

where A_{GTi} and A_{EXi} are the areas of the extracted buildings and GT building, respectively, and n is the number of the buildings that were TPs.

Subsequently, an overlap perimeter for each building was established by applying Eq. 7.

$$E_{overlap} = \frac{\sum_{j=1}^{m} L_{overlap\_j}}{L_{GT}} \quad \text{Eq. 7}$$

where $L_{overlap\_j}$ is the overlap length in the extracted building, m is the number of the overlap lengths, L_{GT} is the total boundary length of the ground truth building.

A positional accuracy can be described in term of the accuracy and conciseness of the extracted building, which is performed through distance and orientation errors between the extracted boundary lines and the ground truth boundary lines [19]. The distance error (d) was defined as the average distance between two end points of L_{EX} to its matched line derived from L_{GT}, while the orientation error (α) was the angle between L_{EX} and L_{GT}. The overall distance and orientation errors of the extracted buildings in the data set were expressed as Eq.s 8 and 9:

$$E_{dist} = \frac{\sum_{j=1}^{m} L_{EXj} d_j}{\sum_{j=1}^{m} L_{EXj}} \quad \text{Eq. 8}$$

$$E_{orient} = \frac{\sum_{j=1}^{m} L_{EXj} \alpha_j}{\sum_{j=1}^{m} L_{EXj}} \quad \text{Eq. 9}$$

where L_{EXj} was the length of the extracted boundary line, and d_j and α_j were the distance and angle between the extracted and reference boundary lines (respectively), and m was the number of boundary lines in the extracted building of interest. In these error measurements, the length of the extracted boundary line was introduced to avoid a heavy penalization for

short extracted boundary lines [19]. Subsequently, average and standard deviation can be used to measure a distribution of these quantities.

3.2 Task B

The goal of Task B was to extract and reconstruct 3D building models from a subset of the data, as shown in Fig. 4. The goal of the 3D model was to achieve LoD3 as defined according to the CityGML standard definition [8] but with a minimum achievement requirement of LoD2. Notably, only building geometry was considered for evaluating the submitted results. For LoD2, the generated building models were to consist of all visible exterior walls, as well as major roof components (e.g. roof surfaces and small dormers) but without roof overhangs or façade details. For LoD3, the reconstructed building models were to include all façade openings with respect to doors and windows; however details of these openings (e.g. window frames or lintels) and other architectural detailing were not required.

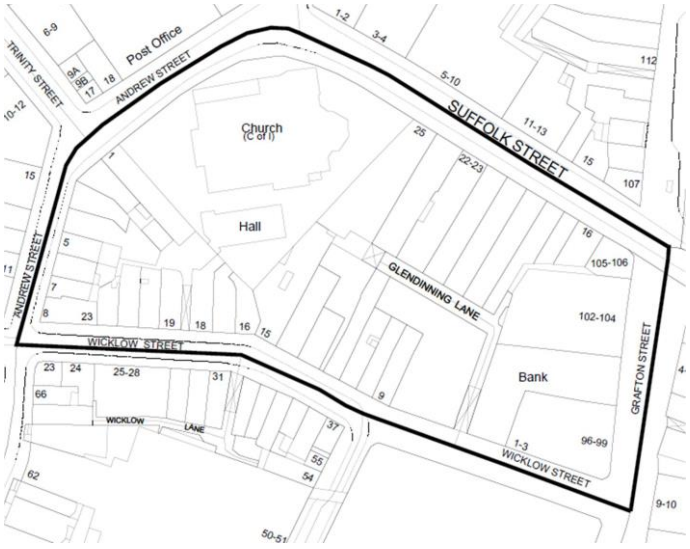


Fig. 5. Plan view of the Task B study area

The results were to be submitted in a DXF file format, compatible with a standard CAD platform. The geometry of each building model was to be stored as either a polygon or a 3D solid. Each component of the building model (e.g. exterior vertical walls or roofs) was to be arranged in a specific layer as part of the DXF file. In addition, the participants were asked to submit a brief description of the approach implemented to reconstruct the 3D building models.

The reference models including 3D building models with LoD3 were created based on independently measured architectural drawings published by Dublin City Council [20]. These drawings were submitted by the building owners and were generated from traditional survey methods. To evaluate the submitted results, the generated building models were initially mapped onto the reference models based upon each building's centroid, as computed using the building corners at the ground level and a normal vector of the street façade. Then the iterative closest point (ICP) algorithm [21] was employed to obtain a final registration model, which was minimized using the root mean square of the Euclidean distances between the generated building corners and the nearest corners derived from the reference model at the ground level (Fig. 5). Similar to Task A, the evaluation process was to measure not only the capacity of the method to detect and reconstruct individual buildings but also the quality of the generated building models. The former metric is described through completeness, correctness, and quality quantities (as defined in Task A). The later metric involves shape matching and accuracy with respect to the reference building. The evaluation process was divided into evaluation of two topics: (1) generated building surfaces consisting of vertical façade surfaces and roof planes and (2) opening details (only applied from the generated building models with LoD3).

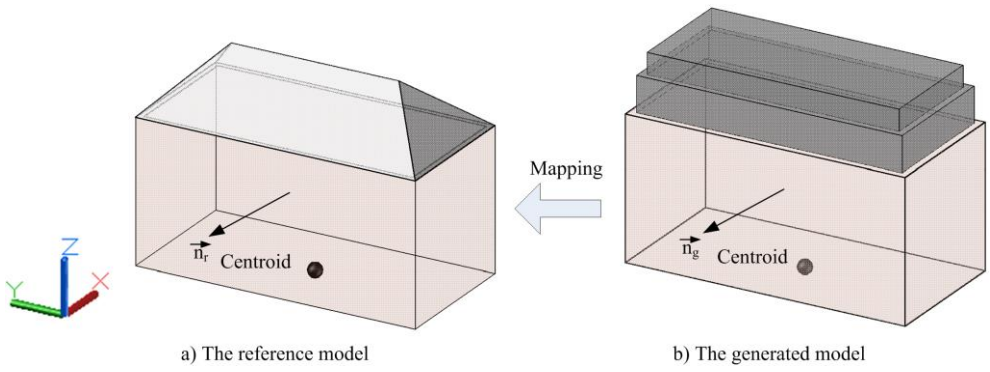


Fig. 6. Mapping the generated model to the reference model for building evaluation

To assess the quality of the reconstructed building models, indicators involving a shape similarity (standing for an identical shape) and errors with respect to distance and orientation (describing the building model's accuracy) were used. First, similar to Task A, after mapping the generated model to the reference one, the matched plane (S_{RM}) of the generated model's plane (S_{GM}) was determined from the reference model. The S_{RM} was then matched plane, if (1) the angle between the normal vectors of the S_{RM} and S_{GM} was less than the angle threshold; (2) the S_{RM} had the shortest distance to S_{GM}; and (3) the overlap area between the

projected S_{GM} onto the S_{RM} was greater than zero. Notably, if the S_{GM} overlapped multiple planes of the reference model, the S_{GM} was split into sub-planes corresponding to the number of overlapping planes. Second, the shape similarity was computed as per Eq. 7, where the overlap area was used. Third, the distance error between S_{GM} and S_{RM} was calculated as the average distance from the vertices of S_{GM} to their projections onto S_{GT}. Next, the orientation error was considered as the angle between the normal vectors of S_{GM} and S_{RM}. Finally, the distance and orientation errors of each reconstructed model were computed by using Eqs. 8 and 9, where the representation of the surface area was replaced with a line segment length.

The building models with LoD3 were to be evaluated with respect to the quality of the reconstructed openings in terms of the number of openings and their locations versus those of the actual geometries. Validation metrics proposed by Truong-Hong [22] were used to measure any differences between the opening positions (openings' corners) and dimensions (width and height) through mean and standard deviation values. In order to get these measurement indicators, the generated building facades containing the openings were mapped onto those of the respective reference models, using the normal vector of the surface and the centroid surface at the ground level, which was computed from edges' coordinates at the ground (Fig. 6).

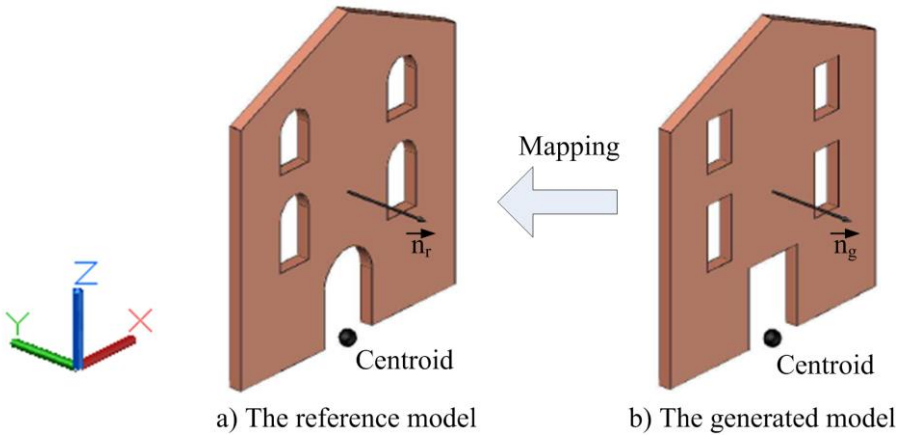


Fig. 7. Mapping openings to evaluate their accuracy

4. Building outline generation

As participants were only given 3 months, Task A was limited to submission of segregated ALS datasets for each building. However, the evaluation process required the building outline represented by a set of line segments. Thus, these were generated by the contest organizers. Figure 8 shows the applied workflow. Notably, the method only reconstructed primary

boundary lines for a length of boundary point segments larger than 1m and did not fill any gap between the generated boundary lines.

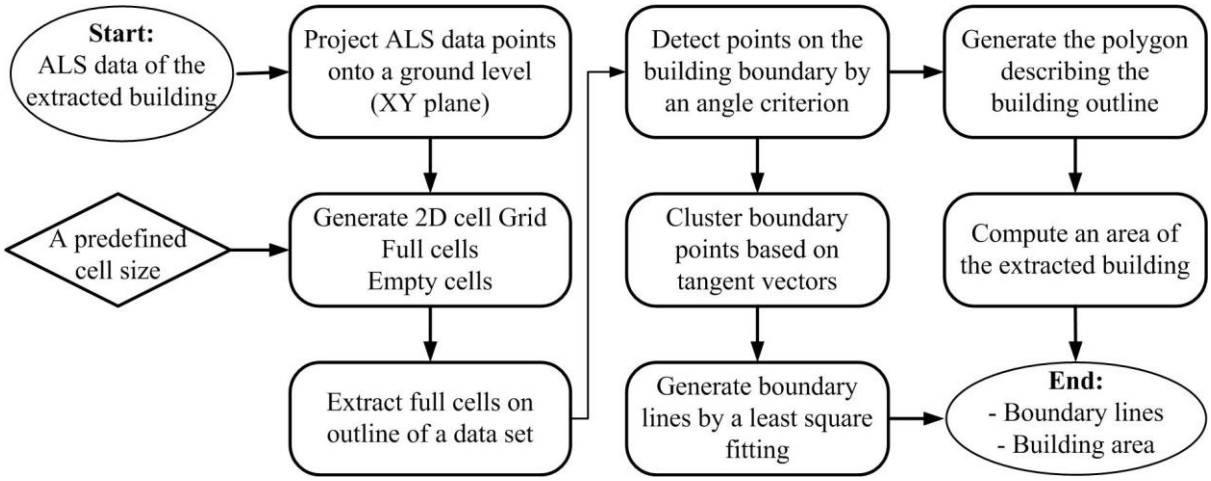


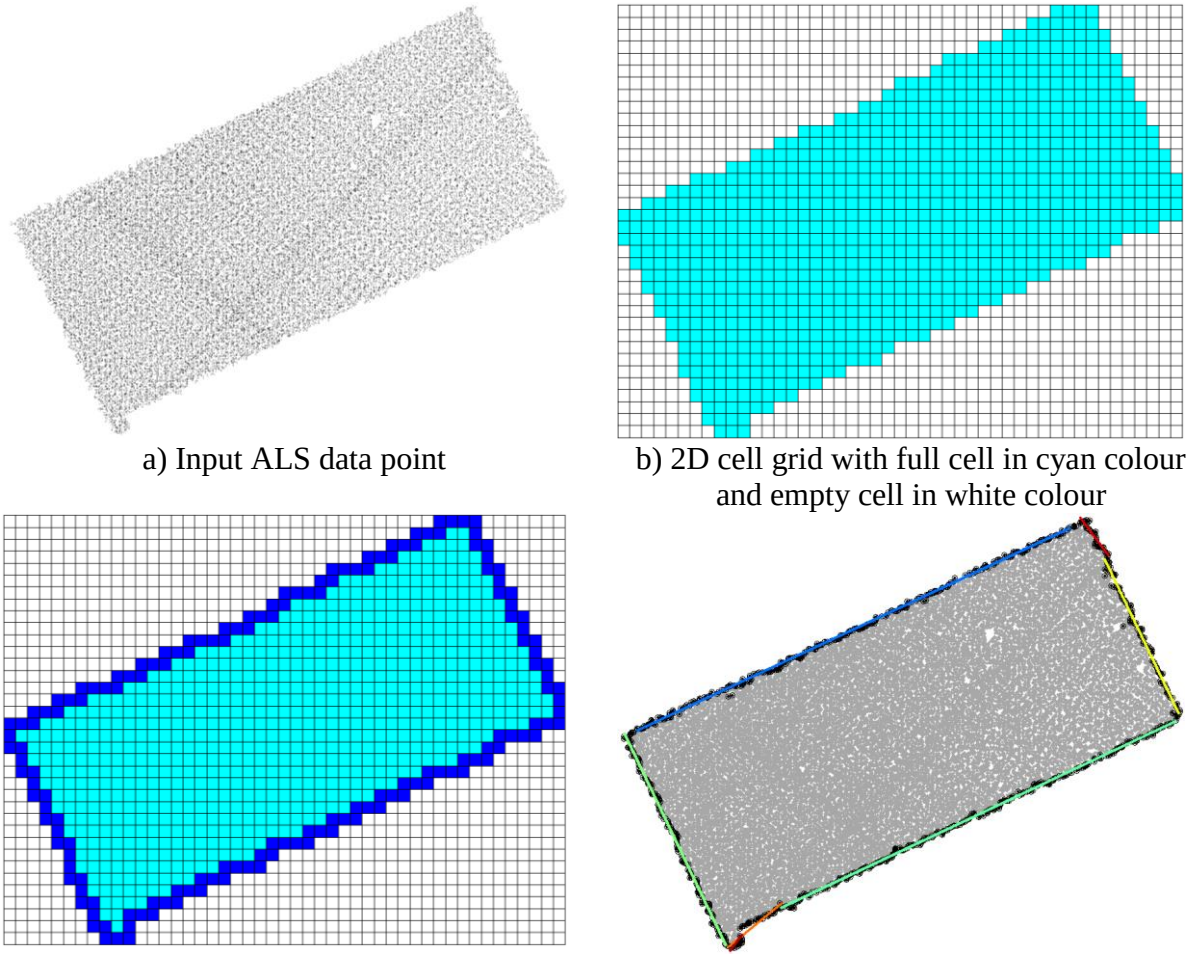
Figure 8. Workflow to generate boundary lines of extracted buildings

In the first stage of the proposed method, the ALS data points of each building were projected onto the ground level (known as the xy-plane in the global coordinate system). Data points on the building outline were a prerequisite for generating boundary lines. Various algorithms associated with the criteria have been proposed to extract those boundary points. Many use an angle criterion [23, 24], a half-disc criterion [25], or a Delaunay triangulation mesh [26, 27]. In this proposed method, the angle criterion was employed to extract the boundary points. However, an existing angle criterion is time consuming, because all points in the data set must be checked. To overcome this aspect, a combination of a 2D cell grid and an angle criterion was introduced.

After the ALS building data were projected onto the ground level (Fig. 9a), the 2D cell grid was employed to represent the data set, in which the cell size of 0.5m was empirically chosen. The cell was classified as either “full” or “empty”. The cell was “full”, if it contained at least one data point and was otherwise empty (Fig. 9b). As the aim of the algorithm was to reconstruct building boundary lines, the data points on each building’s outline had to be extracted. Unlike other angle criterion-based algorithms, in this method only data points that were contained in a full cell on the perimeter of the data set were checked as to their status as boundary points (Fig. 9c).

The data points within these full cells were considered candidate boundary points. The angle criterion was then employed to determine the boundary points from these candidate points.

With a given point, p_i selected from the candidate points, the k-nearest neighbouring (kNN) points, **q** was searched for a kNN of 20 points [23]. Subsequently, Cartesian coordinates of neighbouring points, **q**, were then transformed into relative cylindrical coordinates, with the local origin set at a given point p_i. If an angle between two consecutive neighbouring points, $\alpha_{i,i+1} = \angle q_i q_{i+1}$, (the difference between their azimuths) exceeded a given threshold, the given point was classified as a boundary point. In this implementation, the angle threshold of 90 degrees was selected. For more details of this process, please refer to Truong-Hong et al. [23]. Results of the boundary point searching are shown in Fig. 9d.



a) Input ALS data point

b) 2D cell grid with full cell in cyan colour and empty cell in white colour

c) Extracted full cells containing candidate boundary points

d) Boundary points and boundary lines

Figure 9. Illustration of boundary lines of the extracted building

From a set of boundary points, the polygon describing the contour boundary of the building was generated to support the building area computation, where vertices of the polygon were considered the boundary points. Subsequently, an adopted region-growing method was employed to extract the boundary points belonging to the same boundary line. As part of this,

the tangent vector of each boundary point was computed. This was done by searching the kNN points of the given boundary points, and then a principal component analysis was employed to estimate the tangent vector. Based upon experience by the authors, a neighbourhood of 10 kNNs is sufficient at this step and was used in this implementation. An initial seeding point was randomly selected from the boundary points, and a 10 of kNN was searched. The kNN point(s) were considered as target point(s) (i.e. in the same region with the seeding point) and were added into the seeding point group, if the angle between the seeding point and the kNN point(s) was smaller than the angle threshold, which was empirically chosen as 5 degrees. The process was iteratively applied until all boundary points were selected as seeding points. Finally, the boundary points of each region were used to generate the boundary line by using a least squares method, where only segment lengths larger than 1m were considered. Notably, only the straight line segment was used to represent to the building outline, as illustrated in Fig. 9d.

5. Results and discussions

The contest was launched in May 2014. At the time of the IQmulus workshop (July 8), results for Task A were received from the Computer Graphics Systems Group at the Hasso-Platter-Institute at University of Potsdam; no submissions were received for Task B. The proposed method to detect 3D point clouds of the buildings was based on the point cloud topology and did not require per-point attributes or representative training data. The method adopted the smoothness constraint region growing proposed by Rabbani et al. [28] and the local point connectivity to segment LiDAR data set. The vegetation areas were grouped into small segments, while the other objects (e.g. building roofs) were grouped into larger segments. Next, a type of surface category of each segment was assigned based on the structure, size, and proximity of the segments. In this, large segments located below other ones were recognised as the ground, while the building points were determined from the remaining segments based on size and roughness. For details of the classification method see Richter et al. [29].

The method automatically extracted the data points belonging to each building, and a total of 1889 data files of the buildings were submitted. The building boundary lines were generated by using the above outlined organizers' method. The extracted building outlines were mapped onto the existing 2D footprint for evaluation (Fig. 7). Results of the location

evaluation (completeness, correctness, and quality) of the extracted building are showed in Tables 2 and 3.

Fig. 10. Overlay reconstructed building detection onto 2D footprints with areas filled in colour are the actual 2D footprints, and the black polygons are building outlines from the submission.

Table 2. Defining Assessment quantities

Quantities	Total buildings	Total area (m^2)
TP	794	228,466.8
FP	378	149,792.8
FN	41	11,744.38

Table 3. Resulted evaluation

Comparative quantities	Comparative results in term of the number of the buildings (%)	Comparative results in term of the building area (%)
Completeness	95.1	95.1
Correctness	67.7	60.4
Quality	65.5	58.6

As shown in Table 3, although the completeness was approximately 95%, correctness and quality were less than 70%. As part of the ISPRS benchmark project on urban object extraction [7], alternative methods tested with data from Toronto, Canada with only 6 points/m^2 were able to achieve 75.5% completeness and 70.1% correctness and quality.

Additionally, recent work by Awrangjeb et al. [30] on building detection with an object-based evaluation using Area 4 data from Toronto, Canada showed 100% completeness and 83.6% correctness and quality. The differences may be due to either the proposed method or the complexity of the built environment. In terms of the contest dataset from Dublin Ireland, the buildings have a high degree of similarity with respect to their geometry (building height and roof configuration) and many share at least one vertical wall. Thus, extraction of an individual building is difficult. For example, with the considered region in Fig. 8a, the submission extracted five buildings, but one (shown as a yellow polygon) is actually an extension and not an independent building. Moreover, when a roof involves several surfaces with similar geometric properties (e.g. size and orientation), the algorithm may fail to detect that these belong to one building. For example, in Fig. 8b, the building roofs have several flat surfaces, but the submitted method considered each flat surface as a separate building. In summary, although the test dataset had a very high density, extraction of individual urban buildings was still a major challenge due to architectural attributes of the study area. Arguably, accuracy of individual building data extraction can be improved by combining roof profiles and information of the building facade derived from the distribution of a point cloud on the façade.

a) Failure to extract point clouds of adjacent buildings (11 buildings from GT)

b) Over-extracted individual buildings (2 buildings from GT)

Note: Blue polygons describe 2D footprints; red polygons are contours of building outlines from the submission; and the yellow filled polygon in Fig. 11a showed the incorrect detection of the building outline, which is a part of one of the buildings.

Fig. 11. Problematic extraction of the building outline

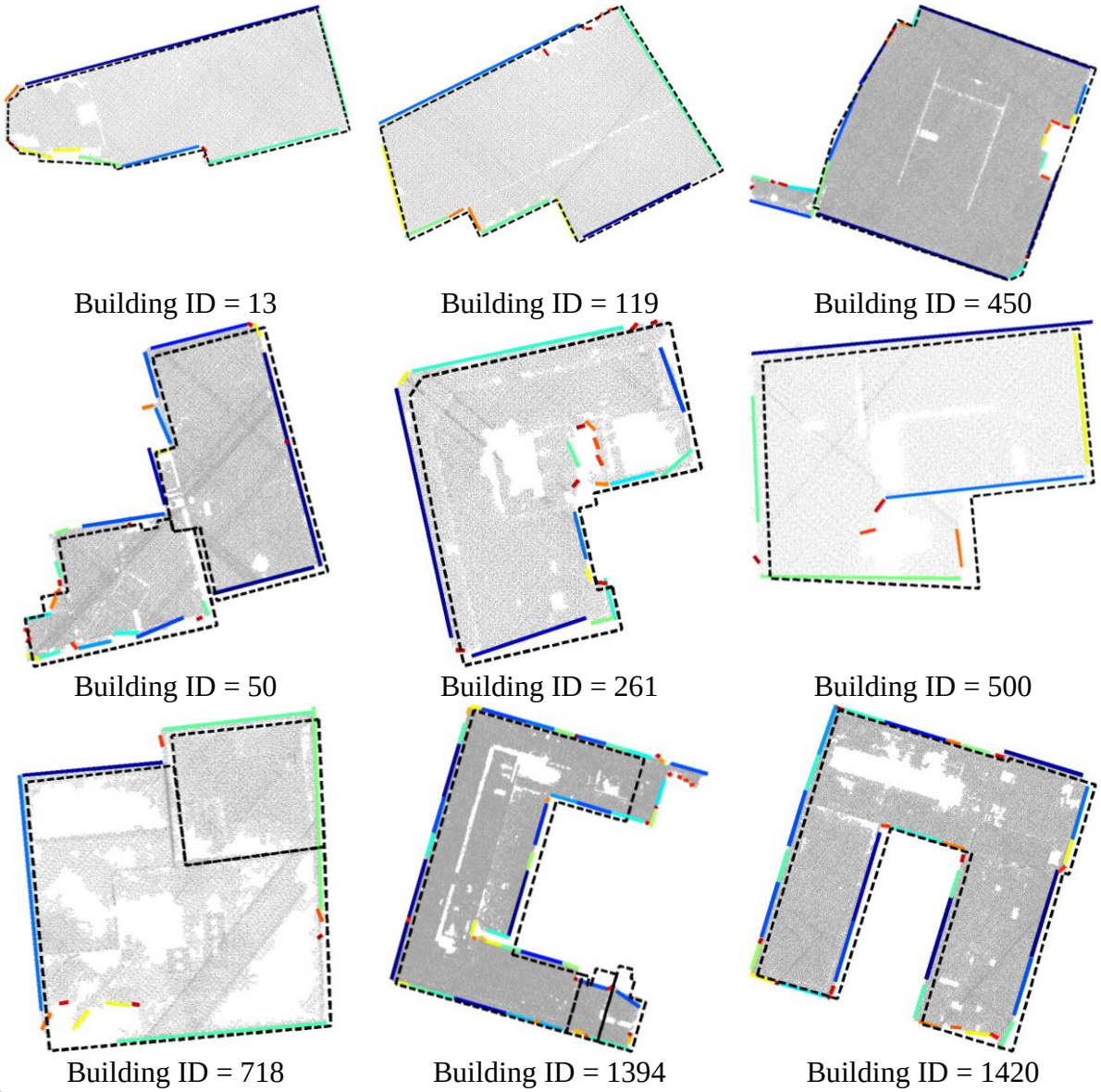
To evaluate quality of individual building detection, the data from 50 buildings were selected to represent various levels of complexity. Against these, the submitted files were compared in terms of shape similarity and positional accuracy. The boundary lines of each building were generated by the proposed method. The summary evaluation is shown in Table 4. Although the absolute average area difference was relatively small (3.59m^2) when compared to the reference building, the results showed the difference in areas varied in a wide range, where the standard deviation (std) was 66.75m^2, and the minimum and maximum differences were respectively -222.07m^2 and 206.44m^2. Otherwise, the submitted results showed good agreement between the extracted building outlines and the reference ones. The average overlap perimeter was 0.76 (std = 0.13), while the distance and orientation errors were 1.10m (std = 0.89m) and 3.09 degrees (std = 1.75 degrees), respectively.

Table 4. Resulted evaluation in terms of a shape similarity and positional accuracy from 50 selected buildings

Aspect	Shape similarity		Positional accuracy	
	Area difference (m^2)	Overlap perimeter rate	Distance error (m)	Orientation error (degrees)
Average	3.59	0.76	1.10	3.09
Minimum	-222.07	0.41	0.14	0.63
Maximum	206.44	0.97	4.21	8.39
Standard deviation	66.75	0.13	0.89	1.75

To illustrate detailed quality of the extracted building, nine buildings representative of typical cases in the study area were selected for further evaluation. An overlap of the extracted boundary lines and GT ones are illustrated in Fig 12, and quantification of the evaluations are shown in Table 5. Absolute area differences between the extracted buildings and GT was small, with most less than 16m^2; however, the difference of Building ID 1394 was 109.77m^2 due to over-extraction, thereby unintentionally incorporating data points from 3 other buildings. The average and standard deviation of the absolute area differences of the 9 buildings were -16.31m^2 and 37.08m^2, respectively; however, if the values from Building ID 1394 are considered as outliers, the average of absolute area differences was only -4.62m^2 (std = 12.93m^2). Furthermore, a high rate of overlap length between the boundary lines of the extracted building and GT ones was found, with the average exceeding 80%. Building ID 52, 718 and 1394 had small overlap length rates (< 0.72%), because the data set contained points from other buildings. In terms of positional accuracy, the evaluation shows that both distance and orientation errors were small, with the average of 1.25m (std = 1.18m) and 2.15 degrees

(std = 1.16 degrees), respectively. If Building ID 1394 is removed, the average of distance error is reduced further to only 0.882m (std = 0.430m).



* Dashed lines are boundary lines of GT and solid lines are the extracted buildings

Figure 12. Illustration of boundary lines of the extracted buildings of interest (*)

Table 5. Resulted evaluation in term of a shape similarity and positional accuracy

	Shape similarity				Positional accuracy	
Building Id	A_{EX} (m^2)	A_{GT} (m^2)	δA (m^2)	Overlap length rate	Distance error (m)	Orientation error (degrees)
13	555.7	543.9	-11.8	0.824	0.364	2.469
50	878.6	883.9	5.3	0.593	1.292	3.598
119	812.1	817.3	5.3	0.876	0.468	1.170
261	524.5	529.9	5.4	0.830	1.491	4.439
450	2456.4	2462.0	5.6	0.878	1.198	1.978
500	247.6	245.6	-1.9	0.835	0.500	1.142
718	751.7	735.5	-16.1	0.582	0.684	1.300
1394	2301.0	2191.2	-109.8	0.720	4.214	1.769
1420	1626.5	1597.8	-28.7	0.893	1.058	1.515

In summary, shape similarity and positional accuracy metrics are proposed to measure the quality each extracted buildings from aerial LiDAR data. In reality, accuracy of input of the LiDAR data (data collection and missing data) and of building outlines has some effect on the quality of the extracted building. However, with a special flight plan (low flight attitude and multiple overlap flight scan) designed to minimize the self-shadowing and street shadowing made insufficient sampling data only a minor issue. These metrics mostly measure the accuracy of the building outlines generated from the submitted method, which may be over- or under-extraction of building data points. Notably, participants were only asked to submit the data points of individual building (because of time constraints), and the building outline was generated by the contest organizers. Therefore, the total errors derived from these metrics involve errors from the submitted and the organizers' methods. However, the evaluation process does reflect the quality of the submitted method, as the organizers' method has been documented elsewhere to be able to reconstruct high accuracy of building outlines [23].

6. Conclusions

Over the last decade, automatic approaches have been proposed to reconstruct and benchmark 3D building models. However, evaluations were predominantly on extremely low-density data sets with only minimal points on building facades, which no longer fully reflects the state of data availability. This paper presents the objectives of the track "Urban 3D Model Generation" in the IQmulus contest 2014 related to automatic building detection

and reconstruction. This 2007 test dataset consisted of ALS data captured over 1km^2 of the Dublin's city centre with data density up to 225 points/m^2 and was designed to maximize vertical surface data capture.

An evaluation strategy was proposed to benchmark the results in terms of the capacity of the submitted method in detecting and reconstructing building outlines/models and quality of the models with respect to geometrical accuracy. While that approach detected most buildings in the study area, correctness and quality quantities were only around 65%. The average area difference between the detected building outline and GT was 3.59m^2 (std=66.75m^2), while the matching overlap between their perimeters exceeded 76%. In terms of the positional accuracy, the off-set distance averaged 1.10m with an orientation error of 3.09 degrees. These results show that even with an extremely dense LiDAR dataset, building detection in dense urban areas is still a major challenge, especially where the buildings have similar geometries and are closely abutted. Notably, the test datasets remain available, and further submissions are welcome for evaluation at www.sites.google.com/site/iqmuluscontest2014/.

Acknowledgments

Data acquisition was generously supported by Science Foundation Ireland Grant 05/PICA/I830 and Ireland's Environmental Protection Agency Grant 2005-CD-U1-M1. This work was funded by the European Union's ERC-2012-StG_20111012 Project 307836. The authors gratefully acknowledge this support.

References

[1] Brenner C. Building Extraction. In: Vosselman G, and Maas, H-G., editor. Airborne and Terrestrial laser scanning. Scotland, UK: Whittles Publishing; 2013.

[2] Perera GSN, Maas H-G. Cycle graph analysis for 3D roof structure modelling: Concepts and performance. ISPRS Journal of Photogrammetry and Remote Sensing. 2014;93:213-26.

[3] Xiong B, S. Oude Elberink, Vosselman G. A graph edit dictionary for correcting errors in roof topology graphs reconstructed from point clouds. ISPRS Journal of Photogrammetry and Remote Sensing. 2014;93:227-42.

[4] Laycock RG, Day AM. Automatically generating large urban environments based on the footprint data of buildings. Proceedings of the eighth ACM symposium on Solid modeling and applications. Seattle, Washington, USA: ACM; 2003. p. 346-51.

[5] Zhou Q-Y, Neumann U. 2.5D Dual Contouring: A Robust Approach to Creating Building Models from Aerial LiDAR Point Clouds. In: Daniilidis K, Maragos P, Paragios N, editors. Computer Vision – ECCV 2010: Springer Berlin Heidelberg; 2010. p. 115-28.

[6] Kaartinen H, Hyyppä J, Gülch E, Vosselman G, Hyyppä H, Matikainen L, et al. Accuracy of 3D city models: EuroSDR comparison. the ISPRS Workshop Laser scanning: ISPRS; 2005. p. 227-32.

[7] Rottensteiner F, Sohn G, Gerke M, Wegner JD, Breitkopf U, Jung J. Results of the ISPRS benchmark on urban object detection and 3D building reconstruction. ISPRS Journal of Photogrammetry and Remote Sensing. 2014;93:256-71.

[8] Gröger G, Kolbe TH, Nagel C, Hafele KH. OpenGIS City Geography Markup Language (CityGML) Encoding Standard (OGC 12-019). Version 2.0.0. OGC 12-019. Open Geospatial Consortium. Available at: : http://www.opengeospatial.org/standards/citygml. Accessed date: March 11, 2014

[9] Hinks T, Carr H, Laefer D. Flight Optimization Algorithms for Aerial LiDAR Capture for Urban Infrastructure Model Generation. Journal of Computing in Civil Engineering. 2009;23:330-9.

[10] Clarke J, Laefer DF. Generation of a Building Typology for Urban Tunnelling Risk Assessment. Bridge and Concrete Research in Ireland. Dublin, Ireland2012.

[11] Hinks T. Geometric Processing Techniques for Urban Aerial Laser Scan Data: University College Dublin; 2011.

[12] Laefer DL, O'Sullivan C, Carr H, Truong-Hong L. Aerial laser scanning (ALS) data collected over an area of around 1 square km in Dublin city in 2007. UCD Library, University College Dublin. 2014. Available at: http://dx.doi.org/10.7925/DRS1.UCDLIB_30462

[13] Arachchige NH, Maas HG. Automatic Building Facade Detection in Mobile Laser Scanner point Clouds. 2012.

[14] Pfeifer N, Rutzinger M, Rottensteiner F, Muecke W, Hollaus M. Extraction of building footprints from airborne laser scanning: Comparison and validation techniques. Urban Remote Sensing Joint Event, 20072007. p. 1-9.

[15] Awrangjeb M, Fraser C. An Automatic and Threshold-Free Performance Evaluation System for Building Extraction Techniques From Airborne LIDAR Data. Selected Topics in Applied Earth Observations and Remote Sensing, IEEE Journal of. 2014;PP:1-15.

[16] Chuiqing Z, Jinfei W, Lehrbass B. An Evaluation System for Building Footprint Extraction From Remotely Sensed Data. Selected Topics in Applied Earth Observations and Remote Sensing, IEEE Journal of. 2013;6:1640-52.
[17] Rutzinger M, Rottensteiner F, Pfeifer N. A Comparison of Evaluation Techniques for Building Extraction From Airborne Laser Scanning. Selected Topics in Applied Earth Observations and Remote Sensing, IEEE Journal of. 2009;2:11-20.
[18] Shufelt JA. Performance evaluation and analysis of monocular building extraction from aerial imagery. Pattern Analysis and Machine Intelligence, IEEE Transactions on. 1999;21:311-26.
[19] Okorn B, Xiong X, Akinci B, Huber D. Toward automated modeling of floor plans. Symposium on 3D Data Processing, Visualization and Transmission. Paris, France 2010. p. 8.
[20] Dublin City Council. Planning Application. Available at: http://www.dublincity.ie/Planning/PlanningApplication/Pages/Planning%20Application.aspx. Accessed date: Dec 20, 2010

[21] Besl PJ, McKay ND. A method for registration of 3-D shapes. Pattern Analysis and Machine Intelligence, IEEE Transactions on. 1992;14:239-56.
[22] Truong-Hong L, Laefer DF. Validating Computational Models from Laser Scanning Data for Historic Facades. Journal of Testing and Evaluation. 2013;41:16.
[23] Truong-Hong L, Laefer DF, Hinks T, Carr H. Combining an angle criterion with voxelization and the flying voxel method in reconstructing building models from LiDAR data. Computer-Aided Civil and Infrastructure Engineering. 2012;28:112-29.
[24] Bendels GH, Schnabel R, Klein R. Detecting Holes in point set surfaces. Journal of WSCG. 2006;14.
[25] Becker S, Haala N. Refinement of Building Facades by Integrated Processing of LIDAR and Image Data. PIA07 - Photogrammetric Image Analysis. 2007;Munich, Germany, 19-21 September, 2007:36(3/W49A), 7-12.
[26] Truong-Hong L, Laefer DF, Hinks T, Carr H. Flying Voxel Method with Delaunay Triangulation Criterion for Façade/Feature Detection for Computation. ASCE Journal of Computing in Civil Engineering. 2012;26:691–707.
[27] Pu S, Vosselman G. Extracting windows from terrestrial laser scanning. ISPRS Workshop on Laser Scanning and SilviLaser 2007. Espoo, Finland, September 12-14, 20072007. p. 320-5.

[28] Rabbani T, Heuvel Fvd, Vosselmann G. Segmentation of point clouds using smoothness constraint. International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences. 2006;36:248-53.
[29] Richter R, Behrens M, Döllner J. Object class segmentation of massive 3D point clouds of urban areas using point cloud topology. International Journal of Remote Sensing. 2013;34:8408-24.
[30] Awrangjeb M, Lu G, Fraser C. Automatic Building Extraction From LIDAR Data Covering Complex Urban Scenes. ISPRS Technical Commission III Symposium. Zurich, Switserland2014. p. 25-32.