Available online at www.sciencedirect.com
ScienceDirect
ELSEVIER
ISPRS Journal of Photogrammetry & Remote Sensing 62 (2007) 43–63
PHOTOGRAMMETRY & REMOTE SENSING
www.elsevier.com/locate/isprsjprs

Data fusion of high-resolution satellite imagery and LiDAR data for automatic building extraction ☆

Gunho Sohn *, Ian Dowman

Department of Geomatic Engineering, University College London, Gower Street, London WC1E 6BT, United Kingdom

Received 7 December 2005; received in revised form 17 December 2006; accepted 3 January 2007
Available online 15 February 2007

Abstract

This paper aims to present a new approach for automatic extraction of building footprints in a combination of the IKONOS imagery with pan-sharpened multi-spectral bands and the low-sampled (~0.1 points/m^2) airborne laser scanning data acquired from the Optech's 1020 ALTM (Airborne Laser Terrain Mapper). Initially, a laser point cluster in 3D object space was recognized as an isolated building object if all the member points were similarly attributed as building points by investigating the height property of laser points and the normalized difference vegetation indices (NDVI) driven from IKONOS imagery. As modelling cues, rectilinear lines around building outlines collected by either data-driven or model-driven manner were integrated in order to compensate the weakness of both methods. Finally, a full description of building outlines was accomplished by merging convex polygons, which were obtained as a building region was hierarchically divided by the extracted lines using the Binary Space Partitioning (BSP) tree. The system performance was evaluated by objective evaluation metrics in comparison to the Ordnance Survey's MasterMap®. This evaluation showed the delineation performance of up to 0.11 (the branching factor) and the detection percentage of 90.1% (the correctness) and the overall quality of 80.5%.

Keywords: Building extraction; LiDAR; IKONOS; Fusion; Binary space partitioning

1. Introduction

One of the ultimate aims in photogrammetry is to generate an urban landscape model (ULM) to show the objects and landcover of an urban area in three dimensions (Dowman, 2000). This task requires *object extraction*, which detects the object of interest and extracts its geometric boundary from remote sensed data. A large number of natural and cultural objects could be involved in constructing the ULM. As the most prominent feature in an urban environment, acquiring accurate and up-to-date extraction of building objects has significance for urban planning, cartographic

☆ Significantly enhanced version of a paper published in the proceedings of International Archive of Photogrammetry and Remote Sensing: Sohn, G., 2004. Extraction of buildings from high-resolution data and airborne LiDAR. 35 (B3), pp. 1036–1042.

* Corresponding author. Present address: Department of Earth and Space Science and Engineering, York University, 4000 Keele St., Toronto, Canada, ON M3J 1P3. Tel.: +1 416 650 8011; fax: +1 416 736 5817.
E-mail addresses: g.sohn@ge.ucl.ac.uk, gsohn@yorku.ca (G. Sohn), i.dowman@ge.ucl.ac.uk (I. Dowman).

0924-2716/$ - see front matter © 2007 International Society for Photogrammetry and Remote Sensing, Inc. (ISPRS). Published by Elsevier B.V. All rights reserved.
doi:10.1016/j.isprsjprs.2007.01.001

mapping, civilian and military emergency response. In tradition, the building extraction has mainly relied on manual photo-interpretation with the support of DPWs (Digital Photogrammetric Workstations), which still remains in an expensive process, especially when a large amount of data must be processed. Thus, the full automation of extracting buildings has been regarded as an active research topic in photogrammetry and computer vision community (Ameri, 2000).

1.1. Background

In general, automatic building extraction can be achieved through certain hierarchical processing stages. The first stage starts with extracting cues (e.g., points, lines and regions), which are locally distinguishable geometric, or chromatic properties extracted by low-level feature extraction algorithms from view-centered data (building cue extraction). Then, those building cues that are related only to a building are found (building detection), and the final stage ends up with reconstructing the building boundaries in an object-centered model (building description). All the tasks mentioned above are equally important for the development of an ideal building extraction system. A lot of research efforts have been made towards solving this problem. However, the success of automatic building extraction is still a distant goal. Rather than directly extracting building outlines, some researchers have used existing 2D ground building plan, thus reducing the difficulties of building outlines (Brenner, 2000; Suveg and Vosselman, 2004). There are several reasons to explain this difficulty which are discussed in the following sections.

1.1.1. Scene complexity

Most of the scenes usually contain very rich information which provides a large amount of cues with geometric or chromatic co-similarity to buildings, but belong to non-building objects. This makes the building detection task difficult. In earlier studies, some heuristic knowledge was used for solving this problem. For instance, it is assumed that building cues can be found in an area with certain brightness (Park et al., 2000) or specific colours (Moons et al., 1998) or pre-defined type of line grouping structure (Jaynes et al., 1994) or nearby shadow (Shufelt and McKeown, 1993). However, there are many situations in complex urban areas where those pre-specified heuristics are not valid. Since the building object is generally located above the terrain with certain height, 3D information extracted from stereo analysis plays an important role to eliminating erroneous cues. In this context, the building cues such as 3D corner (Fischer et al., 1998), 3D lines (Baillard and Zisserman, 2000) and 3D planar polygons (Ameri, 2000) have been investigated for the purpose of building extraction. In an alternative way, the complexity of the building detection can be greatly reduced by subtracting terrain feature from a complex scene as background information (Baillard and Maitre, 1999; Weidner and Forstner, 1995). In the light of this, various techniques for extracting terrain surfaces from optical imagery, IfSAR and LiDAR DSM have been recently reported (Dowman, 2004). There is, however, no unified filtering algorithm, which is robust to all different landforms (Sithole and Vosselman, 2004).

1.1.2. Incomplete cue extraction

There is always a significant loss of relevant building cues due to occlusion, poor contrast, shadows and disadvantageous image perspective. Under these circumstances, the cues may be highly fragmented or completely missed, which lead to wrong building description. There may be two ways to recover those incomplete cues in order to fully describe the building boundary; model-driven and data-driven method. The former approach assumes that most of building outlines can be reconstructed by fitting a specific building model to the cues extracted and verifying them with certain evidences (Lin and Nevatia, 1998). The benefit of this method is to reconstruct the building boundary with minimal use of cues, which is a useful property, especially in the case that sufficient cue extraction is difficult to obtain with given data sources. However, the method requires additional information for verifying models. It is not always easy to collect all the verification cues due to the aforementioned scene complexity. The latter approach performs the building description process by minimal use of a prior knowledge of specific building structure and aims to delineate polyhedral building shapes. This method requires very dense collection of building cues from multiple data frames (Baillard and Zisserman, 2000), or from a high-quality of 3D measurement such as LiDAR data (Vosselman, 1999). The method emphasizes the representation of wide variations in building shapes, but may fail where all significant features of a building object cannot be extracted by low-level vision algorithms. As discussed above, both methods have their own pros and cons. It is a challenging research topic to combine two different approaches in an optimal way so that their weakness can be compensated by each other (Brenner, 2004).

1.1.3. Sensor dependency

The building extraction system is highly influenced by the sensor used. The reason for this is that the primary data used limits knowledge which can be available for

extracting buildings. For instance, a building extraction system using 3D depth information is very different from a system that relies only on monocular imagery. The primary data to support the building extraction is available from a variety of sources with different resolution (e.g., optical airborne, spaceborne, LiDAR, microwave imaging sensors, and cartographic map). Each data source has its own strength and weakness as the primary cue for the building extraction. An important research issue is to fuse different data sources so that the weaknesses of individual cues are compensated for by others (Dowman, 2004). However, it is not a trivial task to determine how to optimally select primary cues from multiple sensors in the fusion process and how to combine them at each process of building detection and description.

1.2. Data fusion

Recent developments of modern technologies diversify the methods for acquiring the primary data in building extraction. Such new comers are IKONOS satellite and LiDAR. Over the last five years, both sensors have become powerful tools to replace or complement conventional data acquisition techniques. The rapid emergence of IKONOS satellite and LiDAR has created an urgent need for new techniques for building extraction. Much research efforts have been made to automatically detect and describe buildings from either IKONOS (Haverkamp, 2004, Lee et al., 2003; Sohn and Dowman, 2001) or LiDAR (Vosselman, 1999; Weidner and Forstner, 1995; Maas and Vosselman, 1999). Further investigation is, however, needed to develop a fully operational building extraction system. This is caused by the weakness inherited from each sensor. IKONOS imagery is still many times poorer than aerial images in spatial resolving power and often shows weak reflectance around building boundary. A consequence is a significant loss of geometric cues which leads to adverse effects in building description. On the other hand, the under-sampling nature of LiDAR data acquisition, due to small footprint size of laser beam compared to average point spacing and disadvantageous backscattering from illuminated targets, makes it difficult to extract building edges with height discontinuity.

However, an efficient procedure of data fusion can compensate for the weakness of both sensors for building extraction (Schenk and Csatho, 2002). It is because they naturally have the complementary properties of each other; the IKONOS imagery is easier to extract building boundaries than in LiDAR and can provide a variety of information such as intensity, colours, and textures, while LiDAR can reliably attribute accurate 3D information to those cues extracted from optical imagery. There have been a few recent studies to extract buildings in a combination of IKONOS imagery and LiDAR data. Complementary fusion principles for extracting building outlines were driven by integrating LiDAR data with multi-spectral IKONOS imagery (Kim and Muller, 2002). Although a satisfactory result was reported by the authors, exploitation of the synergy of both sensors mainly focused on detecting buildings, and sophisticated technique for accurate building description was not considered.

Rottensteiner et al. (2005) employed the Dempster–Shafer theory as a data fusion framework for building detection from airborne orthophoto with multi-spectral bands and LiDAR data. The building object is detected on a per-pixel classification by probabilistically investigating various properties (e.g., colour, height variance and surface roughness) of each pixel, which is later validated in a region-wise manner. Instead of chromatic information, Guo and Yasuoka (2002) combined monochromatic IKONOS imagery with LiDAR data for extracting buildings. After isolating building regions by subtracting terrain surface from LiDAR data, an active contour model was fitted to building outlines by the guidance of edges extracted from both LiDAR data and IKONOS imagery. However, reconstructed outlines still remain to be un-regularized. Chen et al. (2004) detected step edges from LiDAR data, which are geometrically improved with the support of linear feature extracted from airborne imagery with 10-centimetres ground sampling distance. Those refined lines are grouped in a closed polygon by a splitting-and-merging algorithm. Although the proposed algorithm has an advantageous aspect in implicitly constructing topological relations across features extracted, the method did not show how to cope with erroneous lines for achieving an objective optimality. This problem may become significant, especially when IKONOS imagery or poor quality of LiDAR data is used as primary data source. This paper presents a new technique of automatic building extraction from multi-spectral IKONOS imagery in a combination of LiDAR with low point density. A particular research interest will be given to reconstructing building outlines with global optimization criteria, especially when sufficient rectilinear lines are not able to detect from primary data sources.

2. Method overview

Fig. 1 illustrates a workflow of the major components and their subsequent processes in the automated building extraction system proposed in this paper. The system requires IKONOS imagery with multi-spectral

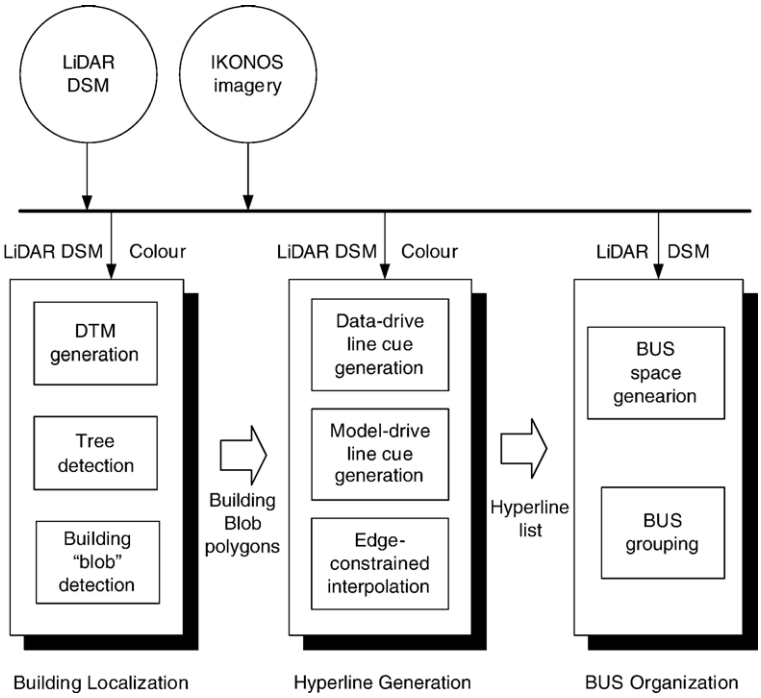

Fig. 1. Schematic workflow of the overall process of the proposed building extraction system.

bands, which is re-sampled at 1-metre resolution on the ground, and LiDAR data with horizontal point spacing of about 3 m. It is pre-assumed that both data are geo-referenced. The suggested building extraction system consists of two major processes; building detection and building description.

For building detection, the proposed system follows a focusing strategy suggested by Baillard and Maitre (1999), in which feature classes are hierarchically categorized and a classification algorithm is selected according to the feature class targeted. This process allows the subsequent building description process to focus only on one building structure. First, on-terrain features and off-terrain features are differentiated according to whether those features are located either on the terrain surface or above the terrain surface. This classification is done as terrain surface is reconstructed by a newly developed algorithm, called RTF (Recursive Terrain Fragmentation) filter. Once the terrain information is extracted, outlying points with a height of more than a pre-defined threshold from the generated terrain surface are further classified as high-rise features. At the finest level, vegetated points are excluded from the high-rise features by computing the normalized difference vegetation indices (NDVI) from IKONOS multi-spectral bands. As a result, LiDAR points which only belong to building structures can be obtained. Further processing allows the individual objects to be bounded with rectangles, which are then fed into the building description process with the final labels of LiDAR points.

The next step is to reconstruct building outlines based on the result of the building detection process. This is achieved by a BUS (Building Unit Shape) organization method. This description process aims to delineate polyhedral shape of buildings. To this end, generic building shape is represented as a mosaic of convex polygons. Each convex polygon is referred as the BUS. A rectangle resulting from the building detection is used as an initial BUS. Linear features around building boundary are extracted from IKONOS image with the support of the classification result of LiDAR points. In order to compensate completely or partly missed data-driven lines, model-driven lines are additionally produced by fitting specific building models to LiDAR points. Both data-driven and model-driven lines are employed as hyperlines to halve the initial BUS, which leads to two sub-convex polygons. This binary partitioning process continues until no child polygon is generated by the hyperlines, and ends up with a set of polygons (i.e., BUSes). This recursive generation of convex polygons is implemented by a Binary Space Partitioning (BSP) algorithm, in which a binary tree is incrementally expanded as child polygons are generated by the

hyperlines. The final result of the BSP tree is a BUS adjacent graph where each node represents a BUS and each arc means the connectivity between neighbouring BUSes. A BUS grouping process merges only BUSes, which belong to building structures, and eliminates spurious lines besides building boundaries. As a result, building outlines are reconstructed.

3. Test dataset

For the current research a LiDAR DSM was provided by Infoterra Co., which covers a sub-site of Greenwich industrial area with the size of 1.2×1.2 km^2 ($\sim$1,180,073.6 m^2). The LiDAR DSM was acquired by the first pulse of Optech's 1020 airborne laser sensor. The data has been converted from OSGB36 (plan) and OSD Newlyn (height) to UTM/WGS84. The LiDAR DSM contains a total of 113,541 points, which approximately corresponds to a point density of 0.1 (points/m^2), i.e., one point per 3.2×3.2 (m^2). The height of the study area varies from 0.21 m to 39 m. As can be seen in Fig. 2(a), the Greenwich LiDAR DSM shows a typical urban environment. The height of the terrain in this scene ranges from 0.21 m–14 m, and the highest terrain height can be found in the South–West corner in Fig. 2(a). A number of industrial buildings with different sizes spread over the study area. The height of large buildings ranges from 7 m–39 m. Some low objects with height of 4 m–8 m can be found in the Greenwich dataset. In addition, a group of tall trees with various heights are located on the street. A row of terraced houses can be found at the lower left corner of the scene. Fig. 2(a) also shows very irregular surface over some rooftops, though they are formed in planar roof surfaces. This problem is highlighted when the Greenwich LiDAR measurements are overlaid with the IKONOS image in Fig. 3(a). Although the point density of the Greenwich data shows 0.1 points/m^2, it is much less over some building rooftops and the ground near buildings. This may be caused by certain factors such as poor weather condition, high absorption by specific material property, occlusion effect, laser scanning pattern and flying height restriction over London, which limit returned laser pulses from dark surfaces, sloped and shinny surfaces. This problem may affect the building extraction process, which will be discussed later.

Over the Greenwich study area, a pan-sharpened multi-spectral (PSM) IKONOS image is also provided by Infoterra Co. The PSM image is produced by combing the multi-spectral data with the panchromatic data, and re-sampled with 1-metre ground pixel. This IKONOS PSM image was orthorectified by Space Imaging Co. to satisfy the positional accuracy ($\sim$1.9 m) of Precision Product of Space Imaging. To meet these high-quality accuracy levels, Space Imaging uses high-precision ground control and precise terrain models to create Precision products. Fig. 2(b) shows the Greenwich IKONOS PSM image, in which the red channel is replaced with the near-infrared channel while the green channel with the red one respectively.

4. Building detection

This section presents a building detection method to localize individual buildings by sequentially removing dominant urban features which are not relevant to buildings. The success of building detection is important to make the remaining building description process focused to building objects individually.

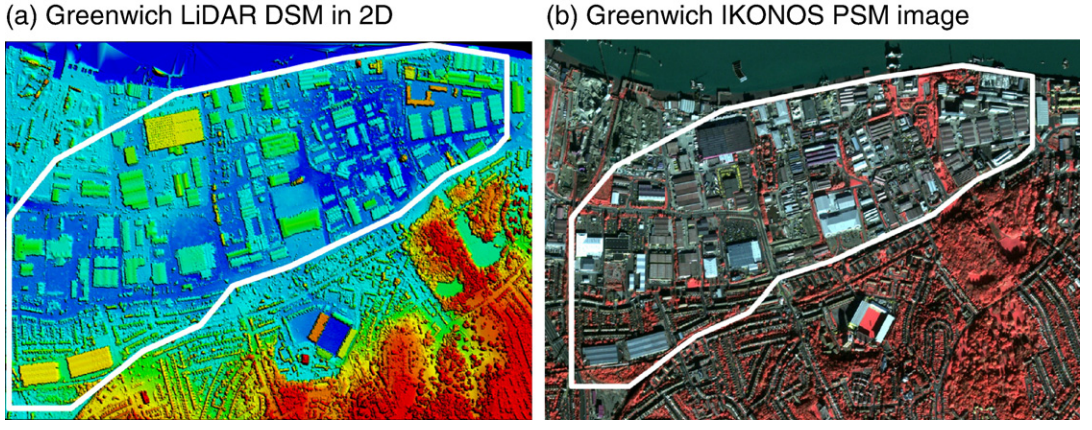

Fig. 2. Greenwich test datasets.

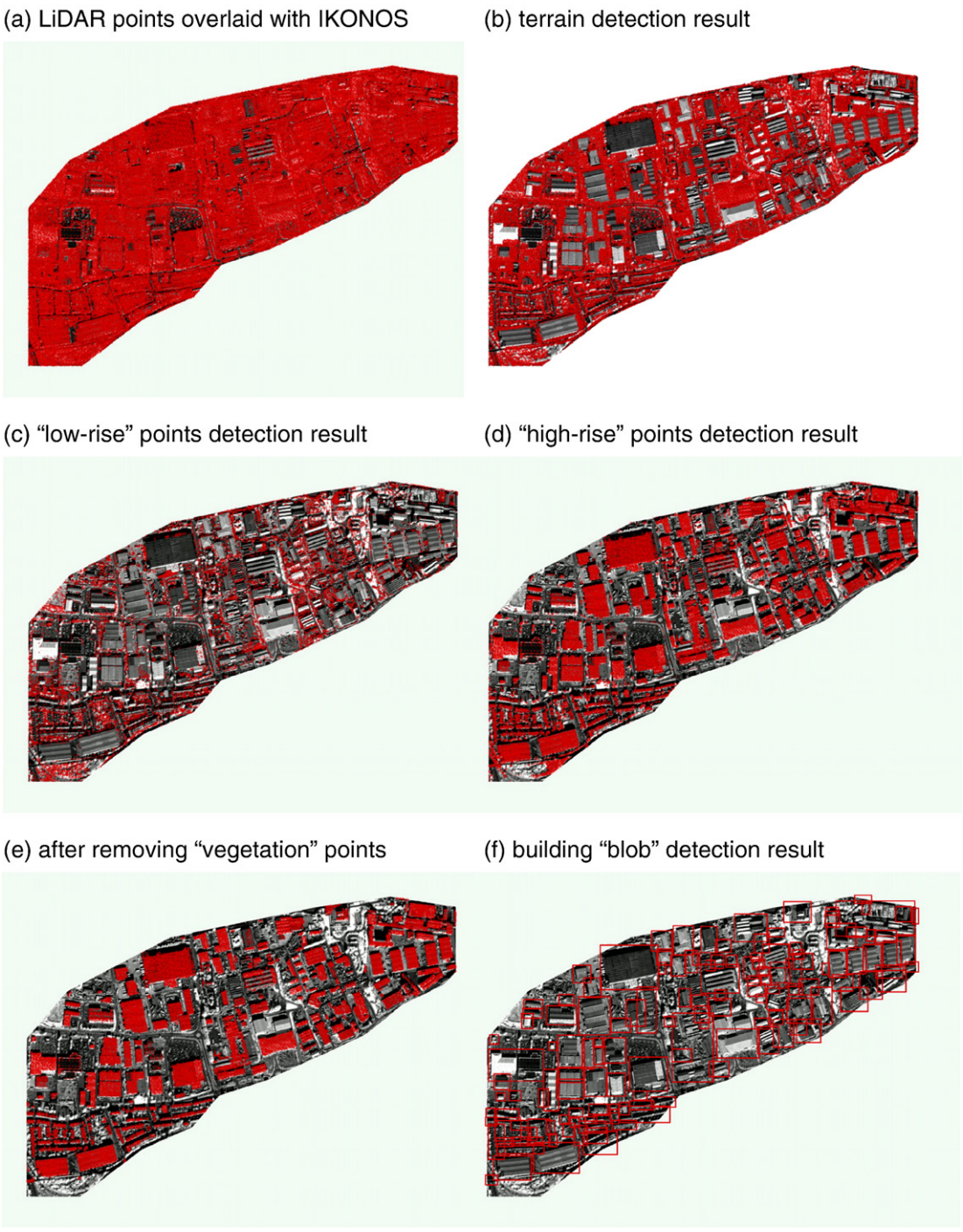

Fig. 3. Sequential results of building detection process; white cross means LiDAR points; white rectangles in (f) mean building blob detection results.

4.1. Coarsest feature classification

The first step of building detection process is to differentiate on-terrain points from off-terrain points in the Greenwich LiDAR data. To this end, a LiDAR filter, called recursive terrain fragmentation (RTF) filter, was developed to classify on-terrain points from a cloud of LiDAR points. This filter assumes that generic terrain surface is a set of piecewise planar surfaces, and a plane terrain model (PTM) is defined as an elementary model

for reconstructing a mosaic of planar terrain surfaces from the LiDAR measurements. A final terrain surface is reconstructed in an iterative hypothesis–test scheme. The entire LiDAR space is initially hypothesized by a single PTM. If the hypothesized area does not satisfy the conditions of the PTM, the initial terrain model will be fragmented into smaller sub-regions by obtaining a terrain point from underlying LiDAR points. Since a number of LiDAR points could be considered as the terrain point, an optimum solution of the terrain fragmentation is achieved by MDL (Minimum Description Length) principle. This fragmentation process continues until all the fragmented regions are verified as PTM. At the end of the process, every LiDAR point is labelled as either terrain or non-terrain point. Since the development of RTF filter is beyond the scope of this paper, details of this technique will not be described in the following sections, but can be found in Sohn and Dowman (2002) and Sithole and Vosselman (2004). Fig. 3(b) shows the on-terrain points detected by the RTF filter from Fig. 2(a). In this figure, some terrain segments which are not densely covered by the filtered on-terrain points show poor quality of the Greenwich LiDAR DSM. Fig. 4 shows the terrain reconstructed by the RTF filter from the Greenwich LiDAR DSM. As shown in the figure, the DTM successfully removed non-terrain objects by the RTF filter despite the fact that underlying terrain surface consists of different terrain slopes and building objects with a range of different sizes are located.

4.2. Intermediate feature classification

Once either an on-terrain or off-terrain label is assigned to every LiDAR point, off-terrain label points are re-classified into more detailed feature classes, i.e., low-rise or high-rise feature class. The determination of this intermediate level of feature classification is made by comparing height attributes of off-terrain points with pre-specified height threshold. Fig. 5(a) shows a height histogram of the result of the coarsest feature classification. As can be seen in this figure, the height histogram of on-terrain points obviously shows single dominant peak. It suggests that the underlying terrain surfaces of the Greenwich dataset is successfully detected by the RTF filter. It is, however, difficult to distinguish between two classes (i.e., low-rise and high-rise classes) from the height histogram of off-terrain points since their heights are corrupted by the terrain surface. This problem can be solved by subtracting the height of underlying terrain surface from the one of off-terrain points. As shown in Fig. 5(b), it becomes easier to distinguish between the low-rise and the high-rise class when the heights of off-terrain points are normalized by the terrain surface. Although a height threshold to classify off-terrain points into either low-rise or high-rise feature class is rather arbitrarily selected as 4 m in this study, Fig. 5(b) shows the selection of this height criterion could be automated by a statistical analysis of the normalized height histogram. Fig. 3(c) and (d) shows the final results of the intermediate level of feature classification, in which classified low-rise and high-rise points are overlaid with IKONOS imagery.

4.3. Finest feature classification

The high-rise feature class is generally a mixture of tree and building objects. For building detection, the trees must be excluded from the high-rise features,

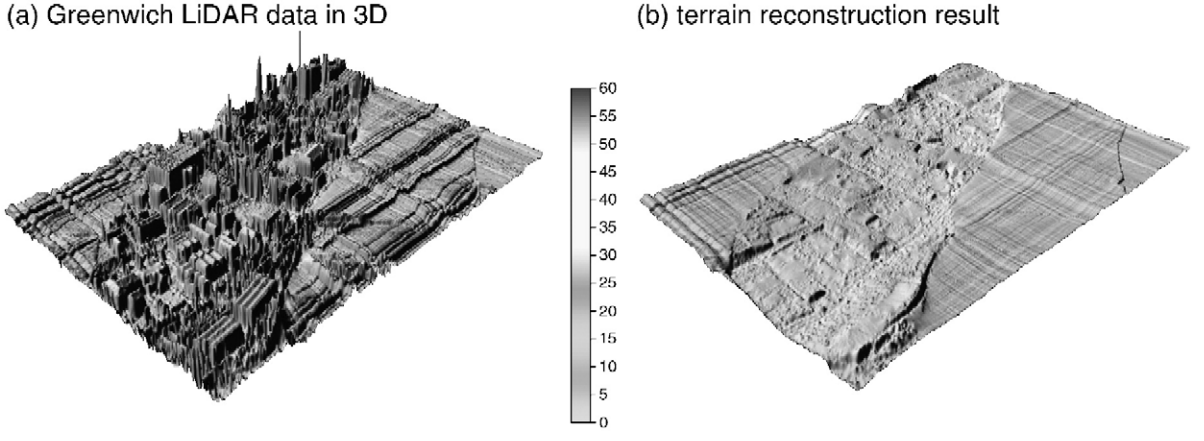

Fig. 4. Greenwich terrain surface reconstructed by RTF filter.

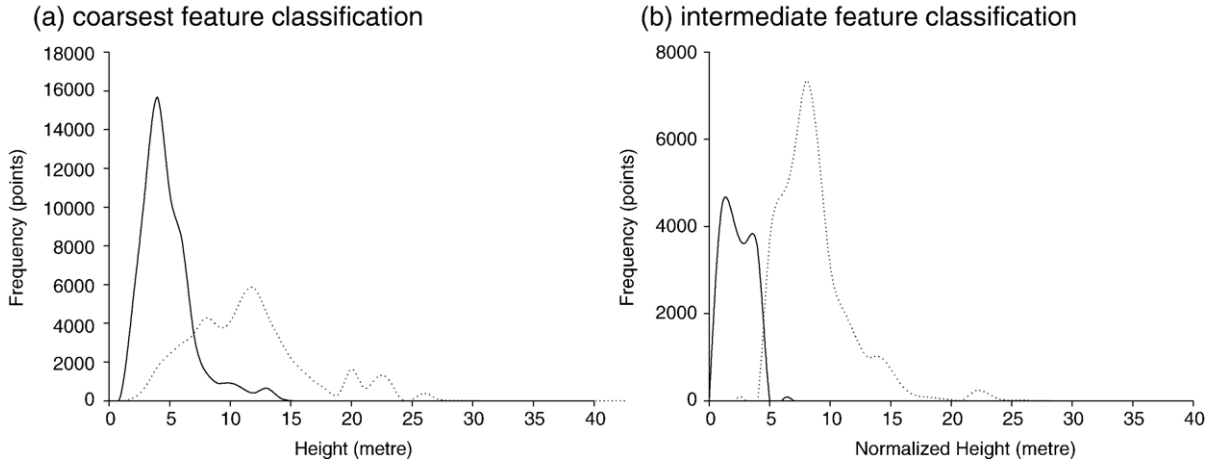

Fig. 5. Height histograms of coarsest and intermediate feature classification; (a) on-terrain feature (solid line) and off-terrain feature (dash line) and (b) low-rise feature (solid line) and high-rise feature (dash line).

by which the isolation of individual buildings can be achieved. The discrimination between buildings and trees is accomplished through the analysis of the normalized difference vegetation index (NDVI) of the IKONOS multi-spectral information. The NDVI is an established vegetation index that is widely used in the thematic mapping applications by looking at the combinatorial reflectance in the red and near-infrared bands and is defined as; $NDVI=(NIR-R)/(NIR+R)$, where NIR and R represent data from IKONOS red and near-infrared channel respectively. Since vegetated area generally generates relatively high reflectance by near-infrared channel than by visible red channel, higher NDVI values indicate larger amount of green vegetation present in the pixel. Once the NDVI map is calculated (Fig. 6(a)), the high-rise features are then, divided into building and tree class using a global threshold technique of the NDVI map. This method assumes that the NDVI distribution for the high-rise feature class will have two clear bimodal distributions, one for the building and one for the tree. By back-projecting the high-rise points into IKONOS image space and incorporating the NDVI information with them, the NDVI for each high-rise point is calculated from a small mask constructed around the back-projected point; the tree label is assigned to a high-rise point if any pixel of the masked points has the NDVI value larger than a pre-specified threshold, i.e., δ_{NDVI}, otherwise, the building label is assigned. Finally, vegetated high-rise points can be removed, while building label points are retained. Fig. 6(b) shows a binary image when the NDVI map is thresholded by δ_{NDVI} of 0.8. Fig. 3(e) displays building label points detected when the mask size of 5×5 was used over Fig. 6(b).

4.4. Building blob detection

Isolating the building label points and making them into individual building objects is rather straightforward. Once tree points are differentiated from building points by the NDVI method described in the previous section, those points classified into the on-terrain, low-rise, and tree objects are altogether assigned non-building labels. Then, building points surrounded by the non-building labels, are grouped as isolated objects (called building blobs) by a connectivity analysis in TIN (Triangulated Irregular Network), which was modified based on an algorithm suggested by Lumia et al. (1983). As a result, 170 building blobs can be found from Fig. 3(f) after removing small blobs whose member points is less than 30 points. Further processing allows the individual building blobs to be bounded with rectangle polygons, and these polygons are then fed into the building extraction process, which will be discussed in the next section.

5. Building description

This section describes a building description method, which extracts building outlines based on the result of the building detection process presented in the previous section. A BUS organization is introduced as a new middle-level of feature grouping method. IKONOS imagery and LiDAR points play complementary roles to extract lines around building boundary. Then, those lines extracted from both sensors recursively partition the rectangles bounding building blobs into convex polygons until certain termination criteria are met. Finally, building outlines are reconstructed by merging polygons which are verified as part of building structure.

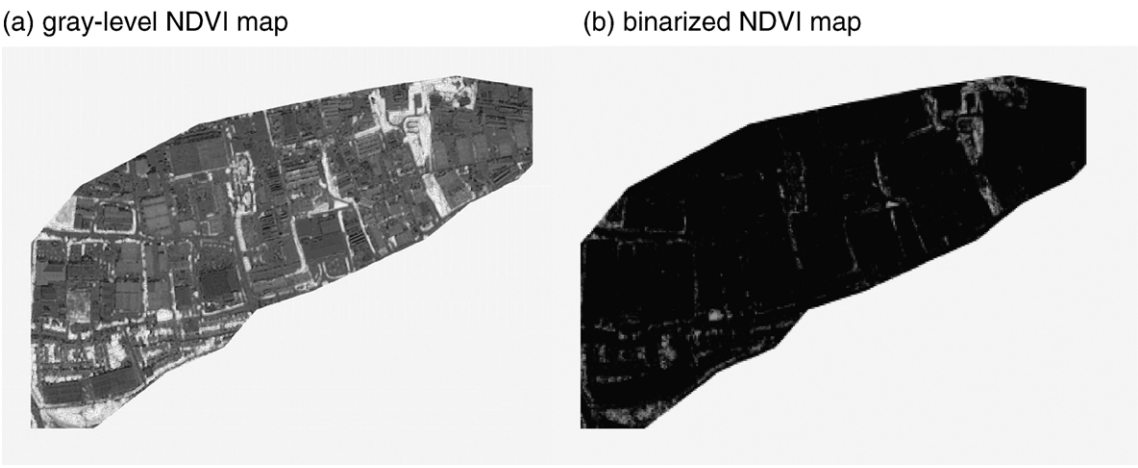

Fig. 6. Results of the finest feature classification.

5.1. Data-driven line cue generation

As a primary cue for building extraction, straight lines are extracted from IKONOS imagery by the Burns algorithm (Burns et al., 1986). However, since extracted line features include a number of extraneous line segments, uncorrelated to building saliencies, it is necessary to filter those distracting features so that only focused lines with significant length located around building boundaries remain. To this end, non-building (including on-terrain, low-rise, and tree) and building points labelled by the previous classification process are used to determine whether or not line primitives can be considered as boundary lines.

First, straight lines extracted by the Burns algorithm are filtered by a length criterion, by which only lines larger than pre-specified length threshold, l_d, remain for further processing. Then, two rectangle boxes with certain width, l_w, are generated along two orthogonal directions to the line vector filtered in length. The determination of boundary line can be given if non-building and building points are simultaneously found in both boxes or if only building-label points are found in one of the boxes and no LiDAR point can be found in the other box. The latter boundary line condition is considered if a low density LiDAR dataset is used.

As a final line filtering process, a geometric disturbance corrupted by noise is regularized over boundary lines. A set of dominant line angles of boundary lines is analyzed from a gradient-weighted histogram which is quantized in 255 discrete angular units. In order to separate a weak, but significant peak from other nearby dominant angles, a hierarchical histogram-clustering method is applied. Once the dominant angle, θ_d, is obtained, lines with angle discrepancies which are less than certain angle thresholds, θ_{th}, from θ_d are found. Then, their line geometries are modified as their angles are replaced with θ_d. These modified lines do not contribute to the succeeding dominant angle analysis and the next dominant angle is obtained. In this way, a set of dominant angles is obtained, by which geometric properties of boundary lines can be regularized. Fig. 7 shows boundary lines extracted by the support of labelled LiDAR points, after being filtered in length (l_d=5 m) and geometrically regularized in angle threshold (θ_{th}=30°), where a box width of l_w is chosen as 5 m for boundary line verification. The result shows the boundary lines can be successfully extracted as the intensity line cues. Since the deficiency of LiDAR measurements on the ground nearby building boundaries in the Greenwich dataset, distracting lines located on the ground are also detected as intensity line cues. These lines will be, however, excluded during an upper level of cue generation, referred as polygon cue generation, which will be described in Section 5.3.

5.2. Model-driven line cue generation

Although a satisfactory result to extracting boundary lines could be obtained by the method presented in the previous section, the line cues produced in a data-driven way do not always provide sufficient cue density covering all the parts of building edges. This is because significant boundary lines may be missed due to low contrast, shadow overcast, and occlusion effects, especially when a 1-metre satellite imagery of a complex scene is solely used. However, LiDAR data is relatively less affected by those disadvantageous factors than the optical imagery. Thus, new lines are virtually extracted from LiDAR space in order to compensate for the lack of data-driven line

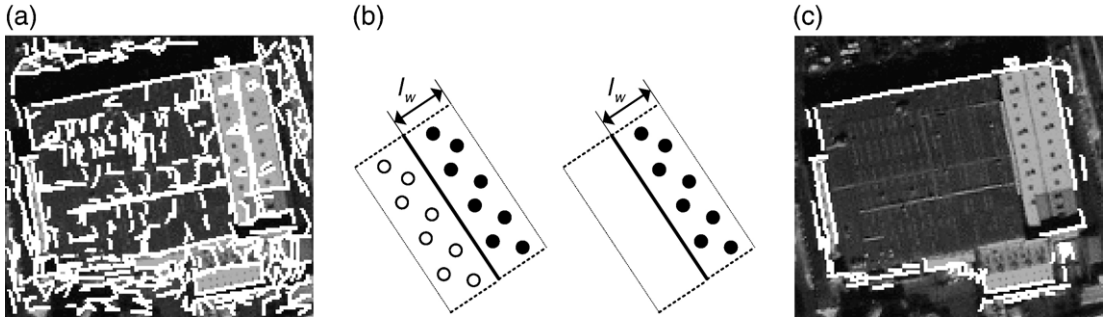

Fig. 7. Extraction of data-driven cue extraction and verification from IKONOS image; (a) straight lines extraction; (b) boundary line verification (white dots and black dots represent non-building and building points respectively; (c) filtered boundary line.

density by employing specific building models. It is assumed that building outlines are comprised of parallel lines. Based on this, for each data-driven line, parallel lines and "U" structured lines are inferred from LiDAR space. The process aims to acquire at most three model-driven lines starting from a data-driven line, but it could fail to generate any model-driven line. The main idea of the model-driven line detection is that a small virtual box is generated from each data-driven line and it grows over building roof so that building points are maximally captured without including any non-building point.

First, a box growing direction, pointing to the location of parallel boundary line, is determined. To this end, a small virtual box is generated with a width of l_w from the selected data-driven line in the same way of detecting boundary lines described in Section 5.1. To that direction, the virtual box grows until it comes across any on-terrain point (Fig. 8(a)). Then, it shrinks in order to have maximum building points while in its minimum size (Fig. 8(b)). In this way, the virtual box is expanded, but at this time, towards to two orthogonal directions parallel to the boundary line detected (Fig. 8(c)). Thus, "U" structured boundary lines made with the parallel boundary line can be detected. Finally, these three model-driven lines detected are back-projected onto image space and then, their line geometry is adjusted by gradient weighted least-square method.

When l_w is chosen as 5 m, a procedure of detecting three model-driven lines from one data-driven line can be illustrated in Fig. 9(a)–(c). As remaining data-driven line cues are involved, Fig. 9(d) shows the entire model-driven lines detected. Although the assumption of geometric regularity having parallel or "U" structured outlines is made for producing the model-drive lines, this assumption may not be valid in the case that a building shape has no symmetric property or erroneous lines are included as data-driven lines. For this reason, only a portion of model-driven lines could be involved in order to recover significant boundary segments missed in data-driven line generation. It is, however, subject to the degree of complexity of individual buildings and a false alarm rate of boundary line detection as to what percentage of model-driven lines cues can be used. Therefore, a verification process of model-driven lines is necessary. In the following section, it will be automatically determined at polygonal cue generation, whether or not a model-driven line should be used in boundary representation of building shape.

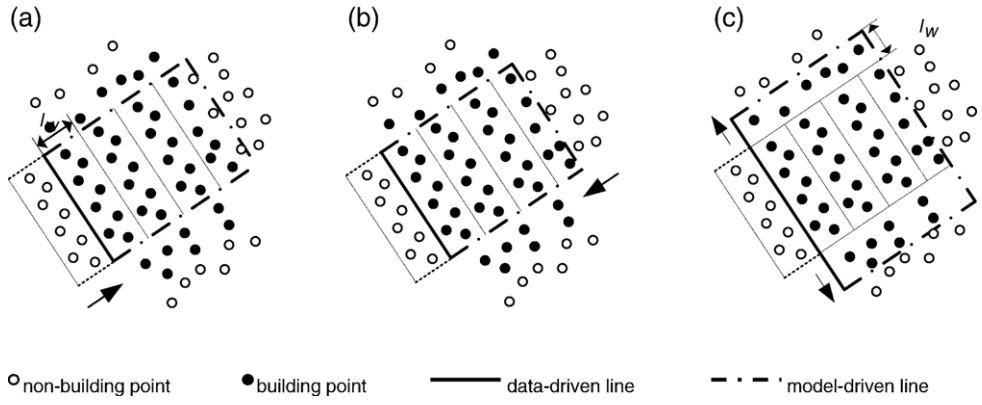

Fig. 8. Illustration of model-driven line cue generation.

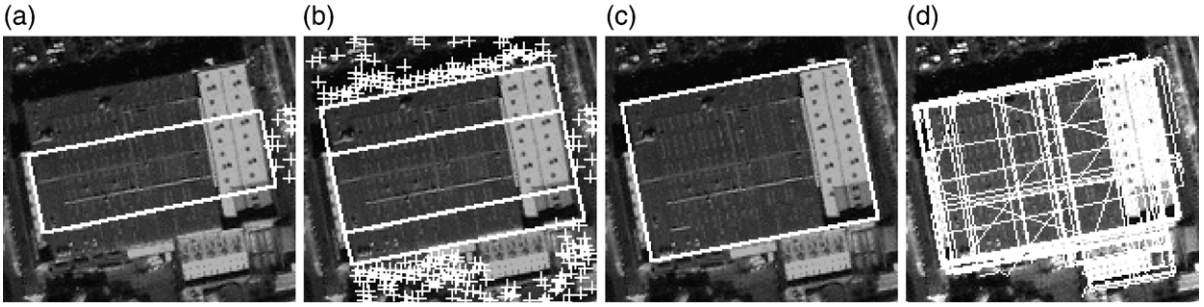

Fig. 9. Results of model-driven line cue generation; white lines represents detected model-drive lines, white arrow points to the virtual box growing direction and white cross represents non-building points; (a) parallel line detection; (b) "U" structure line detection; (c) gradient fitted virtual lines; (d) entire model-driven lines detected.

5.3. Edge-constrained data interpolation

Once the data-driven and model-driven lines are extracted from IKONOS imagery and LiDAR data, both lines are then used to refine a low and irregular density of LiDAR data. As discussed in Section 3, the Greenwich LiDAR data acquired by Optech's ALTM 1020 system provides approximately 0.1 points/m^2. The major problem of the Greenwich data is that the laser points were acquired with extremely irregular point density. This irregular property of the Greenwich LiDAR data can be found when the average distance of each point from its connected points in TIN is measured. It shows that all the laser points are distributed with a distance of 3.5 m in average with the standard deviation σ of 0.88 m. However, 9.6% of the total 113,684 points have larger than 4.38-metre point spacing (1σ level) and its maximum reaches up to 18 m (see Fig. 10(a)). In particular, those points with such extreme irregularity are mainly found over roads, parking lots and building rooftops with dark surfaces where the transmitted laser pulses were weakly reflected so that many laser points were missed.

As we will discuss in the following sections, the spatial distribution of laser points is the major factor which affects final reconstruction results of the suggested methods. Thus, such an irregularity of the laser point density will degrade the overall building description performance, leading to wrong building shapes. To prevent this problem, it is necessary to interpolate the Greenwich LiDAR data for having denser and more regular point density. The interpolation process in current experiment simply increases the numbers of LiDAR data by adding new laser points where triangular sizes in TIN constructed with LiDAR data are larger than certain threshold. However, the construction of TIN is constrained by both data-driven and model-driven lines so that the interpolation of laser points over the rooftops is limited to the inside of building outlines, while the terrain points increase outside buildings only.

All the building boundary lines are detected by extracting the data-driven and model-driven lines as presented in previous sections. Then, starting from the centre point of each line, virtual points are created with the equal distance of 2 m towards both starting and ending points of the lines. The original LiDAR points are integrated with newly generated virtual points in a TIN structure. If a triangle of the TIN is comprised of on-terrain points only or a mixture of on-terrain and virtual points, the triangle is classified as "on-triangle". In a similar way, the triangle is classified as "off-triangle" if its member points include off-terrain points only or a mixture of off-terrain and virtual points. For both "on-" and "off-triangle", if any lateral length of the triangle is larger than certain threshold (5 m), new laser point is calculated at the centre of the triangle and its height is determined as the average heights of member points. In this way, the original laser points increased to 229,866 points and the average point spacing became lowered to 2.5 ± 0.55 m (Fig. 10(b)). This figure suggests that the edge-constrained interpolation method successfully made the Greenwich dataset to be denser and more regular. However, 10% of the data still shows the irregularity of point spacing (larger than 3.5 m). This is because the extreme irregularity of the original dataset made failed to find boundary lines where laser points cannot be interpolated.

5.4. BUS generation

For building description, a major difficulty may lie at constructing topological relations between linear features. It can be accomplished by relying on hard-

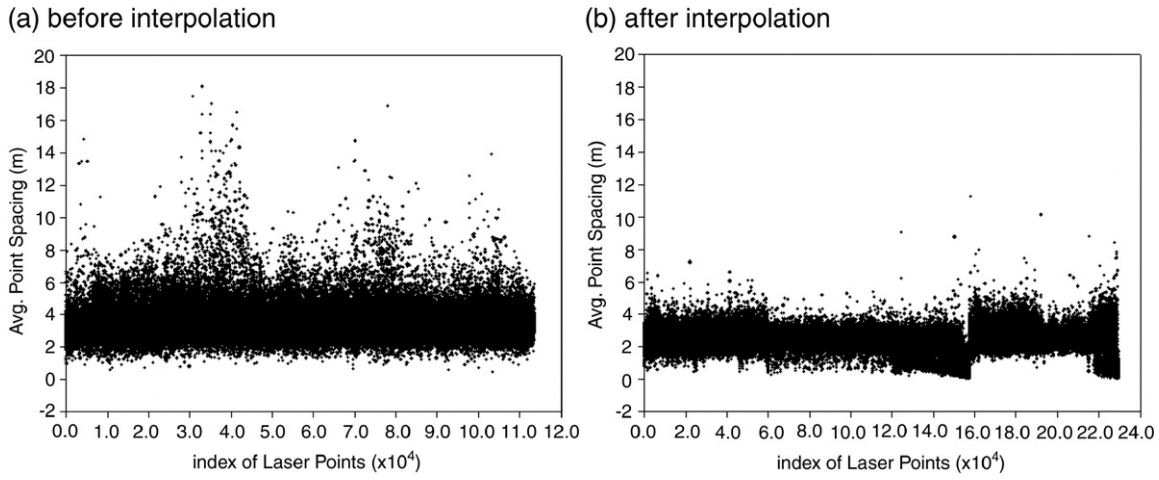

Fig. 10. Edge-constrained interpolation results.

constraints, which explicitly limits the number of neighbouring lines and grouping rules (Kim and Muller, 1999). However, how to determine those constraints are always subjective to various factors including sensor type used, image quality, scene and object complexity. A better idea is to implicitly construct the topological feature relations which can accommodate higher degrees of shape complexity.

5.4.1. BUS partitioning

There may be two ways to extract building outlines, either constructing topological relations from extracted lines (edge-based approach), or detecting outlines from segmented regions (area-based approach). Former method is better to describe a building shape with rectilinear lines, but is sensitive to the quality of lines extracted. Latter approach always provides full description of the scene, but produce building outlines, which are ragged with no regular geometry and do not coincide with real edges. A method suggested for feature grouping in this section is to combine advantages of both approaches, and segments whole scenes by lines extracted. By that, boundaries detected between neighbouring regions can coincide with real discontinuity, and topological relations between lines can be implicitly constructed as similar regions are merged.

The aforementioned hybrid building description method is accomplished by the BSP tree introduced by Fuchs et al. (1980). The Binary Space Partitioning (BSP) tree is a hierarchical subdivision of n-dimensional space into homogeneous regions, using $(n-1)$-dimensional hyperplanes. For 2D case, the subdivision is created using arbitrarily oriented lines, known as hyperlines. The BSP tree was initially developed for computer graphics applications to represent polyhedra in 3-dimensional space, and perform set operations upon them. Currently the BSP tree is widely used for applications, ranging from motion planning, shadow generation, image representation, and image compression (Radha et al., 1991). However, it is rare to find researches related to object description based on the BSP tree. Sohn and Dowman (2001) exploited the feasibility of BSP tree as a tool for building description, which aims to extract building outlines from monocular IKONOS imagery. In this study, the BSP algorithm is newly modified with a different strategy considering the contribution of LiDAR data.

In the framework of the BUS organization, a region surrounded by lines is referred as the BUS, and rectangles resulting from the building detection (Fig. 3(f)) are a special type of the BUS, representing initial shape of buildings. The BUS partitioning based on BSP tree is to decompose the initial BUS into a set of convex polygons by both of data-driven and model-driven lines. Fig. 11 illustrates the overall partitioning scheme to generate BUS space. Suppose that we have an initial BUS bounding a building object, P_0. A line list is prepared by integrating N lines of the data-driven and model-driven line cues, $\{l_i: i=1,\ldots, N\}$, (thick lines in Fig. 11). Also, N hyperlines, $\{h_i : i=1,\ldots, N\}$, computed as P_0 is intersected respectively by each line segment from the line list, and h_i is now defined by an expression of the form, $h_i = a + bx + cy$, where a, b, c are line parameters (thin lines in Fig. 11). When h_i is selected as a hyperline to subdivide P_0, two sub-convex polygons, P_{1+} and P_{1-}, are generated, depending on whether the dot product of each vertex point of P_0 with h_i is positive $h_i > 0$ or negative $h_i < 0$. All vertices comprising P_0 are

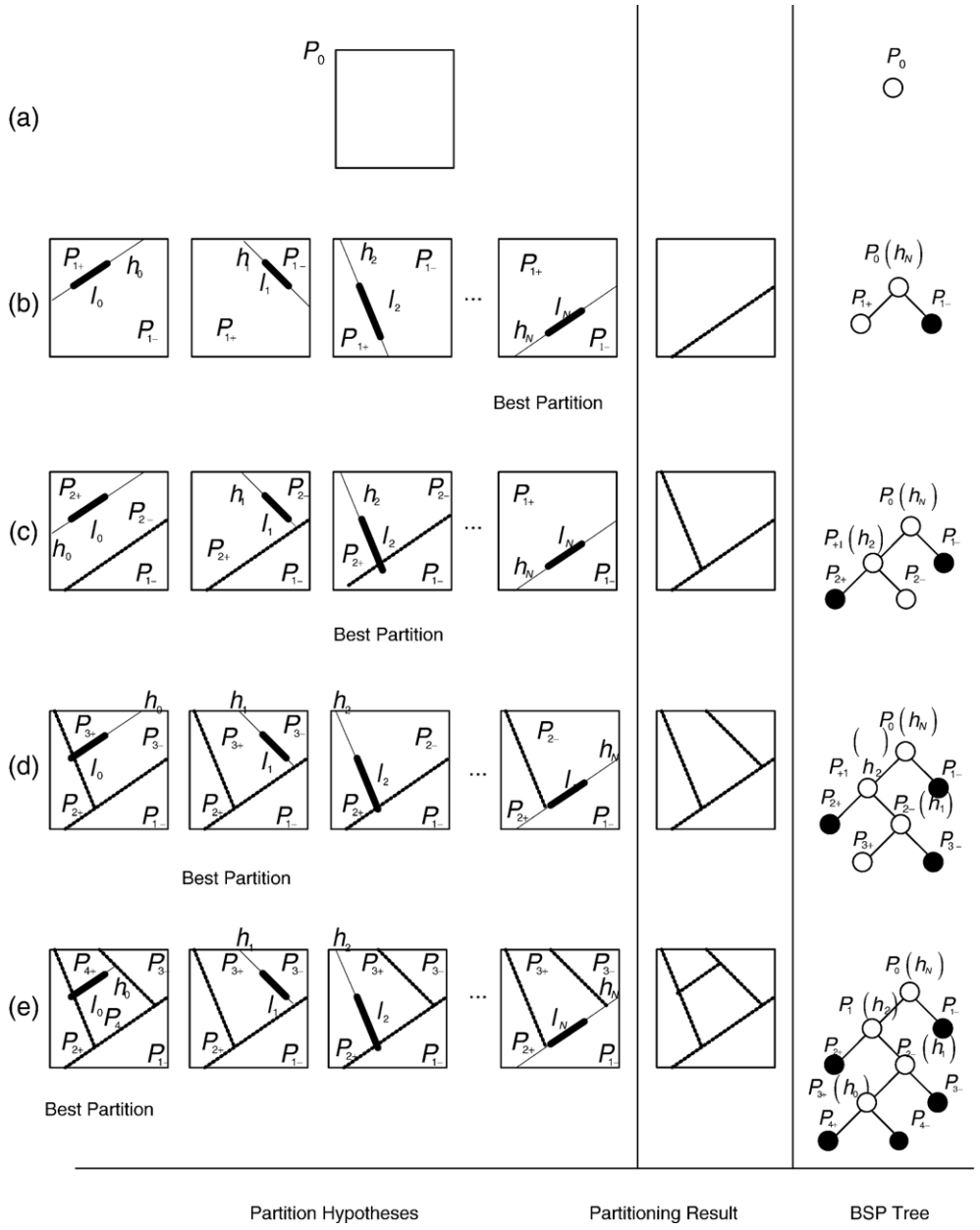

Fig. 11. BSP tree construction; in the BSP tree, black circle and white circle represent “closed” and “open” polygon respectively.

stored as the root node of a binary tree. As an internal node, the root node holds a hyperline and all vertices comprising P_0 (Fig. 11(a)). The two sub-polygons are represented by children of the node (Fig. 11(b)). A leaf node, in contrast to an internal node, holds no hyperline and corresponds to an unpartitioned polygon. This process is repeated again recursively until no leaf node is generated by BUS partitioning. A different sequence of hyperlines will produce different partitioning results. Thus, determining an optimal sequence of hyperlines and eliminating spurious hyperlines governs the success of our building extraction system. This partitioning optimality will be achieved by the hypothesis-and-test framework in a hierarchical manner. In the following two sections, we will discuss a triggering and terminating condition for partitioning in the given polygons, and an optimality criterion to select the hyperline, producing the best partitioning result.

5.4.2. BUS classification

Polygon classification is a process to determine whether or not the partitioning process is triggered over a given polygon, P_i. A polygon, P_i, is classified into four different polygon classes; "empty", "open", "closed", and "garbage" polygon. These polygon classes are predetermined depending on the labelling attributes of the member points of P_i or geometric property of P_i as follows:

- "Empty" polygon: P_i is classified as "empty" polygon if there is no member point within P_i (Fig. 12(a)).
- "Open" polygon: P_i is classified as "open" polygon if the member points of P_i are attributed to both building and non-building labels (Fig. 12(b)).
- "Closed" polygon: P_i is classified as "closed" polygon if the member points of P_i are attributed to only building label (Fig. 12(c)).
- "Garbage" polygon: P_i is classified as "garbage" polygon if P_i is classified as the "open" polygon, but any lateral length of P_i or the area of P_i is less than a certain threshold, i.e., l_{th} and a_{th} respectively (Fig. 12(d)).

A triggering and terminating condition of the partitioning process over P_i is determined depending on the polygon classification result. Basically, P_i is partitioned further if it is classified as "open" polygon, whereas the partitioning process over P_i is rejected if P_i is classified as one of "closed" polygon or "empty" polygon or "garbage" polygon. The "garbage" polygon is also considered as one of the terminating condition in order to prevent the method from partitioning a polygon into too small sub-polygons (over-partitioning) which can be ignorable as parts of the building.

5.4.3. BUS scoring

Once the partitioning of a given P_i is determined through the polygon classification, we need to select a hyperline which actually partitions P_i from all the hyperline candidates generated in Sections 5.1 and 5.2. Thus, it is necessary to have the hyperline selection criterion which provides an optimal partitioning result over P_i.

There are usually many erroneous hyperlines inherited in their formation. For instance, blunders in data-driven lines are caused when the quality of the building detection is low, and undesirable model-driven lines may be produced if building object targeted does not have particular geometric symmetry. Those distracting hyperlines must be eliminated during the partitioning process. This optimality is achieved by the hypothesize-and-test scheme with a partition scoring function. All the hyperlines are tested to partition P_i, and the partitioning result generated by each hyperline is evaluated by a partition scoring function. A hyperline, h^*, with the highest partitioning score is finally selected to partition P_i.

Fig. 11(b) shows a procedure to obtain the best result of the initial polygon, P_0. A hyperline candidate, h_i, is selected from the hyperline list, by which P_0 is divided into the positive and negative planes, P_{1+} and P_{1-}. Then, a partition scoring function, H, evaluates the partitioning results respectively over P_{1+} and P_{1-}, generated by h_i, and then determines a final score for h_i as the maximum value of the partitioning results over P_{1+} and P_{1-}. This scoring function, H, can be described as follows:

$$H(P_0; h_i) = \max(H(P_{1+}; h_i), H(P_{1-}; h_i)) \tag{1}$$

Basically, H in Eq. (1) assigns a maximum score to h_i if it produces the best partitioning result, whereas a minimum score for the worst partitioning result. That is, if P_0 is classified as the "open" polygon, H computes partitioning scores according to the ratio of label distribution of building and non-building labels over P_{1+} and P_{1-}. It aims to assign higher partitioning score when a "closed" polygon with larger area is produced by h_i (Fig. 13). The partition scoring function, H, for "open" polygon can be described by

$$\begin{aligned} H(P_{1+}; h_i) &= \frac{1}{2}\left\{\frac{N_{\text{bld}}(P_{1+}; h_i)}{N_{\text{bld}}(P_0; h_i)} + \frac{N_{\text{non-bld}}(P_{1-;h_i})}{N_{\text{non-bld}}(P_0; h_i)}\right\} \\ H(P_{1-}; h_i) &= \frac{1}{2}\left\{\frac{N_{\text{bld}}(P_{1-}; h_i)}{N_{\text{bld}}(P_0; h_i)} + \frac{N_{\text{non-bld}}(P_{1+}; h_i)}{N_{\text{non-bld}}(P_0; h_i)}\right\} \end{aligned} \tag{2}$$

where $N_{\text{non-bld}}$ and N_{bld} are functions to count numbers of building labels and non-building labels belonging to a corresponding polygon.

Once the partitioning scores for h_i are computed by Eqs. (1) and (2), remaining hyperlines are sequentially selected from the hyperline list and their partitioning scores are also measured by H. As can be seen in Fig. 11(b), a hyperline, h_N, with the maximum partitioning score is finally selected to partition P_0 by

$$h_N = \arg\max_{\forall\{h_i\}} H(P_0; h_i) \tag{3}$$

Then, all the vertices of P_0 and h_N, are stored as a root node of BSP tree, which is expanded as new child nodes with vertices of P_{1+} and P_{1-} are added to the

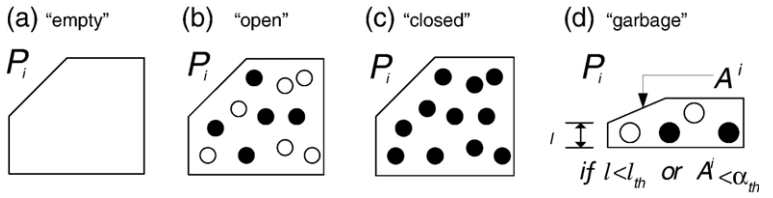

Fig. 12. Illustration of polygon classification; white and dark dot represent non-building and building labels respectively.

root node for further recursive partitioning. The same method used for the partition of P_0 is applied to P_{1+} and P_{1-} respectively, but to only an "open" polygon. In Fig. 11(b), P_{1-} is classified as the "closed" polygon and P_{1+} is classified as the "open" polygon. Thus, the partitioning process over P_{1+} is only triggered, and the partition of P_{1+} by a hyperline, h_2, is selected as the "best" result. The BSP tree of Fig. 11(b) is expanded with new child polygons, P_{2+} and P_{2-}, and the selected hyperline, h_2 (Fig. 11(c)). This process continues until no leaf node of the BSP tree can be partitioned by hyperlines (Fig. 11(d) and (e)).

5.5. BUS grouping

Fig. 14(a)–(c) shows an example how the initial BUS is segmented with a set of convex polygons by the partitioning as described in the previous section. After finishing the process by expanding a BSP tree, final leaves of the BSP tree (i.e., BUSes) are collected. Then, a BUS adjacency graph is computed with the BUSes where each node represents a BUS and each arc means the connectivity between neighbouring BUSes. The BUS grouping process is to merge only nodes in the adjacency graph, which is verified as the "building" polygon. In principle, all the "closed" polygon, which include only building points as their member points are considered as the "building" polygons. In addition, an "open" polygon, which has building labels with a significant ratio (ρ_{th}), is also classified as the "building" polygon, though it includes non-building labels. This heuristic grouping rule for an "open" polygon can be described as follows:

$$\forall P_i = \text{"open" polygon } P_i \mapsto \text{"building" polygon;}$$
$$\text{if } \rho_{pt}(P_i) = \frac{N_{\text{bld}}(P_i)}{N_{\text{mem}}(P_i)} > \rho_{\text{th}} \tag{4}$$

where ρ_{pt} is a point ratio of building labels over the total number of member points of P_i; N_{bld} and N_{mem} are functions to count numbers of building labels and non-building labels belonging to P_i. Once all "building" polygons are found, they are merged together and building boundaries are reconstructed. Fig. 14(d) shows a final result of the BUS grouping process.

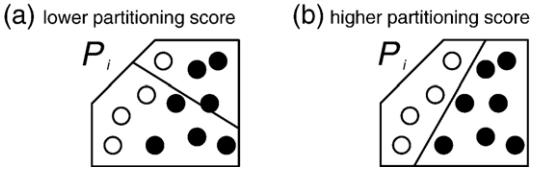

Fig. 13. Illustration of partition scoring; P_i is a parent convex polygon to be partitioned and h_i is a hyperline to partition P_i; white and dark dot represent non-building and building labels respectively.

6. Experimental results

The aim of this section is to assess the positional accuracy of the building boundaries produced by the proposed building extraction technique, and evaluate their cartographic completeness in relation to the reference building polygons provided by the Ordnance Survey. Firstly, the building extraction results over the Greenwich dataset are presented with description of the parameters selected for this technique. Finally, the performance of the proposed building extraction technique is evaluated by comparing the extracted building boundaries to the Ordnance Survey MasterMap®.

6.1. Building extraction result

The BUS organization method presented in the previous section is applied to the building detection result of Fig. 3(f), in which 170 buildings with building labels were automatically detected and bounded with rectangles. Some parameters for building extraction by the BUS organization method were pre-determined. For extracting the data-driven lines, straight lines, whose length is larger than $l_{\text{d}}=5$ (m) are extracted from the Greenwich IKONOS image. Two rectangle boxes with the width, $l_{\text{w}}=5$ (m), are created from each straight line and the building edge lines are found by the labelling

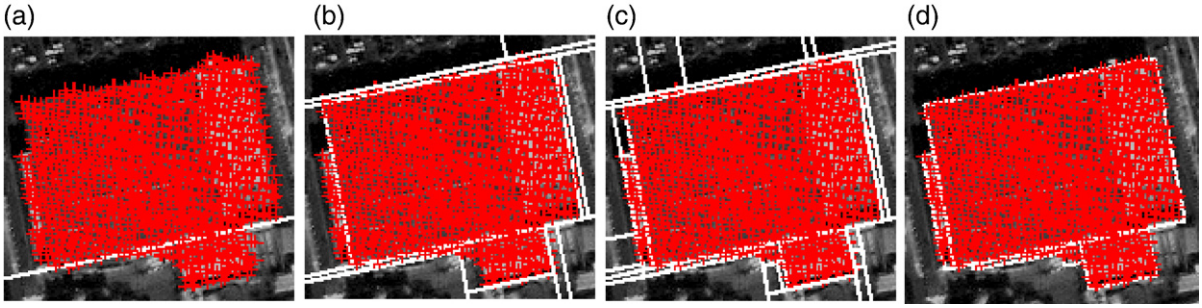

Fig. 14. Illustration of polygon cue generation and grouping; white line and red cross represent selected hyperlines and building-label points; (a) initial partitioning; (b) intermediate parttitioning; (c) final BSP partitioning; (d) "open" polygon filtering.

attributes of the LiDAR points within the boxes. The dominant angles of those edge lines are analyzed with the angular resolution of $\theta_{th}=30°$, by which the geometries of the edge lines are regularized. Then, two rectangle boxes with the width, $l_w=5$ (m), are again created from each edge line in order to determine a box growing direction towards building rooftop, and parallel lines and orthogonal lines for each edge line are created. For the generation of BUS space, the "garbage" polygon is defined with l_{th} of 3 m and a_{th} of 50 m^2. If a LiDAR data with higher point density is used, smaller l_{th} and a_{th} can be selected and thus, higher level of details of building shape can be delineated. For the BUS grouping process, an "open" polygon is verified as a building part if building-label points are included over a given polygon with the point ratio of ρ_{th} larger than 0.6. A level of details to delineate a building shape can be controlled by modifying the parameters used for the BUS grouping process. Since the point spacing irregularity of the Greenwich LiDAR DSM is extremely high, the roles of the parameter for the BUS space generating and grouping process are important in the current implementation. However, the selection of those parameters can be more stabilized if evenly distributed LiDAR with higher point density is used.

Fig. 15(a) shows a building extraction result over the Greenwich dataset (referred as UCL building map), which was generated by the BUS organization method. As can be seen in this figure, the boundaries of most building shapes were successfully reconstructed from initial BUSes generated by the building detection process. In the Greenwich scene, the contrast between the ground and buildings is not so high, and most building rooftops show complex texture and different shapes comprising a number of planar surfaces. In these circumstances, a large amount of erroneous linear cues, which are not relevant to building boundaries, could be generated from each BUS initialized. However, the suggested method efficiently pruned those distracting linear cues, and retained building edge lines. This suggests that an optimal balance between data-driven and model-driven cues for building extraction can be achieved by BUS partitioning. In particular, it can be concluded that the contribution of model-driven line cues is obvious for building extraction. For several building objects, we can observe that extended building structures with small extent were reconstructed even though those parts are occluded or have very low contrast (Fig. 15(b)). It is difficult to obtain sufficient linear cues for describing such a small object, relying on a data-driven process only. This difficulty can be overcome by employing a simple building model with minimal use of cue grouping rules. Furthermore, the recursive part segmentation of initial convex polygon for extracting building boundaries adopts a level-of-detail (LoD) strategy, which extracts significant building part comprising larger size of "building" polygon first and less significant part with smaller extent later. This is suitable for reducing building extraction errors, which are restricted only to minor building structures.

6.2. Quality assessment

The Ordnance Survey provided reference building polygons for this study (Fig. 15(c)). These data were provided in MasterMap® format. Although the detail and accuracy of the OS MasterMap® is ideal for this study, there is an inconvenient characteristic of the OS MasterMap® that hinders the accuracy assessments. The problem with the OS truth data is that it contains some significant differences in comparison to the UCL building map. This is mainly because the OS data was constructed at a different time to the acquisition of the IKONOS image and LiDAR data, from which the UCL data was generated. Since this research has an interest to evaluate the positional accuracy of the UCL building

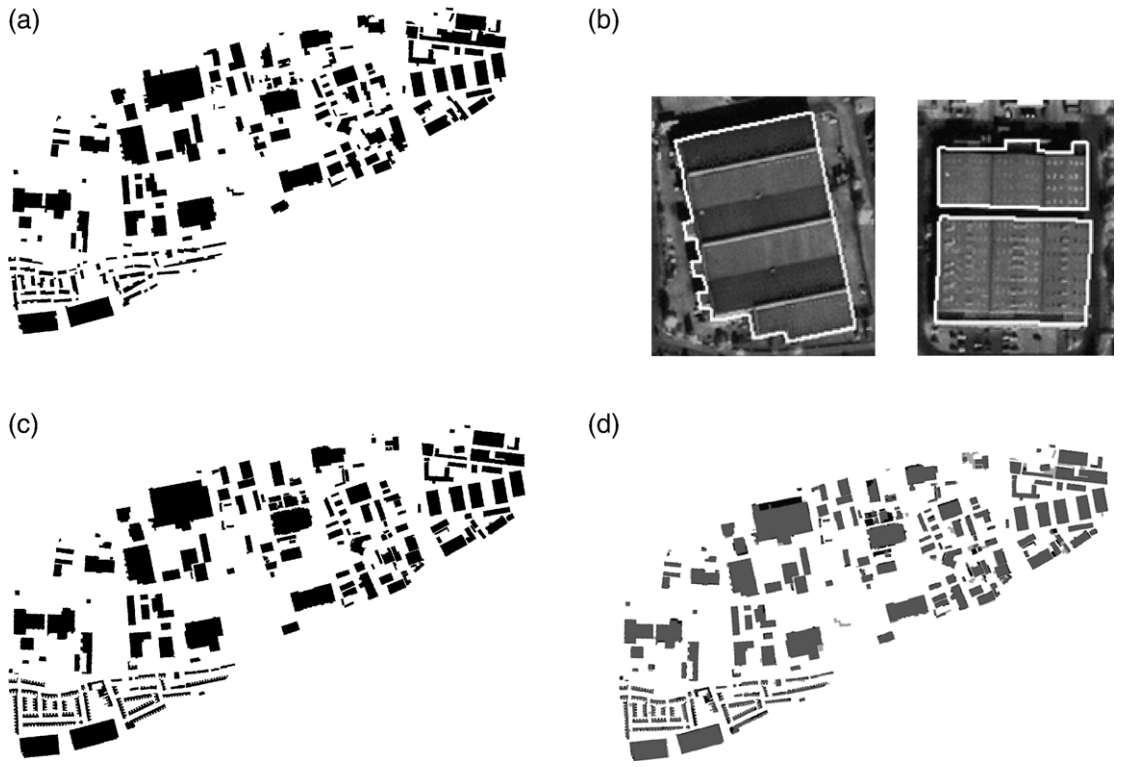

Fig. 15. Building extraction results over the Greenwich dataset; (a) UCL building map; (b) building extraction results of small extended building structures with low contrast; (c) OS MasterMap®; building extraction errors (light grey: true positives; medium grey: false positives; false negatives).

map, rather than the OS MasterMap®, those inherent faults in the OS data were removed from the UCL building maps. In addition, the scale of the features in the OS MasterMap® has been compiled at a larger scale than the one of IKONOS image. As a result, very small features cannot be clearly recognized in the IKONOS image. Thus, those small features, whose member points are less than 30 points, were removed from the OS data.

To evaluate the overall performance of the suggested building extraction system, the automatically extracted buildings (UCL building map) and the reference buildings (MasterMap®) were compared pixel-by-pixel. Based on the comparison result between two datasets, the following objective metrics (Lee et al., 2003; Rottensteiner et al., 2005) were employed in order to provide a quantitative assessment of the developed building extraction algorithm:

$$\begin{aligned}&\text{Branching Factor} : \mathrm{FP}/\mathrm{TP}\\&\text{Miss Factor} : \mathrm{FN}/\mathrm{TP}\\&\text{Completeness}(\%) : 100 \times \mathrm{TP}/(\mathrm{TP}+\mathrm{FN})\\&\text{Correctness}(\%) : 100 \times \mathrm{TP}/(\mathrm{TP}+\mathrm{FP})\\&\text{Quality Percentage}(\%) : 100 \times \mathrm{TP}/(\mathrm{TP}+\mathrm{FN}+\mathrm{FP})\end{aligned} \tag{5}$$

Here, TP (True Positive) is the number of building pixels classified by both datasets, TN (True Negative) is the number of non-building pixels classified by both datasets, FP (False Positive) is the number of building pixels classified only by the UCL building map, and FN (False Negative) is the number of building pixels classified only by the OS MasterMap®.

Each metric mentioned above provides its own quantitative measure for evaluating the overall success of building extraction algorithms. The branching factor and miss factor are closely related to the boundary delineation performance of the building extraction system. The branching factor increases when the algorithm incorrectly over-classifies building pixels, while the miss factor increases when the algorithm misses the building pixels. The completeness (often, called building detection percentage) and the correctness are the ratio of the correctly classified building pixels (TP) with respect to the total building pixels classified by the OS MasterMap® (TP+FN) and by the automatic process (TP+FP) respectively. Both factors indicate a measure of building detection performance. Higher building detection rate can be achieved as either less missing building pixels or less over-classified building pixels are produced by the algorithm. The

quality percentage combines aspects of both boundary delineation accuracy and building detection rate to summarize system performance. To obtain 100% quality, the algorithm must correctly classify every building pixel, without missing any building pixel (FN=0) and without over-classifying any background pixel (FP=0).

Table 1 shows the pixel classification results acquired by comparing the UCL building map to the OS MasterMap®. Based on this result, the evaluation measures on the UCL building map were computed by Eq. (5). The suggested building extraction technique showed 0.11 delineation performance in branching factor, while the miss factor presented slightly poor performance (0.13). This indicates that the number of building pixels over-classified by the algorithm is less than the number of missed building pixels. Also, the automatic process detected buildings at the rate of 88.3% in completeness (building detection percentage), while 90.1% in correctness. By the similar reason to the aforementioned delineation performance, we achieve higher correctness performance than the completeness due to the tendency of the developed system to producing less false positive pixels than false negatives. Finally, the overall success of the technique was evaluated as 80.5% extraction quality (quality percentage).

The quality assessment results of building extraction systems using various primary datasets have been reported by many researchers. Lee et al. (2003) developed a method to extract outlines from IKONOS imagery with multi-spectral bands. An unsupervised classification results based on ISODATA algorithm was used as evidences for detecting buildings, and a squaring algorithm regularized building outlines constrained with two dominant directions. They reported the technique achieved a building detection rate of 64%, a branching factor of 0.39 and a quality percentage of 51%. Kim and Muller (2002) integrated LiDAR with multi-spectral bands of IKONOS for building extraction. Parish (2002) evaluated the performance of the system, which showed that contribution of LiDAR data improved the building detection rate up to 86% compared to Lee et al.'s work, but ignorance of geometric regulation in building description process resulted in worse branching factor of 1.0, thereby rather poor quality percentage of 64%. Based on the theory of Dempster–Shafer, Rottensteiner et al. (2005) detected buildings from LiDAR data and an airborne orthophoto with multi-spectral bands, which were re-sampled at 1 m resolution. When the reference map was compared to detected buildings in a per-pixel evaluation, the system detected buildings with the completeness of 94% and the correctness of 80%. In a per-building evaluation, they reported that 95% of the buildings larger than 50 m^2 were detected, whereas about 89% of the detected buildings were correct. With multiple high-resolution airborne images (10 cm of ground sampling) and a high-quality elevation model, Fradkin et al. (2001) reported that they achieved a building detection rate of 80% and a quality percentage of 78%, and showed a branching factor can increase up to 0.1–0.37 depending on image quality and perspective view used. It is difficult to make direct comparison since the aforementioned techniques were applied to scenes with different scene complexity and primary data sources. However, in relation to other researches, it can be concluded that the building detection performance achieved by the suggested technique is acceptable.

Table 1
Pixel classification results

Pixel classification	Pixels
True positive (TP)	282,726
True negative (TN)	253,5114
False positive (FP)	31,142
False negative (FN)	37,517

6.3. Error analysis

Although the developed building extraction system showed a satisfactory result, it also produced certain number of building extraction errors, which should be reduced to achieve a more accurate extraction of building objects. The errors apparent in the result generated by the technique can generally be divided into three categories:

6.3.1. Building detection error

Most of the false negative pixels in the evaluation metrics (black pixels in Fig. 15(d)) were generated by under-detection of buildings. This problem is partly caused as the NDVI classification described in Section 4.3 tends to over-remove building points, in particular over the terraced houses with long and narrow structure (see lower left part in Fig. 15(d)), and is partly caused as LiDAR data is acquired with extremely low density over some buildings (see upper middle part in Fig. 15(d)). A consequence of both effects is to produce very small regions, which makes the algorithm fail to recognize them as isolated buildings by given threshold in the building detection stage. The false negatives can be reduced more if LiDAR data are sufficiently acquired with higher density. However, it is still hard to safely remove trees, especially when they are closely located nearby buildings (Baillard and Maitre, 1999). This

problem could be solved with additional data, either by using a height difference between the first and last returns of LiDAR or employing surface roughness characterising vegetated objects or segmenting LiDAR intensity data (Maas, 2001).

6.3.2. Building delineation error

Medium grey-coloured pixels in Fig. 15 represent the false positive errors, indicating the delineation quality of building outlines. Those errors are caused by several factors. These include the inherent planimetric accuracy of input data (i.e., IKONOS image, LiDAR data, and OS reference data), a degree of point spacing irregularity and density of LiDAR data. Most of boundary delineation errors were deviated from the OS reference data with one or two pixels where LiDAR measurements were sufficiently acquired over buildings. However, as LiDAR points were acquired with less point density, more errors were produced around its boundaries. This is caused by the fact that the detection of data-driven line and model-driven line is more difficult over building rooftops with coarser point density than the one with denser point density. As a result, mis-location of both lines leads to the generation of delineation errors around building boundaries. Also, a significant lack of LiDAR measurements over rooftops could make the suggested automatic process produce many "empty" polygons so that building shapes reconstructed are intruded by removing them. There is the reverse case of the former problem. In the BUS grouping, the presence of non-building points around building outlines were used as important evidence to verify the "building" polygons. If there is a significant deficiency of non-building points on the ground near buildings, erroneous "closed" polygons are generated, and could be mistakenly verified as the "building" polygons. A consequence is to extrude original building shapes as a reconstruction result. Therefore, uniform acquisition of LiDAR data with higher point density will make the proposed system more controllable, and thus reduce the building delineation errors.

6.3.3. Reference data error

Some building extraction errors are presented in Fig. 15(d) due to the inherent faults in the OS building polygons. Although these errors were carefully removed from the UCL building map before performing the quality assessment, the inherent faults in the OS data can still be found, and thus some parts of buildings were missed in the OS MasterMap® although the UCL building map can successfully delineate boundaries of a building based on the result of LiDAR measurements and IKONOS image. This is because the OS data was constructed at a different time to the acquisition of the IKONOS image and LiDAR data, from which the UCL data was generated. The analysis of the reference errors suggests that the developed building extraction technique can also be used for applications detecting changes in an urban environment and supporting map compilation.

7. Conclusions

This paper presented a new building extraction method which automatically detects building objects and delineates their boundaries using IKONOS image and LiDAR data in coherent collaboration of two different datasets. The developed technique consists of a two step procedure: building detection and building description. A building detection method was introduced to reduce a scene complexity in urban areas and simplify the building description process. The technique subsequently detected dominant features comprising the urban scene, and finally isolated buildings from surrounding features. The results prove that terrain information extracted from LiDAR data and chromatic cues provided by multi-spectral bands of IKONOS imagery is important knowledge for detecting buildings. Based on the BSP (Binary Space Partitioning) tree algorithm, the BUS (Building Unit Shape) organization method was proposed as a new middle-level of feature grouping tool for reconstructing building outlines. The approach has favourable characteristics in many aspects. First, the proposed framework can accommodate diversified building shapes. Complexity of the building shape is constrained in a data-driven way. The degrees of geometric freedom, constraining the shape of the building, is determined by analysing dominant directionalities from data-driven lines in IKONOS imagery, which are not limited to pre-specified models. Secondly, classification results of LiDAR points serve as evidences to employ box-type model to produce model-driven lines, thereby completely or partly missed building boundaries can be compensated. Thirdly, internal topological relations amongst linear features are implicitly constructed, and a subjective features grouping relying on pre-defined threshold can be avoided. Finally, optimality of the feature grouping is achieved by a partition scoring function, which makes polygonal cues generated in a level-of-detail strategy. This makes a significant loss of building parts minimised.

The developed building extraction system was applied to a sub-scene of the Greenwich area. The results showed that the suggested system can successfully delineate most buildings in the scene. The overall

quality of the results was evaluated by objective evaluation metrics in comparison to the OS MasterMap®. This evaluation determined the branching factor to be 0.11, the miss factor to be 0.13, the completeness to be 88.3%, the correctness to be 90.1% and the quality percentage to be 80.5%, which in relation to other studies, is satisfactory. However, a qualitative error analysis showed that the extracted building polygons tend to overlook some buildings (false negative pixels) and misclassify some objects as buildings (false positive pixels) because of poor point density of the Greenwich LiDAR data. These errors could be reduced if LiDAR measurements used are more evenly distributed with higher density. Furthermore, the error analysis demonstrated that the technique can automatically find the inherent difference between the new data and the OS MasterMap®, which may be a useful tool for improving a laborious topographic mapping procedure by highlighting problematic areas.

Acknowledgements

We would like to express our gratitude to Professor Peter Muller at University College London and Infoterra, U.K., who made available the IKONOS imagery and LiDAR data referred to in this article. Also, we would like to acknowledge the support of the Ordnance Survey, U.K., who kindly provided the OS MasterMap® for this research.

References

Ameri, B., 2000. Automatic recognition and 3D reconstruction of buildings from digital imagery. PhD thesis, Institute of Photogrammetry, University of Stuttgart, 110 pp.

Baillard, C., Maitre, H., 1999. 3D reconstruction of urban scenes from aerial stereo imagery: A focusing strategy. Computer Vision and Image Understanding 76 (3), 244–258.

Baillard, C., Zisserman, A., 2000. A plane-sweep strategy for the 3D reconstruction of buildings from multiple images. International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences 33 (Part B3), 56–62.

Brenner, C., 2000. Towards fully automatic generation of city models. International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences 33 (Part B3), 85–92.

Brenner, C., 2004. Modelling 3D objects using weak CSG primitives. International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences 35 (Part B3), 1085–1090.

Burns, J.B., Hanson, A.R., Riseman, E.M., 1986. Extracting straight lines. IEEE Pattern Analysis and Machine Intelligence 8 (4), 425–455.

Chen, L., Teo, T., Shao, Y., Lai, Y., Rau, J., 2004. Fusion of LiDAR data and optical imagery for building modelling. International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences 35 (Part B4), 732–737.

Dowman, I., 2000. Automatic feature extraction for urban landscape models. Proceedings of the 26th Annual Conference of the Remote Sensing Society, Leicester. September 12–14, on CD-ROM.

Dowman, I., 2004. Integration of LiDAR and IFSAR for mapping. International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences 35 (Part B2), 90–100.

Fischer, A., Kolbe, T.H., Lang, F., Cremers, A.B., Forstner, W., Plumer, L., Steinhage, V., 1998. Extracting buildings from aerial images using hierarchical aggregation in 2D and 3D. Computer Vision and Image Understanding 72 (2), 185–203.

Fradkin, M., Maitre, H., Roux, M., 2001. Building detection from multiple aerial images in dense urban areas. Computer Vision and Image Understanding 82 (3), 181–207.

Fuchs, H., Kedem, Z.M., Naylor, B.F., 1980. On visible surface generation by a priori tree structures. Computer Graphics 14 (3), 124–133.

Guo, T., Yasuoka, Y., 2002. Snake-based approach for building extraction from high-resolution satellite images and height data in urban areas. Proceedings of the 23rd Asian Conference on Remote Sensing, November 25–29, Kathmandu. 7 pp., on CD-ROM.

Haverkamp, 2004. Automatic building extraction from IKONOS imagery. Proceedings of ASPRS 2004 Conference, May 23–28, Denver. 8 pp., on CD-ROM.

Jaynes, C., Stolle, T., Collins, R., 1994. Task driven perceptual organization for extraction for rooftop polygons. Proceedings of ARPA Image Understanding Workshop, Monterey, November, pp. 359–365.

Kim, T., Muller, J.-P., 1999. Development of a graph-based approach for building detection. Image and Vision Computing 17 (1), 3–14.

Kim, J.R., Muller, J.-P., 2002. 3D reconstruction from very high resolution satellite stereo and its application to object identification. International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences 34 (Part 4) (7 pp., on CD-ROM).

Lee, D.S., Shan, J., Bethel, J.S., 2003. Class-guided building extraction from IKONOS imagery. Photogrammetric Engineering and Remote Sensing 69 (2), 143–150.

Lin, C., Nevatia, R., 1998. Building detection and description from a single intensity image. Computer Vision and Image Understanding 72 (2), 101–121.

Lumia, R., Shapiro, L., Zuniga, O., 1983. A new connected components algorithm for virtual memory computers. Computer Vision, Graphics, and Image Processing 22 (2), 287–300.

Maas, H.-G., 2001. The suitability of airborne laser scanner data for automatic 3D object reconstruction. In: Grün, A., Baltsavias, E.P., Henricsson, O. (Eds.), Automatic Extraction of Man-made Objects from Aerial and Space Images, vol. III. Balkema Publishers, Lisse, pp. 345–356.

Maas, H.-G., Vosselman, G., 1999. Two algorithms for extracting building models from raw laser altimetry data. ISPRS Journal of Photogrammetry and Remote Sensing 54 (2–3), 153–163.

Moons, T., Frere, D., Vandekerckhove, J., Gool, L., 1998. Automatic modeling and 3D reconstruction of urban house roofs from high resolution aerial imagery. Proceedings of the European Conference on Computer Vision, pp. 410–425.

Parish, B., 2002. Quality assessment of automated object identification in IKONOS Imagery. M.Sc. Thesis, University College London, London.

Park, W., Kwak, K., Su, H., Kim, T.G., 2000. Line rolling algorithm for automated building extraction from 1-meter resolution satellite images. Proceedings of the International Symposium on Remote Sensing, Kyungju, November 1–3, pp. 31–36.

Radha, H.R., Leonardi, R., Vetterli, M., Naylor, B.F., 1991. Binary space partitioning tree representation of images. Journal of Visual Communication and Image Representation 2 (3), 201–221.

Rottensteiner, F., Trinder, J., Clode, S., Kubik, K., 2005. Using the Dempster–Shafer method for the fusion of LIDAR data and multi-spectral images for building detection. Information Fusion 6 (4), 283–300.

Schenk, T., Csatho, B., 2002. Fusion of LiDAR data and aerial imagery for a more complete surface description. International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences 34 (Part 3), 310–317.

Shufelt, J.A., McKeown, D.M., 1993. Fusion of monocular cues to detect man-made structures in aerial imagery. Computer Vision and Image Understanding 57 (3), 307–330.

Sithole, G., Vosselman, G., 2004. Experimental comparison of filter algorithms for bare-earth extraction from airborne laser scanning point clouds. ISPRS Journal of Photogrammetry and Remote Sensing 59 (1–2), 85–101.

Sohn, G., Dowman, I.J., 2001. Extraction of buildings from high resolution satellite data. In: Grün, A., Baltsavias, E.P., Henricsson, O. (Eds.), Automatic Extraction of Man-made Objects from Aerial and Space Images, vol. III. Balkema Publishers, Lisse, pp. 345–356.

Sohn, G., Dowman, I.J., 2002. Terrain surface reconstruction by the use of tetrahedron model with the MDL criterion. International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences 34 (Part 3), 336–344.

Suveg, I., Vosselman, G., 2004. Automatic 3D reconstruction of buildings from aerial images. ISPRS Journal of Photogrammetry and Remote Sensing 58 (3–4), 202–224.

Vosselman, G., 1999. Building reconstruction using planar faces in very high density height data. International Archives of Photogrammetry and Remote Sensing 32 (Part 2), 87–92.

Weidner, U., Forstner, W., 1995. Towards automatic building reconstruction from high resolution digital elevation models. ISPRS Journal of Photogrammetry and Remote Sensing 50 (4), 38–49.