

Article

An Object-Based Bidirectional Method for Integrated Building Extraction and Change Detection between Multimodal Point Clouds

Chenguang Dai ¹, Zhenchao Zhang ^{1,2,*}  and Dong Lin ³ 

¹ School of Surveying and Mapping, Information Engineering University, Zhengzhou 450001, China; cgdai2008@163.com

² Faculty of Geo-Information Science and Earth Observation (ITC), University of Twente, 7514AE Enschede, The Netherlands

³ School of Space Information, Space Engineering University, Beijing 101400, China; dong.lin@tu-dresden.de

* Correspondence: z.zhang-1@utwente.nl; Tel.: +86-150-9330-3012

Received: 31 March 2020; Accepted: 22 May 2020; Published: 24 May 2020



Abstract: Building extraction and change detection are two important tasks in the remote sensing domain. Change detection between airborne laser scanning data and photogrammetric data is vulnerable to dense matching errors, mis-alignment errors and data gaps. This paper proposes an unsupervised object-based method for integrated building extraction and change detection. Firstly, terrain, roofs and vegetation are extracted from the precise laser point cloud, based on “bottom-up” segmentation and clustering. Secondly, change detection is performed in an object-based bidirectional manner: Heightened buildings and demolished buildings are detected by taking the laser scanning data as reference, while newly-built buildings are detected by taking the dense matching data as reference. Experiments on two urban data sets demonstrate its effectiveness and robustness. The object-based change detection achieves a recall rate of 92.31% and a precision rate of 88.89% for the Rotterdam dataset; it achieves a recall rate of 85.71% and a precision rate of 100% for the Enschede dataset. It can not only extract unchanged building footprints, but also assign heightened or demolished labels to the changed buildings.

Keywords: change detection; building extraction; point clouds; object-based; airborne laser scanning; dense image matching

1. Introduction

Object extraction and change detection are two of the most important tasks in remote sensing [1,2]. Object extraction derives topographic information from one single epoch, whereas change detection compares remote sensing data from two epochs to derive change information. Airborne photogrammetry and airborne laser scanning (ALS) are two widely-used techniques with respect to the acquisition of remote sensing data over urban scenes. Remote sensing data, acquired from different techniques, vary in data dimensionality, accuracy, noise level and data gap level [3,4].

A comparison in the data quality between ALS and airborne photogrammetry was referred to [5,6]. The main product from laser scanning is usually three-dimensional (3D) point clouds. In contrast, airborne photogrammetry produces geo-referenced imagery, point clouds, Digital Surface Models (DSMs) and orthoimages. In photogrammetry, point clouds are generated through dense image matching (DIM) [7,8]. Generally, ALS point clouds are more accurate than DIM point clouds in terms of vertical accuracy. The former usually contains less noise than the latter. However, DIM provides, not only geometric information in point clouds or DSM, but also spectral information in the orthoimage. Previous work suggests that the spectral information from orthoimage is complementary to the

geometric information for object extraction or change detection tasks [9–11]. That is, even though the accuracy and noise level in DIM data are less satisfying than those of ALS data, the spectral information can fill the gap to some extent.

Extracting topographic objects and detecting topographic changes in urban scenes are fundamental tasks in urban planning and environmental monitoring. This paper aims to extract building footprints and detect building changes between ALS data and photogrammetric data. This is applicable to the situation of several mapping agencies, where laser scanning data are already available as archive data, while aerial images are routinely acquired every one or two years for updates. When the remote sensing data, available at different epochs, are heterogeneous (i.e., with different platforms and sensor characteristics), such heterogeneity makes change detection challenging.

The tasks of building extraction and change detection are closely associated. Building changes include new building, demolished building and heightened building. Tran et al. [1] suggest that most change detection methods apply two steps for change detection: Firstly, extract objects from both epochs; secondly, compare the two epochs for change information. In this case, object extraction is explicitly implemented before change detection. Since change detection aims to detect the change “from object A to object B”, it is necessary to identify what the object is in both epochs in an explicit or implicit manner. In the meantime, the accuracy of object detection affects the change detection results in a sequence. Therefore, this paper aims at integrating building extraction and change detection in a single workflow. The contributions are as follows:

- We propose an unsupervised method for integrated building extraction and change detection between ALS data and photogrammetric data. The outputs contain not only building footprints, but also building change information. This method fuses geometric and spectral features for object extraction, and applies bidirectional object-based analysis for change detection.
- We propose Vertical Plane-to-Plane Distance (VPP) measure to indicate the height change between two heterogeneous point clouds. This measure proves effective in indicating vertical building changes.
- The acquired building footprints and change information are visualized in a single map. The experimental results on two data sets are evaluated at the pixel level and object level. Despite data noise and the differences between multimodal point clouds, the proposed method is capable of extracting buildings and detecting changes with high accuracy.

This paper is organized as follows: Section 2 reviews the related work on point cloud-based semantic segmentation and change detection. Section 3 presents our method. Section 4 provides details on the study areas and experimental settings. Section 5 presents the results and discussion. Section 6 concludes the paper.

2. Related Work

2.1. Point Cloud Classification

Point cloud classification refers to assigning a category label to each point in a point cloud. Point cloud classification methods can be divided into four categories: Rule-based classification, classification based on handcrafted features, classification with contextual features and deep learning-based classification.

Rule-based classification takes handcrafted features as geometric constraints and statistical rules [12–17]. Vosselman et al. [13] extract parameterized shapes (i.e., planes, spheres, cylinders) from the laser points using 3D Hough transform. For example, planes are extracted by 3D Hough transform aided by normal vectors calculated on the point cloud surface. A sequence of surface growing, connected segment merging and majority filtering is applied to cluster the laser points into ground, vegetation and buildings. After parameterized shapes are extracted, the classification is implemented on the extracted shapes instead of individual 3D points. Axelsson [18] classifies the ALS

point clouds into ground and non-ground points using geometry-based analysis: First, the lowest point clouds are selected as seed ground points. Then, the neighboring points are added to the initial ground surface if their distances to the surface, and angles to the plane, meet certain criteria.

The supervised classification, based on handcrafted features, is the most widely-used classification method. The principle is to extract multiple features from the point cloud and then use a classifier for recognition. Compared with the rule-based methods, its advantage is that the complex classification rules and thresholds are automatically designed by the classifier. Guo et al. [19] extract echo features and full-waveform features from the laser points, and classify the features with Random Forests. Weinmann et al. [20] comprehensively analyze the effects of different feature combinations, neighborhood sizes for feature extraction, classifiers and feature selection. Hackel et al. [21] propose an efficient point cloud classification method, which takes full account of the randomness of point cloud distribution, use K-nearest point search to determine the optimal proximity distance, and extract features based on feature vectors. However, the main question within this method is understanding how contributive features and proper classifiers are selected.

In order to make the classification map smooth and preserve the details, contextual information is added to the classification model [22]. In this manner, mutual influence of the neighboring objects is explicitly incorporated into the model. Niemeyer et al. [23] use the Conditional Random Field (CRF) framework to classify laser points. The unary term of CRF is calculated by the Random Forest with point features, and the binary term is the relationship features of the extracted neighboring points. The results are improved when contextual features are incorporated into the model. Vosselman et al. [24] propose a classification method, based on a CRF model with the features from a single segment, and the relationship between two segment as inputs. The classification results are better than the point-based classification with Random Forest, due to the full consideration of neighborhood features.

The advantage of deep learning-based classification is that it exempts the process of manual feature extraction, feature selection and classification. Deep learning-based classification is divided into five categories based on different point cloud representations: Multi-view image, 2.5D DSM, voxel, raw point cloud and point cloud graph. Among the five categories, multi-view image, 2.5D DSM and voxel-based methods are indirect methods, where the Convolutional Neural Networks (CNNs) are working on multi-view images [25,26], voxels [27] or 2.5D [28,29] data rather than raw point clouds. The classification results are then transformed to the raw point clouds. Obviously, point cloud transformation requires more computational effort and causes information loss, which hinders accurate classification. Deep learning can also work directly on the raw point cloud or graphs [30,31]. Qi et al. [32] propose PointNet for the classification and recognition of point clouds. The basic idea is to use Multi-layer Perceptron (MLP) to extract the point cloud features layer by layer, and then connect the features for classification. PointNet++ [33] differs from PointNet in that it extracts not only global features but also multi-scale local features.

Although deep learning-based methods exempt the selection of features and classifiers, a large number of training samples are still required and the hyper-parameters in the neural networks should be determined, which are labor-intensive and complicated. This paper aims to extract objects with an unsupervised method based on geometric features.

2.2. Point Cloud-Based Change Detection

Change detection is the process of identifying differences in an object by analyzing it at different epochs [34]. Change detection can be performed either between 3D data or by comparing 3D data of a single epoch to a 2D map [12,35]. Zhan et al. [36] classify the change detection methods into two categories, based on the workflow: Post-classification comparison and change vector analysis.

In post-classification comparison, independent classification maps are required for both epochs. Change detection is then performed by comparing the response at the same location between the two epochs. When the data of two epochs are of different modalities, both training and testing have to be performed at each epoch separately, thus requiring a large computational effort. Vosselman et al. [12]

propose a method to update 2D topographical maps with ALS data. The ALS data were first segmented and classified. The building segments were then matched against the building objects in the maps to detect the building changes.

Change vector analysis relies on extracting comparative change vectors between the two epochs and fuses the change indicators in the final stage [37–39]. Change vector analysis conducts a direct comparison between two epochs, which is different from post-classification comparison. The most widely-used change vector analysis between 3D data sets is DSM surface differencing, followed by point-to-point or point-to-mesh comparison [40–42]. However, traditional change vector analysis is sensitive to data problems and usually causes many false detections, especially when the data of two epochs are in different modalities.

Du et al. [43] detect building changes in the outdated DIM data using new laser points, which is the reverse setup compared with our work. Height difference and gray-scale dissimilarity are used with contextual information to detect changes in the point cloud space. Finally, the preliminary changes are refined based on handcrafted features. The limitation of the approach raised by the authors is that the boundary of changed buildings could not be determined accurately. Additionally, the method requires human intervention and prior knowledge in multiple steps. Zhou et al. [44] propose a two-step method to detect and update building changes between ALS data and multi-view images. Firstly, LiDAR-guided edge-aware dense matching is proposed to derive accurate partial changes. Secondly, hierarchical dense matching is applied to derive complete changes and update 3D information. This method omits some new or demolished buildings due to the failure of disparity extraction in the repetitive texture.

Recently, deep CNNs have demonstrated their superior performance in image-based change detection ([36,45,46]). Our previous work [29] applies Siamese CNN in change detection between ALS data and DIM data. The two types of point clouds are converted to raster images and are then fed into a Siamese CNN for change detection. However, this method can only extract zigzag change boundaries instead of fine object boundaries. This method also requires many training samples, which may not be available in some applications. In contrast, this paper aims for an unsupervised method for integrated building extraction and change detection. There is no need for large training samples and sharp change boundaries can be obtained.

3. Materials and Methods

This paper aims at an object-based method for integrated building extraction and change detection. The inputs and outputs of our method are shown in Figure 1. The old epoch contains an ALS point cloud. The new epoch contains a DIM point cloud and orthoimage. The output is the integrated map which contains building footprints and change information. To be specific, this paper detects three types of building changes in real scenarios: *building heightened*, *building new* and *building demolished* as shown in the right of Figure 1. *Building heightened* indicates that a building exists in two epochs but has changed in height due to construction work. *Building new* indicates that no building exists in one epoch and a new building is newly-built in the new epoch. *Building demolished* indicates that a building exists in old epoch and demolished in the new epoch.

The proposed method for integrating building extraction and change detection is shown in Figure 2. Suppose that the ALS and DIM point clouds are already registered to the uniform world coordinate [47]. The proposed method is designed based on the characteristics of ALS data and DIM data. ALS point cloud and DIM point cloud show heterogeneous characteristics. Object-based change detection is more robust to point cloud noise and data gaps compared to surface differencing method [41]. The major steps are object extraction and bidirectional object-based analysis.

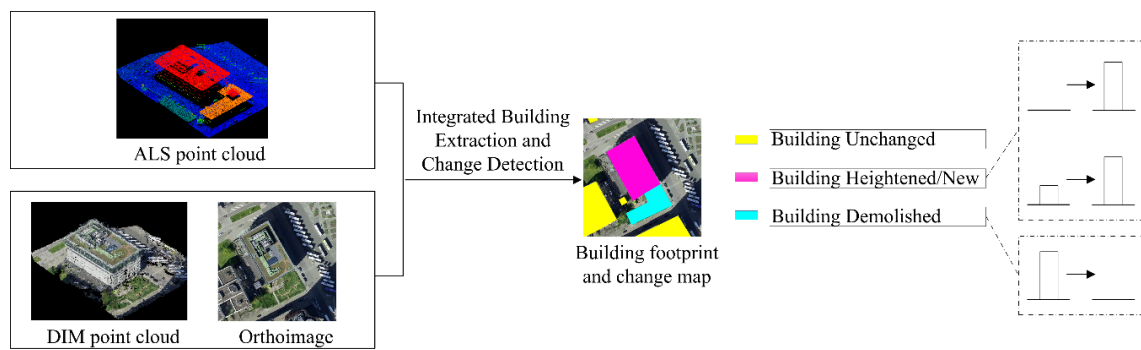


Figure 1. Overview of the input and output of the proposed method.

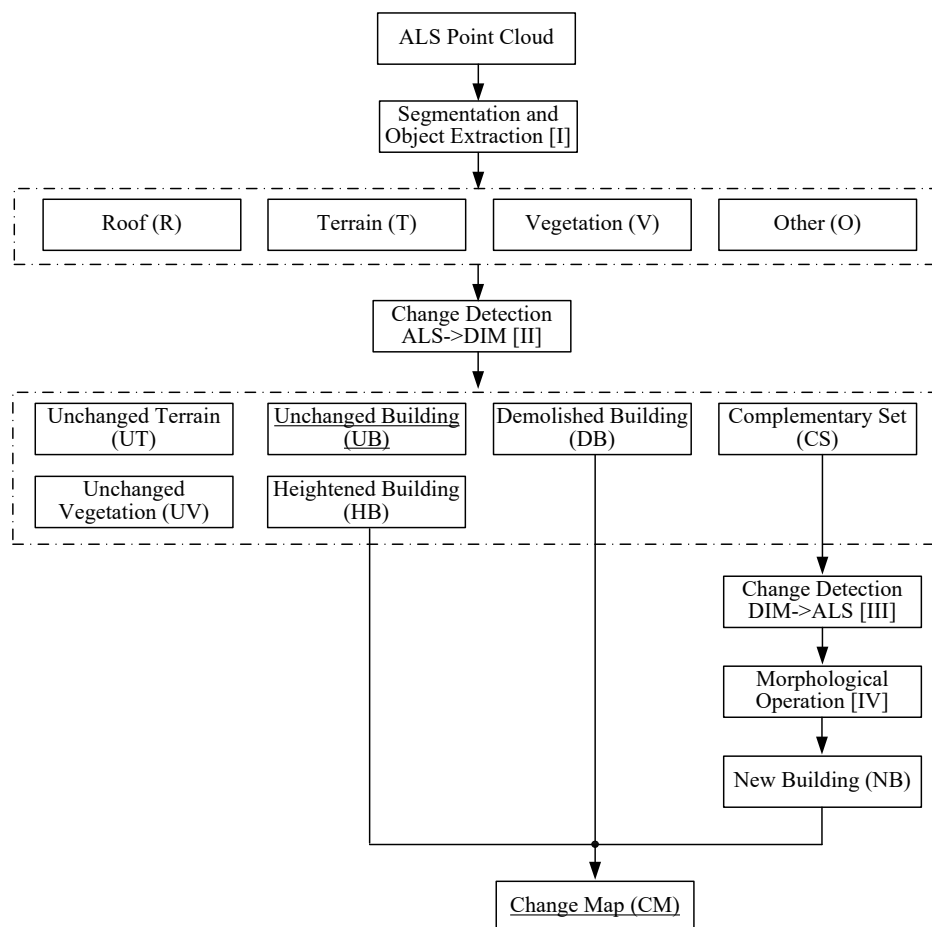


Figure 2. Overview of proposed method. *R, T, V, UT, UB* etc. in the round brackets indicate products. I, II, III and IV in the square brackets indicate four steps. The final products are underlined: unchanged building (*UB*) and change map (*CM*).

ALS point cloud is usually more precise and contain less noise than the DIM point cloud. Therefore, point cloud segmentation and object extraction are supposed to perform well on the ALS data. Object extraction provides initial building footprints, terrain and vegetation locations. When the ALS data are compared to the DIM data on these building footprints, unchanged buildings, heightened buildings and demolished buildings are detected. The remaining regions may still contain newly-built buildings, where buildings do not exist in the old epoch but appear in the new epoch. Therefore, only in the remaining regions, new buildings are detected by comparing the DIM data to the ALS data. The four major steps are as below:

- Step I: The method starts from ALS point cloud and perform point cloud filtering and surface-based segmentation. Terrain (T), roof segments (R) and vegetation (V) are extracted from laser points. Set the whole study area as A and set the remaining irrelevant points as O :

$$A = R \cup T \cup V \cup O \quad (1)$$

- Step II: By comparing the ALS point-based segments to the DIM points, the unchanged building (UB), unchanged terrain (UT), unchanged vegetation (UV), heightened building (HB) and demolished building (DB) are detected. This step is named “Change detection ALS->DIM” since ALS data are used as referred data. Then calculate the complementary set (CS) of the study area A :

$$CS = A - UT \cup UV \cup UB \cup HB \cup DB \quad (2)$$

The complementary set is uncertain regions where new building might be detected in the following steps.

- Step III: In the complementary set, the newly-built buildings (NB) are detected by comparing the DIM data to the ALS data. This step is named “Change detection DIM->ALS” since DIM data are used as referred data.
- Step IV: The newly-built building masks are post-processed by morphological operation. The change map (CM) is acquired by taking the union of heightened building (HB), demolished building (DB) and new building (NB):

$$CM = HB \cup DB \cup NB \quad (3)$$

The four key steps will be explained in details in the following four sub-sections. The proposed method not only extracts unchanged building footprints, but also detect building changes. The heightened buildings (HB) and demolished buildings (DB) are detected in Step II “Change detection ALS->DIM”; The new buildings (NB) are detected in Step III “Change detection DIM->ALS”. All the three types of building changes can be detected after this bidirectional change detection.

3.1. Object Extraction from ALS Point Cloud

Firstly, terrain, roofs and vegetation are extracted from the ALS points based on their geometric properties. The terrain points are usually low and form a smooth surface. The roofs are high and often planar or smooth surfaces. The vegetation canopy usually forms clusters with unordered normal vectors. The object extraction contains four steps: (1) point cloud filtering; (2) surface-based segmentation; (3) segment-based screening; and (4) connected component analysis. Point cloud filtering is to separate non-ground points from ground points. Surface-based segmentation is applied to extract planar or smooth segments from the non-ground points. Segment-based screening is applied to select roof segments. Connected component analysis is applied to extract vegetation clusters.

Progressive TIN Densification is used for ALS point cloud filtering [18]. The method proves effective and robust in filtering non-ground points and also maintaining the local terrain details. The method first selects some initial topographic points as seed points. Then other points are included to or excluded from the terrain points based on the geometric relationship between the neighboring points and the initial topographic surface. The main steps are as follows:

- (1) Select the initial seed points. Construct a square grid with side length L , and select the lowest points in each grid as the initial seed points. Construct a Triangular Irregular Network (TIN) based on seed points;

- (2) Iterate over each laser point until all the non-seed points have been considered. If its distance to the nearest TIN surface d and the included angle α_i ($i \in \{1, 2, 3\}$) to the TIN surface are both less than the thresholds, it belongs to the ground point.

After filtering the ground points from the ALS point cloud, the remaining points are mainly building roofs, walls and vegetation; A small number of remaining points might be cars, railings, or street lights. Since our goal is to detect building footprints and building changes, the roof points need to be extracted. In real urban scenarios, most roof points can be characterized geometrically by planes or smooth surfaces [13]. A surface-based segmentation method is used for planar surface extraction. The method not only works on extracting complete planes, but also break curved surface into small planes. This is especially useful for non-planar roofs, for example a dome.

Surface-based segmentation is a “bottom-up” clustering method. Firstly, 3D Hough Transform is used to extract the “seed planes” from the point cloud. These seed planes are point segments located in the same plane, which are mainly roofs or walls. Due to the impact of point cloud noise or data gaps, a roof might be split into multiple plane segments. Then, the surface growing algorithm is used to analyze the points close to each “seed plane”. If its distance to the nearest point on the plane is less than D and its distance to the fitted plane is less than D_0 , the neighboring point is added to this plane. After new points are added to the plane, the plane parameters are recalculated before testing the next point. After surface growing, the initial seed planes are expanded and large segments are obtained. Figure 3a is the initial ALS point cloud. Figure 3b shows the segmentation results after surface growing where different colors indicate different planes. Some scattered points or vegetation points are not segmented because they do not belong to any plane. These points are displayed in white.

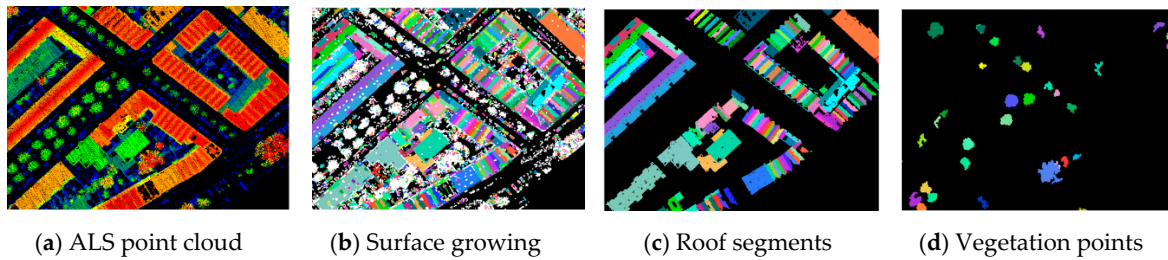


Figure 3. ALS point cloud segmentation and object extraction.

After segmentation, segment-based screening is applied to select roof segments. The planar segments in Figure 3b contains, not only real roof segments, but also wall segments and vegetation points that happen to be located in a plane. Segment-based features are calculated for each segment. Suppose that a segment contains N points, the coordinates of points are (X_i, Y_i, Z_i) , $i \in [1, N]$. The 3D coordinates are used to calculate a covariance matrix M and three eigenvalues $\lambda_1, \lambda_2, \lambda_3$ ($\lambda_1 \geq \lambda_2 \geq \lambda_3$). Some features calculated from the three eigenvalues can characterize the geometry of the segment [5]. These features are used to identify roof segments. In this paper, four features are used to select the roof segments: the segment size N , the inclination angle θ , the normalized height nH , and the residual of plane fitting (RPF) σ . Segment size, or number of points in this segment, is used to eliminate small segments. Inclination angle is used to eliminate wall segments. Normalized height is used to distinguish low segments from high segments. RPF is used to eliminate noisy segments,

$$nH = \frac{\sum_{i=0}^N (Z_i - Z_{i0})}{N} \quad (4)$$

$$\sigma = \frac{\sum_{i=0}^N |d_i|}{N} \quad (5)$$

where Z_{i0} is the ground height interpolated from the neighboring ground points. While d_i is the distance from each point to the fitted plane. A segment is regarded as roof only if it passes the check

based on the four features. The thresholds for the features are set based on trial tests and ALS data quality. The threshold values will be given in the experimental setup section.

In addition, all the points that were not previously segmented are also discarded, i.e., the white points in Figure 3b. Roof extraction results are shown in Figure 3c. Comparison between Figure 3b,c shows that both horizontal roofs and gable roofs can be extracted correctly. Small segments, such as car points and small vegetation segments are excluded.

The next step is to extract vegetation points from the unsegmented points in Figure 3b. Canopy points are adjacent, forming a cluster. Canopy clusters can be easily extracted by connected component analysis method [13]. This method takes the neighboring points into consideration, and groups all the points within a certain distance into the same cluster. After that, small clusters with less than a certain number of points are considered other objects and discarded according to the point cloud density.

In the ideal case, each canopy forms a cluster of points. However, the canopy clusters extracted in Figure 3d show that some canopy clusters are fractal due to missing data or low density. Note that even if some vegetation points cannot be grouped and extracted in this step, our bidirectional change detection method can still guarantee a correct building change detection result. This is because in the change detection from DIM to ALS step, these uncertain remaining regions are re-analyzed again.

3.2. Change Detection: ALS -> DIM

Object extraction divides the ALS point cloud into terrain, roof, vegetation and other classes. When projecting the roof points onto the ground, building footprints are obtained in the ALS data. When a building exists in the old epoch, it could be unchanged, demolished or heightened compared to the DIM data. In this step, the roof points in the ALS data are taken as reference. The neighboring DIM points are compared to each ALS roof segment to detect possible building changes.

To make the change detection robust to point cloud noise and data gaps, the Vertical Plane-to-Plane Distance (VPP) D_p is proposed as an indicator to measure the change scale between ALS roof segment and its corresponding DIM points. For each point, (X_i, Y_i, Z_i) , $i \in [1, N]$ on the ALS roof segment, the elevation on the DIM surface at location (X_i, Y_i) is calculated. Since the DIM point cloud is relatively noisy, a cube with a side length of l is constructed at the center (X_i, Y_i) . The average height of all points in this cube is taken as the DIM elevation Z_i^{dim} at (X_i, Y_i) . Next, calculate the vertical elevation change for each point on the ALS segment, and take the average as the VPP distance:

$$D_p = \frac{\sum_{i=1}^N (Z_i - Z_i^{dim})}{N} \quad (6)$$

The segments with VPP distance D_p greater than threshold are considered as changed building segments, otherwise they are unchanged. Figure 4 shows the schematic diagram of VPP distance measure between an old roof (blue) and a new roof (red). In contrast, point-to-fitted-plane distance (PFP) is also widely used to represent the distance between two planes. Point-to-nearest-point distance (PNP) is calculated by taking the distance from each ALS point to the nearest DIM point and average over all the ALS points. The PNP distance is more prone to point cloud noise and largely affected by the data gaps, since the closest point has to calculate the closest distance. In our case, a building change happens when its height is changed in a vertical direction, so VPP distance can better reflect the practical meaning than PNP distance or PFP distance.

The heightened buildings (HB), demolished buildings (DB) and unchanged buildings (UB) are detected based on VPP distance measure. Then *unchanged terrain* (UT) is detected based on the PFP distance for each terrain point in the ALS data. If the distance from a terrain point to the fitted DIM plane is less than a threshold, this laser point is classified into UT; otherwise it is uncertain and will be judged in the next step.

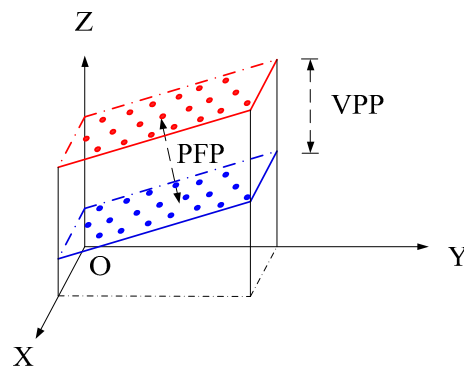


Figure 4. Schematic diagram of VPP distance and PFP distance. The blue segment indicate an old roof in the ALS data. The red segment indicate a new roof in the DIM data. The height of the inclined roof has increased.

The unchanged vegetation (*UV*) is detected by identifying a point as vegetation in both epochs. The previous connected component analysis has detected vegetation in the ALS data. At the vegetation locations, whether the objects are still vegetation is judged with Normalized height nH and normalized vegetation index ($nEGI$). Object-based analysis is more robust in vegetation detection than point-based analysis since the former considers a larger area and less prone to data noise. For each vegetation cluster, whether it is changed or not is determined by the following steps: Firstly, the vegetation points from the same cluster are projected to the ground, and a bounding polygon is constructed with these points in horizontal space. The DIM points within this bounding polygon are applied to verify whether they are vegetation or not. The nH and $nEGI$ are calculated for each DIM point in this polygon and then averaged. If nH and $nEGI$ are both larger than their thresholds, the object in the DIM data is classified into vegetation and this is thus unchanged vegetation (*UV*),

$$nEGI = \frac{2G - R - B}{2G + R + B} \quad (7)$$

where R , G and B indicates red, green and blue value of each pixel, respectively.

3.3. Change Detection: DIM \rightarrow ALS

The remaining regions are obtained by taking the complementary set (*CS*) of *UT*, *UV*, *UB*, *HB* and *DB* (see Equation (2)). The *CS* is an uncertain region, where newly-built buildings (*NB*) are detected, based on the object-based analysis with the DIM data as reference. The key is to extract buildings in the remaining regions. The remaining regions contain, not only newly-built buildings, but also other disturbances. These disturbances are mainly caused by natural vegetation growth or building mis-registration errors. When a vegetation canopy is detected in the ALS data, its boundary is usually larger in the DIM data due to growth. Vegetation change detection in the previous step can only detect the overlapping vegetation regions, but the grown regions cannot be detected. Linear building boundaries might remain in the complementary set due to mis-registration errors between the two point clouds.

The appearance of a complementary set will be shown in the experimental result section. The complementary set suggests that false alarms, such as vegetation boundary change or building boundary change form weak response in the *CS* map, but a real new building change forms a strong response in the *CS* map. Therefore, new buildings can be detected on the binary *CS* map based on point-based features and morphological operation. First, for each undetermined pixel on the *CS* map, nH and $nEGI$ are applied to judge whether it is a building pixel or vegetation pixel. The VPP distance is also calculated between the ALS data and DIM data to indicate height change. To make the two features and VPP distance more robust to noise, features and VPP distance are calculated with all the

neighboring points within a circular neighborhood. Pixels on the CS map are excluded if they meet the following criteria:

- (1) If a pixel is classified into vegetation in the new epoch based on nH and $nEGI$, it is impossible to be a newly-built building and thus excluded.
- (2) If the VPP distance is smaller a threshold, the pixel is not changed and excluded.

In this case, quite many irrelevant pixels are excluded from the CS map. Next, the remaining pixels are further processed with morphological operation to extract newly-built buildings.

3.4. Morphological Operation

The remaining CS map contains not only strong response for newly-built buildings, but also false alarms. The false alarms present the following patterns: Small isolated clusters, elongated artefacts along building edges, small holes on the new building masks. A combination of morphological closing and opening is applied to fill holes and eliminate small or elongated artefacts [48]. This proposed workflow is as follows. Firstly, process the binary CS map with morphological closing and opening in sequence. Their thresholds are T_{close} and T_{open} , respectively. Secondly, connect the neighboring pixels in their 8-neighborhood to form a complete changed object. Remove those objects whose length is smaller than T_{length} . T_{length} is determined by the minimum size of the changed buildings we aim to detect. New buildings are detected by morphological operation and disturbances are eliminated.

4. Experiments

4.1. Descriptions of the Experimental Data

The experiments are implemented on two study areas from The Netherlands. The specifications of the study areas are shown in Table 1. The first study area is located in Rotterdam, which is a densely-built port city mainly covered by residential buildings, skyscrapers, vegetation, roads, and waters. The second study area is located in Enschede, which is also a densely-built urban area. Dense image matching in the two study areas are both performed in Pix4Dmapper [49] to obtain point clouds and orthoimage. The ALS point clouds, DIM point clouds and orthoimages for Rotterdam data and Enschede data are visualized in Figures 5 and 6, respectively.

Table 1. Specifications of two study areas.

Study Area	Area (km × km)	ALS				DIM		Orthoimage	
		Year	Density (pts/m ²)	Vertical Accuracy (cm)	Number of Points (million)	Year	Number of Points (million)	GSD (cm)	
Rotterdam	2 × 2	2007	25	±5	45.3	2016	65.1	10	
Enschede	0.5 × 0.5	2007	25	±5	3.5	2011	7.0	10	

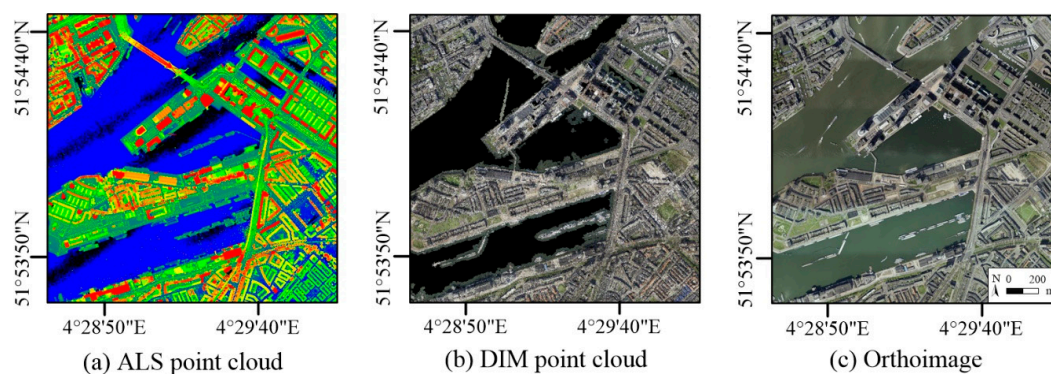


Figure 5. Visualization of the Rotterdam study area. ALS point cloud is colored according to height. DIM point cloud is colored with true color.

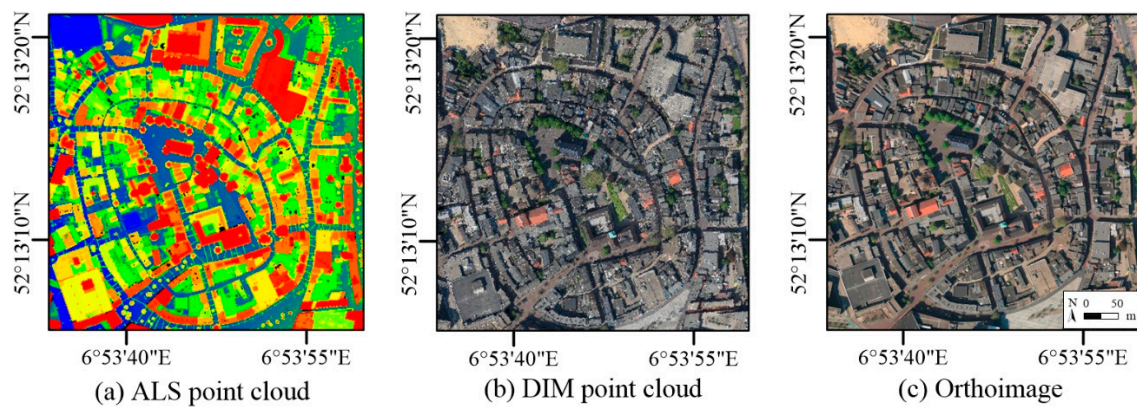


Figure 6. Visualization of the Enschede study area. ALS point cloud is colored according to height. DIM point cloud is colored with true color.

The ALS and DIM data should be registered under the unique coordinate system beforehand. The ALS point cloud was provided under the Dutch national coordinate system (Amersfoort-RD New). The GCP coordinates used in the bundle adjustment were also in the same coordinate system, so the generated DIM point cloud was under the same coordinate system, which guarantees the registration between ALS data and DIM data.

4.2. Experimental Setup

The hyper-parameters listed in Table 2 are set based on the data quality and trial experiments. In the point cloud filtering, since the terrain of two study areas is generally smooth, the threshold d in the Progressive TIN densification is set to 1 m and α is set to 30° . Only those candidate points close to the TIN plane with distance smaller than 1 m and with angle smaller than 30° are classified into terrain. The two thresholds can separate non-ground points and ground points and preserve terrain details.

Table 2. Parameter settings and descriptions.

Step	Threshold	Value	Description
Filtering	d	1 m	Distance from a candidate point to the TIN plane
	α_i	30	Angles between the point-to-TIN-vertex line and the TIN facet, $i \in \{1, 2, 3\}$
Segmentation	D	1 m	Distance from a candidate point to the nearest point on the segment
	D_0	0.2 m	PFP distance from a candidate point to the segment
	N	50	Number of points in a segment
Segment screening	θ	70	Inclination angle of a segment
	nH	3 m	Average normalized height of a segment
	σ	0.2 m	Residual of plane fitting (RPF)
Change detection	dH	3 m	Threshold for building height change, i.e., VPP distance
	$nEGI$	0.1	Threshold for normalized vegetation index
Morphological filtering	T_{close}	50 px	Morphological closing
	T_{open}	50 px	Morphological opening

In the segmentation step, D and D_0 are set based on trial experiments to extract planes. Only when the distance from a candidate point to its closest point on the segment is less than 1 m and its distance to the fitted segment plane is less than 0.2 m, this candidate point belongs to this segment.

Then planar segments are screened based on the following rules: (1) When $N < 50$, this candidate segment is too small and not likely to be a roof segment. (2) when $\theta > 70$, this segment is likely to be a vertical wall or a railing, etc. (3) When $nH < 3$ m, the segment is too close to terrain and not likely to be

a roof. nH indicates the height of the shortest roofs we aim to detect. (4) When the threshold σ for RPF is larger than 0.2, it contains much noise and is not taken as a roof segment.

During change detection, the segments with dH larger than 3 m are taken as roof height change. Considering the DIM data quality, we only detect building changes with height change larger than 3 m. $nEGI$ assists the discrimination between vegetation and non-vegetation in change detection step. When $nEGI$ of a segment or a pixel calculated on the orthoimage is larger than 0.1, it is taken as vegetation.

In morphological filtering, morphological closing and opening are applied to preserve major true positives and eliminate false positives. T_{close} and T_{open} are set based on trial experiments. Considering the DIM data quality, we aim to detect building changes with side lengths greater than 5 m in real scenarios, which is equivalent to 50 pixels on the orthoimages.

In addition, the proposed method is compared with surface differencing as a baseline method. Surface Differencing is a classic change detection method for point cloud-based change detection, which is also used as a baseline method in the previous literature [40–42]. Firstly, convert two point clouds into DSMs and subtract one DSM from the other. Then, apply the similar morphological operation from our method to post-process the heightened map and lowered map separately. The locations where the height difference exceeds 3 m is determined as a candidate building change. The morphological operation starts with closing and then performs opening. Then, a small connected change masks, with a length smaller than 100 px (i.e., 10 m) are eliminated. Finally, the heightened and lowered masks are merged into final change map. Note that the thresholds for morphological operation in surface differencing is coarser than those used in the proposed method. Our trial experiments show that this setting brings a proper balance between true positives and false positives.

4.3. Evaluation Metrics

The results are evaluated qualitatively and quantitatively. The two missions of building extraction and change detection results are evaluated separately. Our final building footprints and change maps are both 2D products. The ground truth (GT) is prepared by careful visual inspection on the point cloud differencing map aided by the point clouds and orthoimages. Three evaluation measures applied in this paper are taken from the ISPRS benchmark on urban object detection [11]: Recall, Precision, and F_1 -score. Recall indicates the ability of a model to detect all the real changes. Precision indicates the ratio of true changes among all the detected changes. The F_1 -score is a metric to combine recall and precision using their harmonic mean.

$$Recall = TP / (TP + FN) \quad (8)$$

$$Precision = TP / (TP + FP) \quad (9)$$

$$F_1 = 2 \cdot (Recall \cdot Precision) / (Recall + Precision) \quad (10)$$

For example, considering the pixel-based change detection evaluation, True positive (TP) is the number of changed pixels detected correctly. True negative (TN) is the number of unchanged pixels detected as unchanged. False positive (FP) is the number of pixels detected by the algorithm, which are not changes in the real scene. False negative (FN) is the number of undetected changes.

In addition, the change detection results are also evaluated at the object level. For the evaluation of building extraction, the results are evaluated only at the pixel level but not at the object level. Evaluation of building extraction in the object-level is hard because many buildings in our two study areas are mainly closely adjacent and the boundaries between individual buildings are hard to recognize.

5. Results and Discussion

5.1. Results and Discussion of Rotterdam Data

Qualitative results: The results of Rotterdam data from the intermediate steps are shown in Figure 7. From the ALS point cloud, progressive TIN densification is performed, and surface-based

segmentation is performed to generate 3958 segments. After segment screening, 819 roof segments are valid as shown in Figure 7a. Some roofs are represented by a complete segment, while some roofs are broken into several sub-segments. The VPP distance is calculated for each valid segment. Segments with height change larger than 3 m are considered as changed buildings as shown in Figure 7b. The binary change masks after morphological processing are shown in Figure 7c, which include two types of changes: eliminated and heightened buildings.

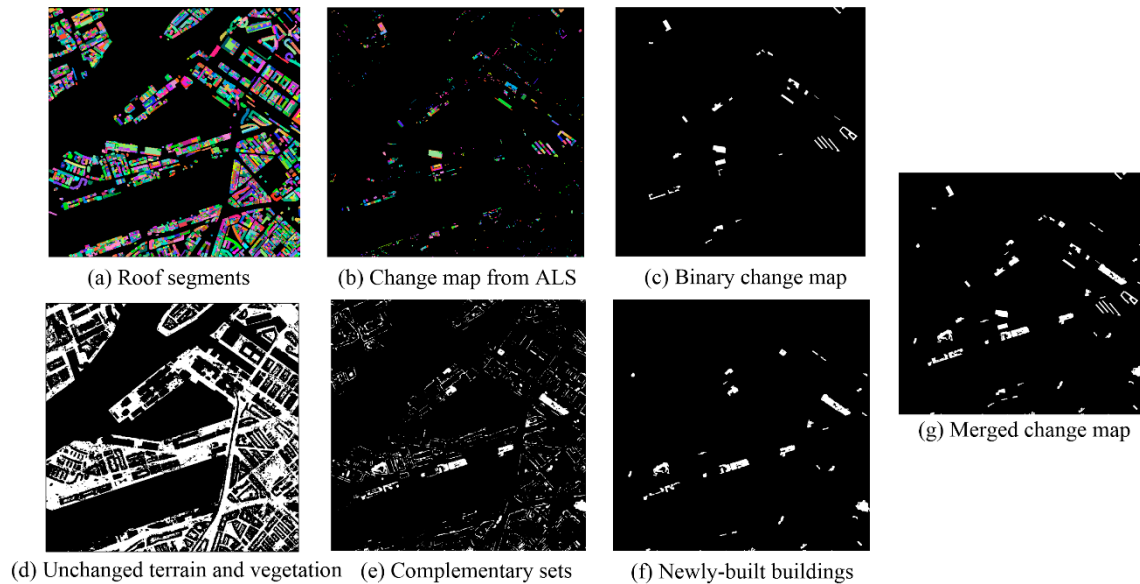


Figure 7. Intermediate products for Rotterdam data.

Figure 7d contains the regions of unchanged terrain (UT) and unchanged vegetation (UV). According to Formula (2), when the ALS roof regions are further excluded, the complementary set CS is obtained as shown in Figure 7e, where newly-built buildings show strong response. The morphological operation is applied to extract new buildings, as shown in Figure 7f. Finally, the bidirectional change detection results are merged into Figure 7g, which contains three types of building changes: Heightened, demolished and newly-built buildings.

Figure 7g is further processed by morphological closing and then opening to eliminate small false positives and fill gaps. The final results are shown in Figure 8. Figure 8a is the ground truth for integrated building extraction and change detection. Yellow indicates building masks; Magenta indicates heightened buildings (incl. newly-built buildings); Cyan indicates demolished buildings. Figure 8b are the results of our method. Figure 8c is the visualization of the errors in building extraction. Red indicates false positives; Blue indicates false negatives. In order to visualize the change detection results separately, Figure 8d shows the ground truth for change detection. Figure 8e is the change detection result from our method. Figure 8f is the result from surface differencing.

Figure 8c shows that the proposed method can successfully extract most building footprints with a few FPs and FNs. Comparing Figure 8d,e shows that most building changes are detected successfully although some FPs appear. It is clear that the detected demolished buildings show sharp boundaries, while some heightened buildings show fuzzy boundaries. The reason is that the boundaries of demolished buildings are determined from the precise ALS data, while the boundaries of newly-built buildings are determined from the relatively noisy DIM data. In contrast, Figure 8f shows that surface differencing brings much more FPs than our method, such as along building edges, on the vegetation surface or shadow.

Six examples of building extraction are visualized in Figure 9. Figure 9a shows that inclined roofs with complicated roof structures are correctly detected. Even though surface-based segmentation breaks the planes into small broken segments, the roof segments are merged into complete roof masks

during morphological operation. Figure 9b,c show two bridges incorrectly detected as roofs. This error is caused by mistakes in point cloud filtering. The bridge is mis-classified into non-ground points and thus remain in the following steps. Similarly, Figure 9d shows some containers or sheds mis-classified into buildings. Figure 9e shows some hedges or box-like structures mis-classified into buildings. Figure 9f shows that false negatives occur on a wavelike roofs. The roof segments are small and contain data gaps, so they are eliminated in the segment screening, which leads to omission errors.

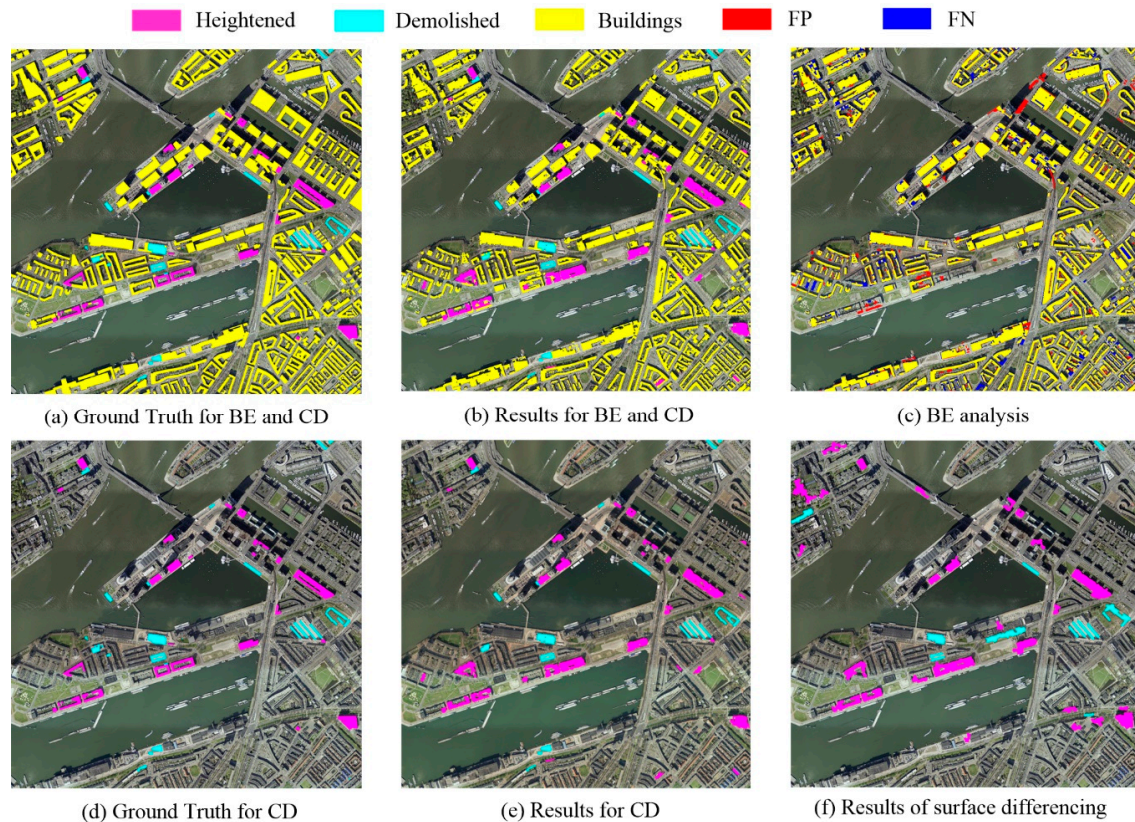


Figure 8. Experimental results of Rotterdam data. (BE: building extraction; CD: change detection).

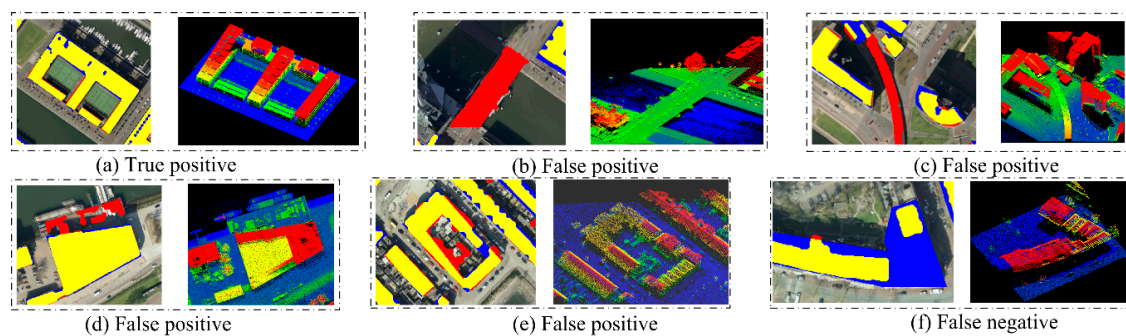


Figure 9. Six examples of the building detection results for Rotterdam data.

Four examples of change detection results are shown in Figure 10. Each example from the left to the right shows the ALS point cloud, DIM point cloud, result from our method and result from surface differencing. Figure 10a shows a demolished building-group. Our method detects the change with sharp boundaries while surface differencing takes the neighboring vegetation changes as building changes and also omits one building change. Figure 10b shows that both our method and surface differencing can detect an independent demolished building. Figure 10c shows a building that is partly heightened and partly demolished. Our method can detect the complicated changes correctly,

while surface differencing omits the demolished building. The area of demolished building is rather small, and thus, eliminated in surface differencing. Figure 10d shows both methods bring FPs in a courtyard. In this shaded region, DIM point cloud is noisy and its height is usually deviated from the true height, so FPs are more likely to appear.

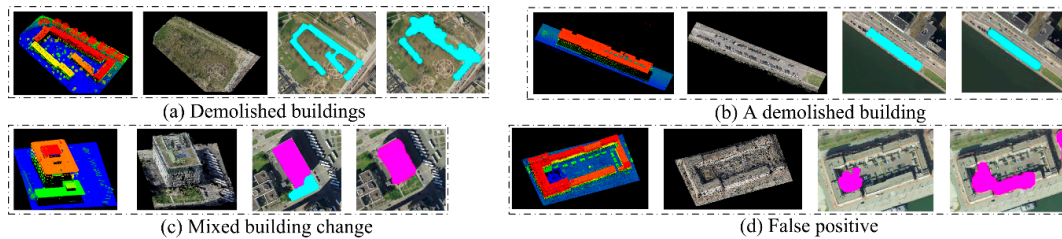


Figure 10. Four examples of the change detection results for Rotterdam data.

Quantitative results: Our building extraction results are evaluated at the pixel level. The method achieves precision of 91.94%, recall of 82.64% and F_1 -score of 87.04%. Although we do not have comparative methods for building extraction, a glimpse at [11] can still give us hints on the performance of our method. The ISPRS benchmark on urban object detection [11] reports that the high-ranking building extraction methods can achieve F_1 -score of 89.8% and 88.9%, depending on the data quality and applied method. Additionally, multispectral features are also available to them. Without multispectral features, we achieve F_1 -score of 87.04% with merely geometric features, which is relatively satisfactory.

The change detection results are evaluated at the pixel level and object level as shown in Tables 3 and 4, respectively. In Table 3, the recalls for heightened buildings and demolished buildings from our method are both above 85%, indicating that most of the two types of changes can be detected successfully. The precision of the heightened buildings is 69.26%, which is much lower than the 95.20% of the demolished buildings. As explained before, the demolished buildings are determined on the ALS data which are more precise, while most of the heightened buildings are determined on the DIM data, which are noisier.

Table 3. Pixel-based change detection results of Rotterdam data (%).

Method	Types	Recall	Precision	F_1 -Score
Proposed method	Heightened	85.94	69.26	76.70
	Demolished	86.88	95.20	90.85
	Overall	86.27	75.70	80.64
Surface differencing	Heightened	88.21	47.10	61.41
	Demolished	58.40	48.52	53.00
	Overall	78.47	47.47	59.16

Table 4. Object-based change detection results of Rotterdam data.

Method	Types	TP	Detected Changes	Detected True Changes	Recall/(%)	Precision/(%)
Proposed method	Heightened	27	31	26	96.30	83.87
	Demolished	25	23	22	88.00	95.65
	Overall	52	54	48	92.31	88.89
Surface differencing	Heightened	27	29	23	85.19	79.31
	Demolished	25	18	13	52.00	72.22
	Overall	52	47	36	69.23	76.60

Considering the results of surface differencing in Table 3, the recall of heightened buildings are much higher than that of demolished buildings. Figure 8f shows that surface differencing brings many FPs in vegetation and shadow, where DIM point clouds are noisier and higher than ALS data due to dense matching errors or natural growth. Then, surface differencing tends to over-classify these

possible regions as heightened buildings. The precisions of heightened and demolished buildings are lower than 61.41% and 53.00%, which are much lower than the proposed method.

Table 4 shows the object-based evaluation results. Since the areas and shapes of building are versatile, it shows that the object-based evaluation measures are not propagational with the pixel-based measures. There are 27 heightened buildings and 25 demolished buildings in total. Our method misses one change and brings five FPs regarding to heightened buildings. It misses the three demolished buildings and brings one FPs in terms of demolished buildings. False alarms are more likely to occur on heightened buildings. In these areas, the height of DIM data are often higher than the true object height due to dense matching errors, so the method incorrectly detects them as new buildings. In addition, Figure 8e shows that most of the missed buildings are small buildings. The small buildings might be omitted in the step of object extraction or morphological operation. Considering the object-based evaluation of surface differencing, the recall and precision of surface differencing for heightened, demolished and overall are both lower than the measures of the proposed method. The recall of surface differencing is lower than the proposed method by 23.08%, and the precision is lower than the proposed method by 12.29%. Specifically, the recall of demolished buildings is only 52%, indicating that nearly half of demolished buildings are missed. Figure 8f shows that surface differencing also misses many small building changes. Compared to our object-based change detection, surface difference is more vulnerable to noise.

5.2. Results and Discussion of Enschede Data

Qualitative results: The intermediate results of Enschede data are shown in Figure 11. Surface-based segmentation is performed on the ALS data to generate 1866 segments. After segment screening, 128 valid roof segments are shown in Figure 11a. Some roofs are represented by a complete segment, while some roofs are broken into several sub-segments. Segments with height change larger than 3 m are considered as changed buildings as shown in Figure 11b. The binary change masks after morphological processing are shown in Figure 11c. Figure 11d shows the union of unchanged terrain (*UT*) and unchanged vegetation (*UV*). When the ALS roof regions are excluded, the complementary set (*CS*) is obtained as shown in Figure 11e, where newly-built buildings show strong response. Morphological operation is applied to extract new buildings as shown in Figure 7f. Finally, the bidirectional change detection masks are merged into the complete change map as shown in Figure 7g.

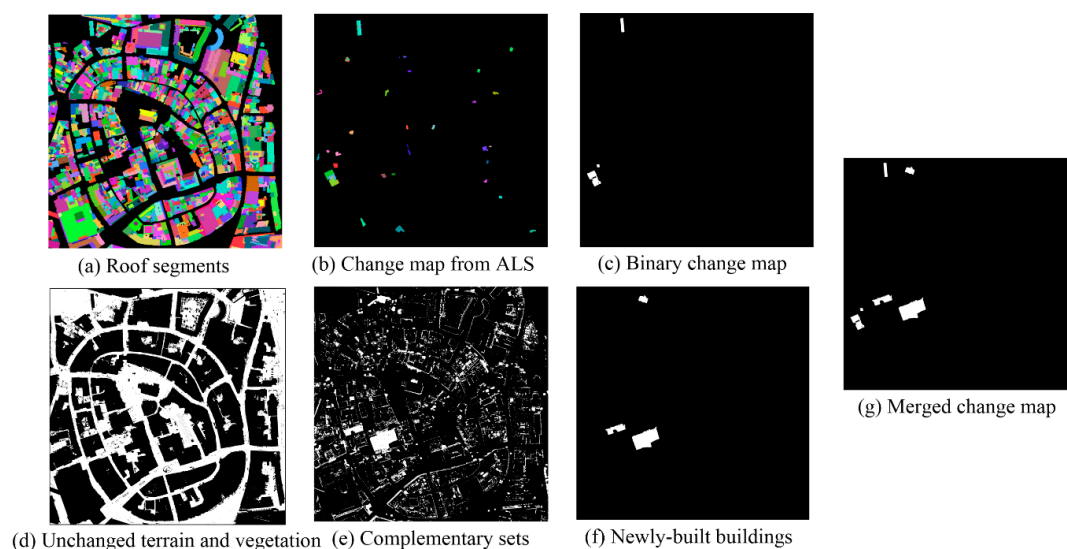


Figure 11. Intermediate products of the method for Enschede data.

The final results are shown in Figure 12. For the definition of each sub-figure and color coding, the readers are referred to Figure 8. Figure 12c shows that the proposed method can successfully extract most building footprints for the Enschede data. A few FPs and FNs lie along the building boundaries or inside the narrow courtyard. Considering change detection, the ALS data and DIM were acquired with an interval of four years, and only four changes are heightened and three buildings are demolished. However, Figure 12b,e still demonstrate that the proposed method is effective in change detection and robust to data noise. In contrast, Figure 12f shows that surface differencing brings more FPs than our method.

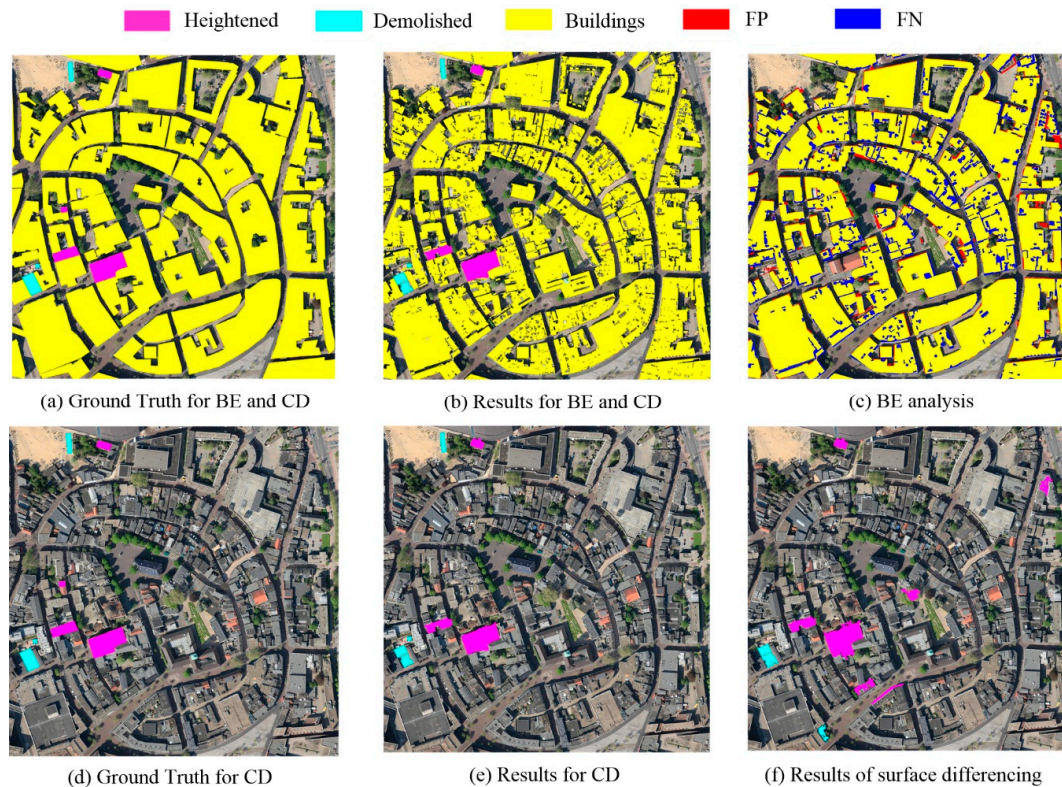


Figure 12. Experimental result of Enschede data. (BE: building extraction; CD: change detection).

Six examples of building extraction are visualized in Figure 13. Figure 13a shows that the footprints of a church are correctly extracted, except some elongated FPs and FNs along the building boundaries. Note that some pixels in the middle of the building footprint are missed due to data gaps in the raw ALS point cloud. Figure 13b shows FPs and FNs along the building boundaries. The ground truth for building footprints is delineated on the orthoimage. These might be some mis-registration errors between ALS data and DIM data. Figure 13c shows a shed with a height of 1.7 m between two buildings. This short and small segment is eliminated in segment screening step which leads to a FP. Figure 13d shows that a truck with a height of 3.3 m is mistaken as a building. There is another truck in the same place in the DIM data, so it is taken as an unchanged building instead of a changed building. Figure 13e shows another church with steep roofs. Previously, Figure 11a shows that the steep roofs are already eliminated in the segment-based screening since the roof segments are broken and small. Figure 13f shows that some pixels on a roof is missed because the data are missing in the ALS point cloud.

Four examples of change detection results for Enschede data are shown in Figure 14. Figure 14a shows a demolished building. Our method detects the change with sharp boundaries, while surface differencing misses this changed building. Figure 14b shows two demolished buildings. Our method detects both changes while surface differencing misses one changed building. Surface differencing

omits one building because this building is small and eliminated in morphological opening. Figure 14c shows that FPs appear in the narrow courtyard in the surface differencing result when the ALS data contain many data gaps. Figure 14d shows that surface differencing causes FPs in the shaded region adjacent to a wall. The FPs in Figure 10d are heightened buildings, while the FPs here are demolished buildings. This can be explained by that dense matching fails in this area covered with shadow and vegetation so the height represented by the DIM data is lower than the true height. This results into a false demolished building.

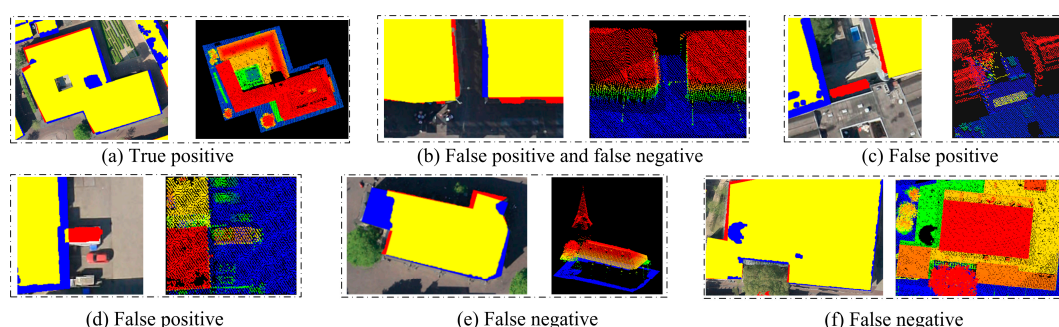


Figure 13. Six examples of the building detection results for Enschede data.

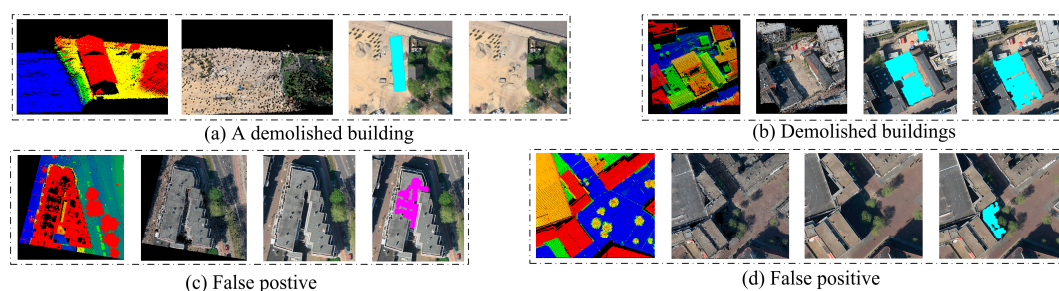


Figure 14. Four examples of the change detection results for Enschede data.

Quantitative results: The building extraction results of Enschede data are further evaluated at the pixel level. The precision reaches 96.28%, recall of 86.63% and F_1 -score of 91.20%. The precision of building extraction for Enschede data is higher than that of Rotterdam data by 4.34%. The recall for Enschede data is higher than that of Rotterdam data by 3.99%. Comparing the error maps in Figures 8c and 12c for the two study areas shows that more building footprints are missed in the Rotterdam data such as small wavelike roofs with complicated shapes. This leads to a lower recall rate in the Rotterdam data. Considering the precision, some bridges, containers or trucks are mistaken as buildings, which results into a relatively low precision. In contrast, although there are also some FPs and FNs in the Enschede results, most of the errors are small dots or elongated lines along the building boundaries. Therefore, the measures of building extraction evaluation for Enschede data are better than those for Rotterdam data.

The change detection results are evaluated at the pixel level as shown in Table 5. In Table 5, the recalls for heightened buildings and demolished buildings from our method are both above 85%, indicating that most of the two types of changes can be detected successfully. The precision of the heightened buildings is 69.26%, which is much lower than 95.20% for the demolished buildings. As explained before, the demolished buildings are determined on the ALS data which are more precise while the newly-built buildings are determined on the DIM data, which are noisier. Considering the results of surface differencing in Table 5, the recall of heightened buildings are much higher than that of demolished buildings. Figure 8f shows that surface differencing brings many FPs in vegetation and shadow, where DIM point clouds are noisy and higher than the ALS data due to dense matching errors or natural growth. Then, surface differencing tends to over-classify these plausible regions as

heightened buildings. Therefore, the F_1 -score of heightened and demolished buildings are much lower than the proposed method.

Table 5. Pixel-based change detection results of Enschede data (%).

Method	Types	Recall	Precision	F_1 -Score
Proposed method	Heightened	86.30	88.77	87.52
	Demolished	82.38	96.90	89.05
	Overall	85.30	90.65	87.89
Surface differencing	Heightened	88.80	56.98	69.42
	Demolished	61.36	70.43	65.58
	Overall	81.79	59.15	68.65

Table 6 shows the object-based evaluation results of change detection. There are four heightened buildings and three demolished buildings in total. Our method misses one small heightened building and successfully detects all the demolished buildings. Surface differencing misses one heightened building and two demolished buildings; It also brings four heightened FPs and one demolished FP. Figure 12f shows that the FN errors mainly appear on small building changes, while FPs usually occur on the vegetation canopy, in the shadow or along the building boundaries. The recall achieved by surface differencing is 57.14% and the precision is only 44.44%, which are much lower than the measures from our method. Our method is more effective and robust to noise than surface difference.

Table 6. Object-based change detection results of Enschede data.

Method	Types	TP	Detected Changes	Detected True Changes	Recall/(%)	Precision/(%)
Proposed method	Heightened	4	3	3	75.00	100
	Demolished	3	3	3	100	100
	Overall	7	6	6	85.71	100
Surface differencing	Heightened	4	7	3	75.00	42.86
	Demolished	3	2	1	33.33	50.00
	Overall	7	9	4	57.14	44.44

5.3. Discussion on Data Quality and Error Sources

Comparing the building extraction results in the two study areas shows that both of the precision and recall of Enschede data are better than the measures of Rotterdam data. Even though more data gaps exist in the Enschede data, compared with the Rotterdam data (see Figures 5a and 6a), the effect of data gaps can be remedied by morphological operation. However, the scene complexity of Rotterdam data is higher than that of Enschede data concerning the presence of bridges, containers and trucks. Our segment-based building extraction method tends to misclassify these square objects into buildings and omit buildings with small or wavelike roofs.

In our method, building extraction and change detection are performed in sequence so the errors of building extraction are propagated to the change detection. When building extraction is complete and correct, the fine building footprints from ALS data can serve as a good reference to implement “ALS->DIM” change detection. The example in Figure 10c shows that when a demolished building is missed in building extraction, this error will finally result into an omission error in change detection.

Concerning the errors in the change maps as shown in Figures 8e and 12e, FPs often appear on the containers, trucks, vegetation, etc. These objects are often mistaken as buildings in building extraction and thus the errors remain in the final change map. FNs often appear when the roofs are small or broken. When large data gaps exist on the changed buildings in the ALS point clouds, FNs may also occur in the change map (e.g., Figure 13f). In shaded regions, such as a narrow road or narrow courtyard, dense matching may fail with no point cloud or generate inaccurate points whose heights

are higher than true values. Therefore, both FPs and FNs may appear on the shaded regions. If a FN happens, our results show that this FN can be a heightened building or a lowered building, depending on whether dense matching failed or is inaccurate.

Comparing our object-based bidirectional method with surface differencing, our method performs better since the change detection is implemented based on each building object instead of individual pixels, which allows to take more contextual information into consideration. Du et al. [39] detect building changes between DIM data and ALS data, and its recall and accuracy rates are 93.50% and 92.15%, respectively. Although their evaluation measures are slightly higher than ours, their workflow is more complicated involving energy minimization and more human intervention. In addition, the outputs of our method contain not only change maps but also building footprints.

In addition, the quality of DIM data affects the change detection results. (1) Concerning the vertical accuracy of DIM data, the vertical accuracy of DIM data determines the minimum height change our method could detect. (2) Concerning the horizontal accuracy of DIM data, the inaccuracy in the DIM data affects the registration between ALS data and DIM data. The mis-registration results into some linear artefacts along building edges during change detection. (3) Concerning the DIM point density, sparse DIM point clouds make accurate localization of building edges impossible, which thus affect the localization of building changes. (4) Concerning the precision of DIM data, the noise in the point clouds hinders accurate representation of the objects. It causes many false positives and false negatives in the final change detection results.

6. Conclusions

We propose an object-based bidirectional method for integrated building extraction and change detection. Change detection between the heterogeneous ALS data and DIM data is vulnerable to dense matching errors, mis-alignment errors and data gaps. Our method starts from object extraction on the precise ALS data. Firstly, progressive TIN densification, surface-based growing, segment screening and connected component analysis are used to extract terrain, roofs and vegetation. Secondly, change detection is performed in a bidirectional manner; heightened buildings and demolished buildings are detected by taking the ALS data as reference, while newly-built buildings are detected by taking the DIM data as reference in the complementary sets. Vertical Plane-to-Plane Distance is proposed as the indicator of height change. Experiments are implemented on two data sets and the results are evaluated at both pixel level and object level. In the object-based evaluation, the recall reaches 92.31% and the precision reaches 88.89% for the Rotterdam data. The recall reaches 85.71% and the precision reaches 100% for the Enschede data.

There are two advantages within our method: Firstly, the building extraction and change detection are achieved in one workflow, and the results are visualized in one merged footprint-and-change map. By means of proper fusion of geometric and spectral features, the method classifies the changed buildings into heightened or lowered with relatively sharp boundaries. Secondly, the method is object-based. Since change detection is performed on the whole roof segment instead of point-to-point comparison, it is relatively robust to point cloud noise and data gaps.

However, our method also presents some limitations. The change detection results are largely dependent on the quality and resolution of the input data. When the data quality and resolution deteriorate, not only the change detection accuracy decreases, but also the sizes of detected buildings decrease. Additionally, the errors in filtering and building extraction are propagated to the final change detection results. The false positives and false negatives from our method are mainly caused by the dense matching errors on the vegetation or shadow. When the accuracy of object recognition from the DIM data improves, the change detection results also improve. Considering future work, spectral and textural features, derived from the original multi-view imagery, might be used to improve the accuracy of object recognition from the DIM data. Change detection might be conducted between ALS data and multi-view imagery directly, instead of two types of point clouds. The proposed method will be validated on other data sets from various urban scenes.

Author Contributions: C.D. and Z.Z. designed and implemented the experiments. D.L. contributed to the analysis and interpretation of the results. Z.Z. and D.L. wrote the manuscript. C.D. conceived of the paper and revised the manuscript. All authors read and approved the final manuscript.

Funding: This research was funded by the China Scholarship Council and National Natural Science Foundation of China (Grant No. 41501482), which are gratefully acknowledged.

Acknowledgments: Prof. George Vosselman and Dr. Michael Ying Yang from the Faculty of Geo-Information Science and Earth Observation (ITC), University of Twente provided much enlightenment and support for this paper, which are gratefully acknowledged. The authors also gratefully acknowledge CycloMedia for the Rotterdam data set.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Tran, T.H.G.; Ressel, C.; Pfeifer, N. Integrated change detection and classification in urban areas based on airborne laser scanning point clouds. *Sensors* **2018**, *18*, 448. [[CrossRef](#)] [[PubMed](#)]
2. Qin, R.; Tian, J.; Reinartz, P. 3D change detection—approaches and applications. *ISPRS J. Photogramm. Remote Sens.* **2016**, *122*, 41–56. [[CrossRef](#)]
3. Remondino, F.; Spera, M.G.; Nocerino, E.; Menna, F.; Nex, F. State of the art in high density image matching. *Photogramm. Rec.* **2014**, *29*, 144–166. [[CrossRef](#)]
4. Mandlbürger, G.; Wenzel, K.; Spitzer, A.; Haala, N.; Glira, P.; Pfeifer, N. Improved topographic models via concurrent airborne lidar and dense image matching. *ISPRS Ann. Photogram. Remote Sens. Spat. Inf. Sci.* **2017**, *IV-2/W4*, 259–266. [[CrossRef](#)]
5. Zhang, Z.; Gerke, M.; Vosselman, G.; Yang, M.Y. A patch-based method for the evaluation of dense image matching quality. *Int. J. Appl. Earth Obs. Geoinf.* **2018**, *70*, 25–34. [[CrossRef](#)]
6. Ressel, C.; Brockmann, H.; Mandlbürger, G.; Pfeifer, N. Dense image matching vs. airborne laser scanning—comparison of two methods for deriving terrain models. *Photogramm.—Fernerkund.—Geoinf.* **2016**, *2*, 57–73. [[CrossRef](#)]
7. Hirschmüller, H. Stereo processing by semi-global matching and mutual information. *IEEE Trans. Pattern Anal. Mach. Intell.* **2008**, *30*, 328–341. [[CrossRef](#)]
8. Rothmel, M.; Wenzel, K.; Fritsch, D.; Haala, N. SURE: Photogrammetric surface reconstruction from imagery. In Proceedings of the LC3D Workshop, Berlin, Germany, 4–5 December 2012; Volume 8, pp. 1–7.
9. Matikainen, L.; Hyypä, J.; Kaartinen, H. Automatic detection of changes from laser scanner and aerial image data for updating building maps. *Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci.* **2004**, *35*, 434–439.
10. Qin, R.; Gruen, A. 3D change detection at street level using mobile laser scanning point clouds and terrestrial images. *ISPRS J. Photogramm. Remote Sens.* **2014**, *90*, 23–35. [[CrossRef](#)]
11. Rottensteiner, F.; Sohn, G.; Gerke, M.; Wegner, J.D.; Breikopf, U.; Jung, J. Results of the ISPRS benchmark on urban object detection and 3D building reconstruction. *ISPRS J. Photogramm. Remote Sens.* **2014**, *93*, 256–271. [[CrossRef](#)]
12. Vosselman, G.; Gorte, B.G.H.; Sithole, G. Change detection for updating medium scale maps using laser altimetry. *Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci.* **2004**, *34*, 207–212.
13. Vosselman, G. Point cloud segmentation for urban scene classification. *Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci.* **2013**, *7*, W2. [[CrossRef](#)]
14. Xie, Y.; Tian, J.; Zhu, X.X. A review of point cloud semantic segmentation. *arXiv* **2019**, arXiv:1908.08854.
15. Rabbani, T.; Van Den Heuvel, F.; Vosselmann, G. Segmentation of point clouds using smoothness constraint. *Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci.* **2006**, *36*, 248–253.
16. Brodu, N.; Lague, D. 3D terrestrial lidar data classification of complex natural scenes using a multi-scale dimensionality criterion: Applications in geomorphology. *ISPRS J. Photogramm. Remote Sens.* **2012**, *68*, 121–134. [[CrossRef](#)]
17. Le, V.H.; Vu, H.; Nguyen, T.T.; Le, T.L.; Tran, T.H. Acquiring qualified samples for RANSAC using geometrical constraints. *Pattern Recognit. Lett.* **2018**, *102*, 58–66. [[CrossRef](#)]
18. Axelsson, P. DEM generation from laser scanner data using adaptive TIN models. *Int. Arch. Photogramm. Remote Sens.* **2000**, *4*, 110–117.

19. Guo, L.; Chehata, N.; Mallet, C.; Boukir, S. Relevance of airborne LiDAR and multispectral image data for urban scene classification using Random Forests. *ISPRS J. Photogramm. Remote Sens.* **2011**, *66*, 56–66. [[CrossRef](#)]
20. Weinmann, M.; Jutzi, B.; Hinz, S.; Mallet, C. Semantic point cloud interpretation based on optimal neighborhoods, relevant features and efficient classifiers. *ISPRS J. Photogramm. Remote Sens.* **2015**, *105*, 286–304. [[CrossRef](#)]
21. Hackel, T.; Wegner, J.D.; Schindler, K. Fast semantic segmentation of 3D point clouds with strongly varying density. *ISPRS Ann. Photogramm. Remote Sens. Spat. Inf. Sci.* **2016**, *3*, 177–184. [[CrossRef](#)]
22. Xu, S.; Pan, X.; Li, E.; Wu, B.; Bu, S.; Dong, W.; Xiang, S.; Zhang, X. Automatic building rooftop extraction from aerial images via hierarchical RGB-D priors. *IEEE Trans. Geosci. Remote Sens.* **2018**, *56*, 7369–7387. [[CrossRef](#)]
23. Niemeyer, J.; Rottensteiner, F.; Soergel, U. Contextual classification of LiDAR data and building object detection in urban areas. *ISPRS J. Photogramm. Remote Sens.* **2014**, *1*, 152–165. [[CrossRef](#)]
24. Vosselman, G.; Coenen, M.; Rottensteiner, F. Contextual segment-based classification of airborne laser scanner data. *ISPRS J. Photogramm. Remote Sens.* **2017**, *128*, 354–371. [[CrossRef](#)]
25. Su, H.; Maji, S.; Kalogerakis, E.; Learned-Miller, E. Multi-view convolutional neural networks for 3d shape recognition. In Proceedings of the IEEE International Conference on Computer Vision, Santiago, Chile, 11–18 December 2015; pp. 945–953.
26. Zhang, L.; Sun, J.; Zheng, Q. 3D Point Cloud Recognition Based on a Multi-View Convolutional Neural Network. *Sensors* **2018**, *18*, 3681. [[CrossRef](#)]
27. Zhou, Y.; Tuzel, O. Voxelnet: End-to-end learning for point cloud based 3d object detection. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–23 June 2018; pp. 4490–4499.
28. Hu, X.; Yuan, Y. Deep-learning-based classification for DTM extraction from ALS point cloud. *Remote Sens.* **2016**, *8*, 730. [[CrossRef](#)]
29. Zhang, Z.; Vosselman, G.; Gerke, M.; Persello, C.; Tuia, D.; Yang, M.Y. Detecting Building Changes between Airborne Laser Scanning and Photogrammetric Data. *Remote Sens.* **2019**, *11*, 2417. [[CrossRef](#)]
30. Winiwarter, L.; Mandlbürger, G.; Schmohl, S.; Pfeifer, N. Classification of ALS Point Clouds Using End-to-End Deep Learning. *PFG—J. Photogramm. Remote Sens. Geoinf. Sci.* **2019**, *87*, 75–90. [[CrossRef](#)]
31. Yousefhusien, M.; Kelbe, D.J.; Ientilucci, E.J.; Salvaggio, C. A multi-scale fully convolutional network for semantic labeling of 3D point clouds. *ISPRS J. Photogramm. Remote Sens.* **2018**, *143*, 191–204. [[CrossRef](#)]
32. Qi, C.R.; Su, H.; Mo, K.; Guibas, L.J. PointNet: Deep learning on point sets for 3d classification and segmentation. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017; pp. 652–660.
33. Qi, C.R.; Yi, L.; Su, H.; Guibas, L.J. PointNet++: Deep hierarchical feature learning on point sets in a metric space. In Proceedings of the Advances in Neural Information Processing Systems, Long Beach, CA, USA, 4–9 December 2017; pp. 5099–5108.
34. Singh, A. Digital change detection techniques using remotely-sensed data. *Int. J. Remote Sens.* **1989**, *10*, 989–1003. [[CrossRef](#)]
35. Matikainen, L.; Karila, K.; Hyypä, J.; Puttonen, E.; Ahokas, E. Feasibility of multispectral airborne laser scanning for land cover classification, road mapping and map updating. *Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci.* **2017**, *3/W3*, 119–122. [[CrossRef](#)]
36. Zhan, Y.; Fu, K.; Yan, M.; Sun, X.; Wang, H.; Qiu, X. Change detection based on deep Siamese Convolutional Network for optical aerial images. *IEEE Geosci. Remote Sens. Lett.* **2017**, *14*, 1845–1849. [[CrossRef](#)]
37. Volpi, M.; Camps-Valls, G.; Tuia, D. Spectral alignment of multi-temporal cross-sensor images with automated kernel canonical correlation analysis. *ISPRS J. Photogramm. Remote Sens.* **2015**, *107*, 50–63. [[CrossRef](#)]
38. Gong, M.; Zhan, T.; Zhang, P.; Miao, Q. Superpixel-based difference representation learning for change detection in multispectral remote sensing images. *IEEE Trans. Geosci. Remote Sens.* **2017**, *5*, 2658–2673. [[CrossRef](#)]
39. Wu, C.; Du, B.; Cui, X.; Zhang, L. A post-classification change detection method based on iterative slow feature analysis and Bayesian soft fusion. *Remote Sens. Environ.* **2017**, *199*, 241–255. [[CrossRef](#)]
40. Tian, J.; Cui, S.; Reinartz, P. Building change detection based on satellite stereo imagery and digital surface models. *IEEE Trans. Geosci. Remote Sens.* **2014**, *52*, 406–417. [[CrossRef](#)]

41. Qin, R. An object-based hierarchical method for change detection using unmanned aerial vehicle images. *Remote Sens.* **2014**, *6*, 7911–7932. [[CrossRef](#)]
42. Xu, S.; Vosselman, G.; Oude Elberink, S. Detection and classification of changes in buildings from airborne laser scanning data. *Remote Sens.* **2015**, *7*, 17051–17076. [[CrossRef](#)]
43. Du, S.; Zhang, Y.; Qin, R.; Yang, Z.; Zou, Z.; Tang, Y.; Fan, C. Building change detection using old aerial images and new LiDAR data. *Remote Sens.* **2016**, *8*, 1030. [[CrossRef](#)]
44. Zhou, K.; Lindenbergh, R.C.; Gorte, B.G.; Zlatanova, S. LiDAR-guided dense matching for detecting changes and updating of buildings in Airborne LiDAR data. *ISPRS J. Photogramm. Remote Sens.* **2020**, *162*, 200–213. [[CrossRef](#)]
45. Mou, L.; Bruzzone, L.; Zhu, X.X. Learning spectral-spatial-temporal features via a recurrent convolutional neural network for change detection in multispectral imagery. *IEEE Trans. Geosci. Remote Sens.* **2019**, *57*, 924–935. [[CrossRef](#)]
46. Daudt, R.C.; Le Saux, B.; Boulch, A. Fully convolutional siamese networks for change detection. In Proceedings of the IEEE International Conference on Image Processing (ICIP), Athens, Greece, 7–10 October 2018; pp. 4063–4067.
47. Dong, Z.; Liang, F.; Yang, B.; Xu, Y.; Zang, Y.; Li, J.; Wang, Y.; Dai, W.; Fan, H.; Hyypä, J.; et al. Registration of large-scale terrestrial laser scanner point clouds: A review and benchmark. *ISPRS J. Photogramm. Remote Sens.* **2020**, *163*, 327–342. [[CrossRef](#)]
48. Haralick, R.M.; Shapiro, L.G. *Computer and Robot Vision*; Addison-Wesley: Boston, MA, USA, 1992.
49. Pix4d. Available online: <https://pix4d.com/> (accessed on 24 May 2020).



© 2020 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).