

Journal of Applied Remote Sensing

RemoteSensing.SPIEDigitalLibrary.org

Building damage assessment using airborne lidar

Colin Axel
Jan van Aardt

SPIE.

Colin Axel, Jan van Aardt, "Building damage assessment using airborne lidar," *J. Appl. Remote Sens.* 11(4), 046024 (2017), doi: 10.1117/1.JRS.11.046024.

Building damage assessment using airborne lidar

Colin Axel* and Jan van Aardt

Rochester Institute of Technology, Chester F. Carlson Center for Imaging Science,
Digital Imaging and Remote Sensing Laboratory, Rochester, New York, United States

Abstract. The assessment of building damage following a natural disaster is a crucial step in determining the impact of the event itself and gauging reconstruction needs. Automatic methods for deriving damage maps from remotely sensed data are preferred, since they are regarded as being rapid and objective. We propose an algorithm for performing unsupervised building segmentation and damage assessment using airborne light detection and ranging (lidar) data. Local surface properties, including normal vectors and curvature, were used along with region growing to segment individual buildings in lidar point clouds. Damaged building candidates were identified based on rooftop inclination angle, and then damage was assessed using planarity and point height metrics. Validation of the building segmentation and damage assessment techniques were performed using airborne lidar data collected after the Haiti earthquake of 2010. Building segmentation and damage assessment accuracies of 93.8% and 78.9%, respectively, were obtained using lidar point clouds and expert damage assessments of 1953 buildings in heavily damaged regions. We believe this research presents an indication of the utility of airborne lidar remote sensing for increasing the efficiency and speed at which emergency response operations are performed. © The Authors. Published by SPIE under a Creative Commons Attribution 3.0 Unported License. Distribution or reproduction of this work in whole or in part requires full attribution of the original publication, including its DOI. [DOI: [10.1117/1.JRS.11.046024](https://doi.org/10.1117/1.JRS.11.046024)]

Keywords: disaster management; building damage assessment; airborne lidar.

Paper 170705 received Aug. 5, 2017; accepted for publication Nov. 22, 2017; published online Dec. 18, 2017.

1 Introduction

Building damage assessment represents an urgent response priority following a natural disaster. This is true because determination of the damage status of buildings allows first responders to be directed to the most important locations, while resources, which are often a limiting factor in emergency response, can be utilized to their full potential. The geographic extent of large-scale disasters and potentially unsafe ground conditions often prohibit field-based assessments from providing a rapid overview of damaged regions.¹ Instead, remote sensing platforms can be utilized to collect two-(2-D) or three-dimensional (3-D) data over the affected area, which can in turn be used to assess damage, either manually by analysts or automatically by algorithms.

Much research has been conducted in building damage assessment using a wide variety of remotely sensed data. In the 2-D domain, optical and synthetic aperture radar (SAR) imagery have been used to detect and classify the damage of buildings after disasters such as earthquakes. For optical imagery, features such as spectra, texture, shape, and building shadows have been used to detect damage in both change detection and postevent analysis.² However, change detection of optical imagery requires precise registration between the two sets of imagery and can result in false alarms due to illumination and color differences. Methods using SAR imagery typically exploit backscattering intensity and phase information to locate damage.³ Success with SAR data in urban areas has been limited due to issues arising from an oblique viewing geometry, occlusions, and multiple scattering from tall buildings.² Some building types are also undetectable in the 2-D domain due to a lack of height information, such as “pancake collapses,” in which one or more stories collapse onto themselves because of structural failures.⁴

*Address all correspondence to: Colin Axel, E-mail: cma4204@rit.edu

3-D data, typically in the form of point clouds, provide accurate height information that facilitates the detection of building damage. Geometric features such as planarity and inclination angle, combined with surface features such as curvature and size, provide insight into the structural condition of a building. Point clouds can be collected using airborne light detection and ranging (lidar) or formed during postprocessing from imagery (structure-from-motion).⁵ Airborne lidar is advantageous for rapid building damage detection for several reasons. The data can be collected day or night and do not suffer from illumination shadows cast by tall buildings. Additionally, the data can be used as soon as it is downloaded from the sensor without the need for extensive postprocessing (as compared to image-based point clouds), which is critical in a time-sensitive scenario such as disaster response. Select research has used coregistered lidar data and optical imagery for building damage assessment using object-based image analysis and obtained improved results over using just lidar data.^{6,7} For example, one experiment used spectral and textural features in addition to lidar-derived height values in an object-based image classification framework and obtained 87% overall classification accuracy for several classes in postearthquake Haiti including damaged buildings.⁷ Although these methods are effective, they need additional processing to coregister the sensing modalities and rely on a collection platform that can collect both modalities simultaneously or require multiple data collections. Other studies have performed change detection on pre- and postevent planar segments, derived from 3-D data, to classify damage based on changes such as volume reduction and inclination change.^{8,9} The challenge with change detection is that it relies on the availability of pre-event data, which is not always the case. Even in cases where the pre-event data are available, precise registration is needed to avoid introduction of damage false alarms, which can be difficult with two 3-D datasets that are often collected by different sensors, and at different point densities. Methods that use only postevent lidar point clouds stand alone and can be applied to the raw point cloud data as soon as it is available from the laser scanner. These methods address the time-sensitive criteria of a natural disaster response plan and will be the focus of this paper. In the next section, a review of the methods in the literature that utilize only postevent airborne lidar point clouds for building damage assessment is presented.

2 Literature Review

A wide variety of techniques for detecting building damage from airborne lidar point clouds are present in the literature. All of the methods are based on extracting features and classifying damage at either the point, segment, or roof level. One of the main distinguishing features is the use of supervised classification versus unsupervised classification.

Methods that use supervised classification require a human operator to manually select labeled data to train a classifier. In the case of building damage assessment, this means supplying the classifier with training data with damage labels. For example, one publication used a linear support vector machine (SVM) to classify a digital surface model (DSM) rasterized from a point cloud collected after an earthquake.¹⁰ The SVM was used for binary classification of pixels into a debris class or intact class, based on eight texture features and height above ground. Buildings containing 30% or more debris pixels were classified as damaged. The study reported an overall classification accuracy of 91.6% on a dataset consisting of 43 buildings, of which only five were damaged. Another experiment performed supervised classification of planar segments, extracted from postearthquake lidar data, using a rule-based classifier and a maximum entropy classifier.¹¹ Five features were computed for each segment that represented its size, height, and planarity. The rule-based and maximum entropy classifiers obtained overall qualities of 56% and 60%, respectively. Similarly, a different publication classified planar segments into damaged or undamaged using three different classifiers: a linear discriminant classifier, linear SVM, and random forests.¹² An initial set of 18 segment features was reduced to a subset of six using forward selection and backward elimination. Notable features from the subset included the ratio of unsegmented points to segmented points, sphericity, and the height above ground. The authors were able to obtain an overall accuracy of 85% on a test set of 698 labeled segments, but did not supply a method for deriving building-level damage from the classified segments. The main limitation of damage assessment techniques based on supervised classification is that they

require a user to manually label training data. As the number of features increases, the number of training samples needed to produce an accurate classification also increases. The training time can create a significant bottleneck in the damage assessment workflow, which could negatively impact the response effort.

Unsupervised classification techniques found in the literature can be divided into two distinct groups, namely those that classify damage at the point level and those that classify damage at the building level. One publication proposed a method in which damaged roof points were detected by comparing the slopes of the lines formed with neighboring points.¹³ If the difference in slopes in both x - and y -directions was greater than some threshold, the point was considered damaged. Buildings containing 16% or more damaged points were classified as damaged. The method achieved an overall accuracy of 73.4%, but was able to correctly classify only 38.5% of the heavily damaged buildings in the dataset. In a continuation of that study, authors of Ref. 14 developed an improved damage assessment technique that used surface normals to identify damaged points. Building roofs were gridded into small tiles and the points within each tile were used to compute a surface normal. A histogram of the angles between the surface normals and a zenith vector was computed; the points in tiles with angles that fell in bins containing less than 20% of all the angles were considered damaged. A damage percentage was computed as the ratio of damaged points to total points in a building. Buildings with a damage percentage of 51% or higher were classified as damaged. The study achieved an accuracy of 68.3% on a dataset of 160 buildings, with many false alarms resulting from hipped roofs. The angles between surface normals and zenith vectors have been utilized to identify damaged points.¹⁵ A surface normal and angle was computed for every point using its nine nearest neighbors. Angular thresholds were derived to classify points in flat roofs and inclined roofs as damaged by examining the angle distributions for buildings labeled as collapsed, partly collapsed, or undamaged. The ratio of the angle standard deviation to the mean absolute deviation was used as an indicator of the severity of damage. Although visual comparisons are presented for buildings with differing levels of damage, no classification results are presented in that paper.

Other damage assessment methodologies compute building-level features in order to achieve a more direct classification. For example, one study classified buildings as damaged if their inclination angle was above a certain threshold.¹⁶ The inclination angle was defined as the angle between the geometric axis of the building (the normal vector of the roof plane for flat buildings, or the sum of the normal vectors of “main” planes for buildings with slanted roofs) and the normal vector of the terrain that the building sits on. However, classification results were presented only for a single undamaged and damaged building. Another publication introduced a damage classification method using a 3-D shape descriptor for buildings.¹⁷ Each building was represented as clusters of contours within a contour tree, and the shape descriptor was computed based on shape similarities within the contours. Buildings were classified as damaged if the shape descriptor met a set threshold. The algorithm was able to correctly classify 87% of the 1875 buildings tested.

Despite the wide range of damage assessment mechanisms present in the literature, the majority of them share similar strategies. For example, planarity, surface normals, and the angles between surface normals and a vertical zenith vector appear in many of the previously mentioned publications. This is because building damage assessment lends itself naturally to an investigation of planarity and surface normals. A manmade surface is typically dominated by planes, but this assumption is often violated when buildings are damaged.¹⁸ Roof points at high or near vertical angles are unusual for most intact buildings and can be used as an indication of damage. Combining aspects of work already in the literature, this paper proposes an automatic building damage assessment methodology that identifies candidate damaged buildings as those with roof points at high angles and then performs a rule-based classification dependent on the planarity and height above ground of roof features. The objective of this research is to assess the feasibility of creating an end-to-end, robust building damage assessment algorithm that is fully unsupervised and requires only a postdisaster point cloud. Based on the work presented in the literature, it is hypothesized that the research objective can be accomplished if appropriate point features are used as indicators of damage, and if building damage is adequately sampled in the lidar point clouds.

Section 3 introduces the dataset and building damage scale used for this research. Section 4 provides a detailed explanation of the methods used to preprocess the point cloud, detect

buildings, and finally assess damage. In Sec. 5, experimental results are presented on airborne lidar data from the 2010 earthquake in Haiti, along with a discussion of the results, while conclusions and future work are addressed in Sec. 6.

3 Data

The study area used to test the proposed algorithm consists of seven sites located in the Haitian cities of Port-au-Prince and Carrefour, two regions that were heavily affected by the 7.0M_w earthquake that occurred on January 12, 2010 (see Fig. 1). These sites were chosen because they contain both a wide range of building types and building damage types. Different construction types include: one- to three-story reinforced concrete buildings, masonry bearing walls, timber frames, and shanty housing made of reinforced concrete and masonry block with corrugated metal roofs.¹⁹ The damage level in buildings range from completely undamaged to fully destroyed, and everything in between.

The building damage assessment method proposed in this paper requires only one input, an airborne lidar point cloud. The lidar data used for development and testing were collected on January 21, 2010 by Kucera International Inc. and the Rochester Institute of Technology (RIT).

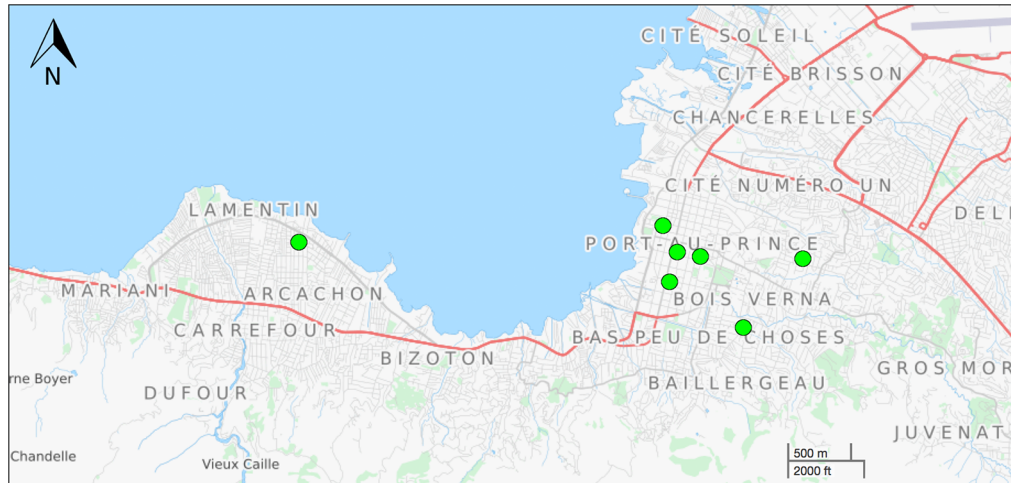


Fig. 1 A map showing the portion of Haiti that was used in this research. The location of the seven test sites are marked by the circles.

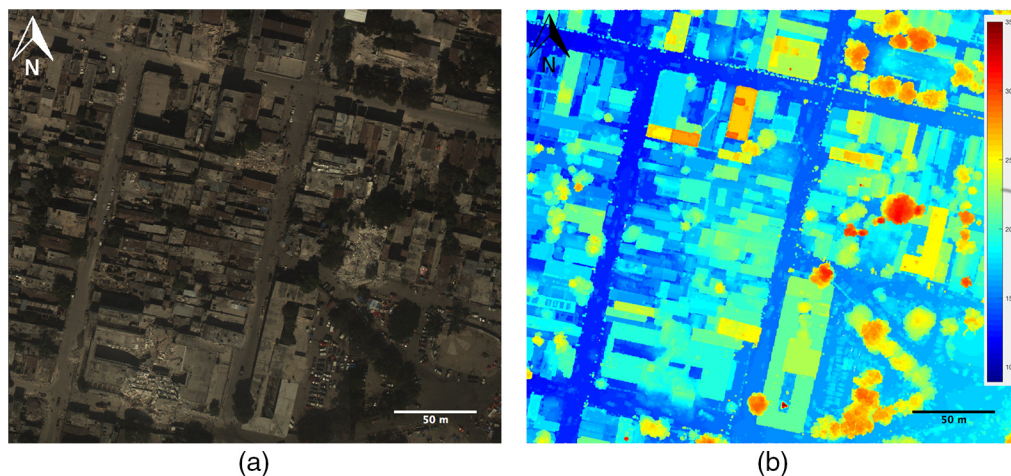


Fig. 2 One of the sites from Port-au-Prince: (a) a WASP image of the scene and (b) the corresponding lidar point cloud of the scene. The cloud is colored by height.

The point clouds for the seven sites have an average point density of 4.2 pts/m² and were captured by a Leica ALS60 at an altitude of ~820 m with a pulse rate of 150 kHz. The vertical point measurement accuracy of the instrument is 0.15 m. Multispectral imagery simultaneously was collected on the aircraft by the Wildfire Airborne Sensing Platform (WASP) system at a resolution of 0.15 m.²⁰ Although the imagery was not used for building segmentation or damage assessment, it served as a reference for visualization and was used in figures throughout the paper. Figure 2 shows both a WASP image and lidar point cloud of one of the scenes from Port-au-Prince.

4 Methods

The proposed workflow ingests an airborne lidar point cloud, segments the cloud into individual building regions, and then classifies the damage level of the buildings. The method can be divided into three components: preprocessing, building segmentation, and damage classification. Figure 3 shows a visual representation of the workflow.

4.1 Point Cloud Preprocessing

The only input to the proposed workflow is an airborne lidar point cloud of a disaster-affected region. The required attributes for each point in the cloud are its x -, y -, and z -coordinates. The point cloud is filtered to remove noise and outlier points using a filtering technique called statistical outlier removal.²¹ The statistical outlier removal algorithm first calculates the mean Euclidean distance, d , between each point, p , and its k -closest neighbors. Statistics of the mean distances are used to characterize the distribution across all of the points in the cloud. Specifically, the mean (μ_k) and standard deviation (σ_k) are calculated, and points that are considered statistical outliers, based on those two values, are removed

$$P^* = \{p \in P | d \leq (\mu_k + \gamma \cdot \sigma_k)\}, \quad (1)$$

where P^* is the entire point cloud after statistical outlier removal and γ is a scalar multiplier to control the severity of point removal.

Ground points are separated from nonground points (i.e., buildings, vegetation, and vehicles) through an implementation of the progressive morphological filter (PMF).²² The PMF operates by using a morphological filter of gradually increasing window size, along with an elevation difference threshold, to separate ground and nonground points. The increasing window size is used to remove nonground objects of increasing size in each iteration, i.e., initially small objects such as bushes, and later large objects such as buildings. One of the main parameters of the PMF is the initial elevation difference threshold, d_{ho} . Points that have a height above the estimated ground surface higher than d_{ho} are classified as nonground points. By setting d_{ho} to be

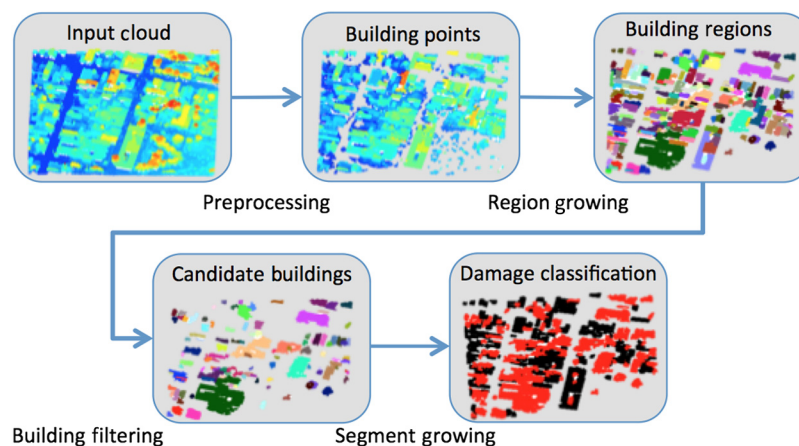


Fig. 3 A high-level flowchart of the building damage assessment algorithm.

very low (we use the vertical accuracy of the lidar system, which is 15 cm in our case), debris around the base of damaged buildings can be correctly classified as nonground, facilitating damage assessment in a later portion of the workflow. Following the PMF, a digital elevation model (DEM) is created by performing a Delaunay triangulation with natural neighbor interpolation on the points classified as ground. A normalized digital surface model (nDSM) of the points classified as nonground is created by subtracting their elevations from the corresponding DEM elevations. The nDSM is stored for use later in the workflow.

The final step of preprocessing is vegetation removal. Vegetation contained in the point cloud can be mistaken for damaged buildings during damage assessment, so it is important to remove as much vegetation cover as possible during preprocessing. Vegetation removal is accomplished using a graph cuts optimization, based on local surface properties of points.²³ Manmade surfaces are typically locally smooth with little variation of surface normals in a small region. Vegetation, on the other hand, exhibits large variations of surface normals and high curvature. Therefore, local surface properties are used to distinguish between vegetation and nonvegetation points. First, the normal vector of each point, p , is calculated using all points in a local neighborhood, N_p , defined by the radius, r . The radius, r , is automatically set as $2 \cdot \mu_k$ to ensure that enough points are used to obtain a reliable normal vector estimate. The neighborhood points of p are obtained using

$$N_p = \{q \in P | d(p, q) < r\}, \quad (2)$$

where q are all of the points in the entire point cloud P , and d is the Euclidean distance between two 3-D points. Eigenanalysis of the covariance matrix of N_p produces the eigenvalues $\lambda_1 < \lambda_2 < \lambda_3$. The eigenvector corresponding to λ_1 is the estimate of the point normal, n . Airborne lidar data are collected from above, so the absolute value of the z -component of n is used to ensure the normal vector points outward from the surface. The curvature, v , of the point can also be computed using

$$v = \frac{\lambda_1}{\lambda_1 + \lambda_2 + \lambda_3}. \quad (3)$$

The neighborhood analysis is taken one step further by computing the distribution of normals in N_p . Eigenanalysis is again used, but this time on the covariance matrix of N_p , resulting in $\lambda_{n_1} < \lambda_{n_2} < \lambda_{n_3}$. The eigenvalue corresponding to λ_{n_2} is representative of the variation of the local distribution of normals around the point. For simplified notation in equations, λ_{n_2} will be represented as f .

A weighted graph, G , of all of the points, v , is constructed using each point and its four nearest neighbors. The weight on each link between two points is the inverse Euclidean distance between the two points. The energy function, E , which is comprised of a data term, D , and a smoothness term, S , is used for optimization, as shown in

$$E(l) = \sum_{p \in \mathcal{V}} D_p(l_p) + \sum_{\{p, q\} \in N_p} S_{p, q}(l_p, l_q), \quad (4)$$

where l is the label assigned to a given point (i.e., 1 for vegetation and 0 for nonvegetation).

The data term [Eq. (5)] accounts for the curvatures and local normal variations for each point in the cost of making a cut

$$D_p(l_p) = s_v \cdot e^{-\frac{v_p^2}{\sigma_v^2}} + sf \cdot e^{-\frac{f_p^2}{\sigma_f^2}}, \quad (5)$$

where s_v is a scalar coefficient for the curvature term, sf is a scalar coefficient for the normal variation term, σ_v is the standard deviation of all of the point curvatures, and σ_f is the standard deviation of all of the normal variations.

The smoothness term [Eq. (6)] controls the cost of neighboring points based on their labels

$$S_{p, q}(l_p, l_q) = ss \cdot \delta(l_p, l_q), \quad (6)$$

$$\delta(l_p, l_q) = \begin{cases} 0 & \text{if } l_p = l_q \\ 1 & \text{if } l_p \neq l_q \end{cases}, \quad (7)$$

where ss is a scalar coefficient.

As a result of the graph cuts optimization, all points are classified as either vegetation or nonvegetation. The points classified as vegetation are removed, which is the last preprocessing step.

4.2 Building Segmentation

After preprocessing, the remainder of the points consists of buildings, debris from damaged buildings, and small objects, such as vehicles. The next step is to segment both damaged and undamaged buildings into individual point clouds for damage assessment. To accomplish this task, a region-growing-with-smoothness-constraint approach is used.²⁴

The inputs to region growing include all of the points (p), their normal vectors (n), and their curvatures (v). The points are sorted in order of increasing curvature and the point with the lowest curvature is chosen as the first seed point. All of the points within a selected radius, r , of the seed point are considered as candidate points in the current region. If the angle, β , between the normal vectors of the seed point and candidate point are below an angular threshold, T_β , then the point is added to the current region. If the curvature, v , of the candidate point is below a threshold, T_v , it is added to a list of potential seed points. After all of the candidate points have been tested, the next seed point in the list is selected as the current seed point and the process is repeated. This sequence of steps is repeated until no more points can be added to the current region. The process is completed once all of the points have been assigned to a region. The value of r is automatically set to $2 \cdot \mu_k$ to ensure that all neighbors belonging to the same surface are considered.

The output of region growing is considered to be the set of buildings to be assessed for damage. A minimum point threshold, T_p , of 100 points is used to remove regions that are too small to be reliably considered for damage. Figure 4 shows an example of building segmentation of a point cloud that has already been preprocessed. Some undersegmentation is present as a result of liberal parameters ($T_\beta = 25$ deg, $T_v = 0.05$) and the close proximity of buildings in the scene. These thresholds were determined through experimentation with several scenes. However, this undersegmentation allows debris to be segmented with buildings and helps to identify partially damaged buildings during the next step of the algorithm.

4.3 Building Damage Assessment

The proposed algorithm performs a classification of each building into one of two categories, damaged or undamaged. There is sometimes a large amount of segmented buildings in a given scene that are completely undamaged. To maintain efficiency, the first step is to identify candidate buildings for damage assessment, using two rules. The first rule involves analyzing the angles, θ , between the surface normals of building points and a horizontal zenith vector (i.e., corresponding to a normal vector of a completely vertical surface). Low values of θ (i.e., corresponding to highly inclined surfaces) have been shown to be indicative of building damage, in both flat and inclined roofs.¹⁵ An angle threshold, T_θ , is used to find potential damage points. If a building has a ratio of points that are below T_θ to total points greater than a threshold, T_d , then that building is marked as a candidate for damage assessment. Figure 5 shows point clouds for an undamaged building and a damaged building colored by θ . The damaged building had several values in the 40 deg to 70 deg range, while the undamaged building points were almost exclusively 85 deg or higher.

The second rule is based on the assumption that damage to buildings often results in debris and portions of the building located at low points around the base of the building. A majority of these points are segmented with the rest of the building points during region growing. Low points are identified as points with a height, h , (taken from the nDSM) below a threshold, T_h . If a building has a ratio of low points to total points above a certain threshold, T_l , then that building is marked as a candidate for damage assessment. Figure 6 shows point clouds for an undamaged

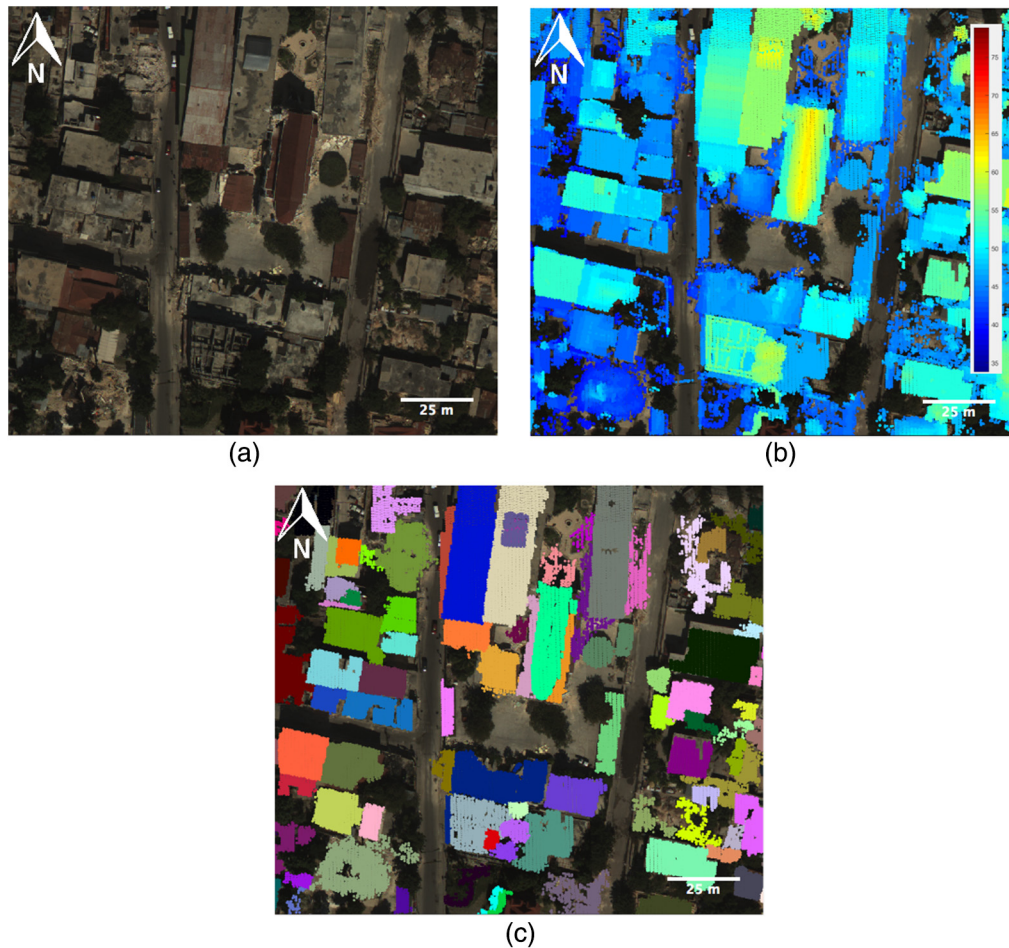


Fig. 4 A point cloud in a region of Port-au-Prince that was affected by an earthquake is segmented into buildings using region growing: (a) a WASP image of the scene and (b) the point cloud overlaid on the image. The point cloud is colored by height; and (c) the individual buildings resulting from region growing overlaid on the image. Each unique color represents one building.

building and a damaged building, colored by h . Points in the undamaged building range from about 9 to 13 m above ground, whereas points in the damaged building range from 0 to 2.5 m above ground.

Equation (8) shows the mathematical notation of the two rules

$$C(i) = \begin{cases} 1 & \text{if } \frac{\#(0_i < T_\theta)}{\#0_i < T_d} \vee \frac{\#(h_i < T_h)}{\#h_i} < T_l, \\ 0 & \text{otherwise} \end{cases}, \quad (8)$$

where $C(i)$ indicates if building i is a candidate for damage assessment (a value of 1 indicates candidacy). The values of the thresholds will be described in Sec. 5.

Each candidate building undergoes evaluation to classify it as damaged or undamaged. Similar to the candidate identification process, two rules are used for damage classification, a planarity rule and a height rule. An assumption is made that undamaged parts of buildings can be represented as planar segments. The points are segmented into planes using the same region growing algorithm described in Sec. 4.2, but with stricter parameters to ensure planarity ($T_\beta = 4$ deg, $T_v = 0.02$). A minimum segment size of 15 points is used to prevent miniature groups of points being counted as true planes. A threshold, T_s , is placed on the ratio of segmented points (those that were grouped into planar segments) to total points (R_s) in a building. If the segmentation ratio is less than T_s , then the building is classified as damaged and removed from the list of candidate buildings. Figure 7 shows the segmentation of planes of an undamaged and a damaged building.

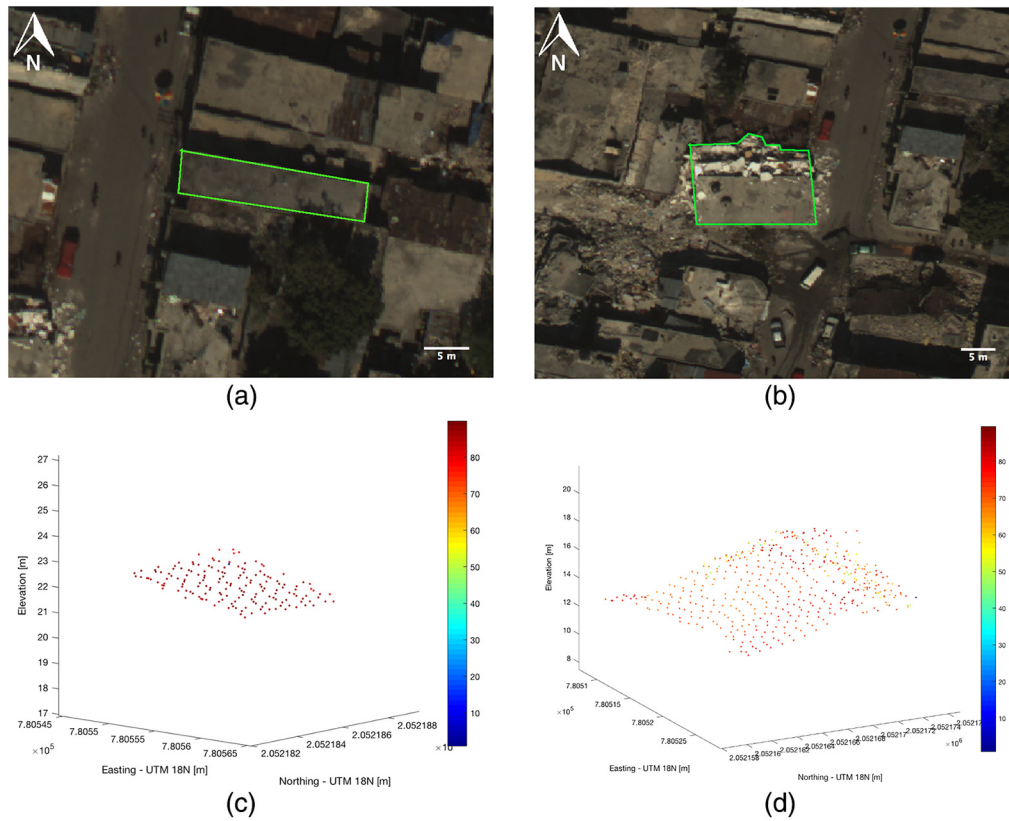


Fig. 5 Surface normal angles (θ) of an undamaged and a damaged building in Port-au-Prince: (a) a WASP image of the undamaged building; (b) a WASP image of the damaged building; (c) the undamaged building point cloud, colored by θ ; and (d) the damaged building point cloud, colored by θ .

In some cases, whole roofs or portions of roofs remain intact, but collapse to the ground due to damage to the structures that support the roof. The planarity rule is unable to detect these types of damage, so a second rule, based on height, is used. Low points are identified as points with a height, h , below a threshold, T_{h2} . If a building has a ratio of low points to total points above a certain threshold, T_{l2} , then that building is classified as damaged. Otherwise, the building is classified as undamaged.

$$Dam(i) = \begin{cases} 1 & \text{if } R_{s_i} < T_s \vee \frac{\#(h_i < T_{h2})}{\#h_i < T_{l2}}, \\ 0 & \text{otherwise} \end{cases}, \quad (9)$$

where $Dam(i)$ is the classification of building i .

4.4 Experimental Validation

4.4.1 Building segmentation and damage assessment reference data

In the days and weeks following the Haiti earthquake on January 2, 2010, several organizations contributed to what became an international humanitarian relief effort. A joint collaboration between the United Nations Institute for Training and Research (UNITAR) and the World Bank (WB) generated building damage assessments for most of the affected regions of Haiti. The assessments were made by manually interpreting pre- and postevent airborne (15 cm) and satellite (50 cm) imagery. The EMS98²⁵ provides a building damage classification scale from grades I to V, in increasing order of damage, and is frequently used with remotely sensed imagery.²⁶ The Haiti damage assessments were classified into four damage categories, roughly corresponding to damage grades I, III, IV, and V from the EMS98: no visible damage,

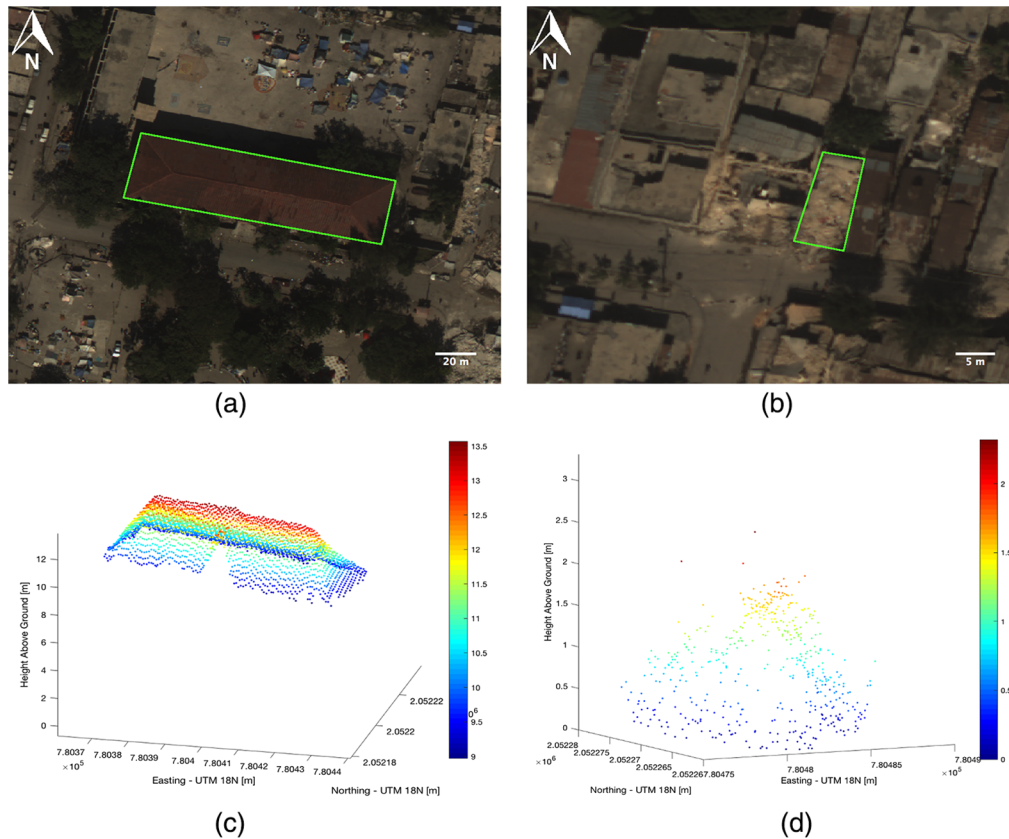


Fig. 6 Heights (h) of an undamaged and a damaged building in Port-au-Prince: (a) a WASP image of the undamaged building; (b) a WASP image of the damaged building; (c) the undamaged building point cloud, colored by h ; and (d) the damaged building point cloud, colored by h .

moderately damaged, severely damaged, and destroyed.²⁷ Figure 8 shows illustrations from the EMS98 that represent the damage categories for masonry and concrete buildings used in assessing the Haiti imagery.

A GIS file containing the building damage assessments by the UNITAR-WB team in point form was obtained and cropped to the seven test sites described in Sec. 3. Building outlines, corresponding to the validation assessment points, were manually traced using 15 cm WASP imagery and attributed with the damage grades from the assessment. These building outlines with associated damage grades served as the reference data for validating the building segmentation and damage classifications of the proposed algorithm.

4.4.2 Building segmentation

The workflow proposed in this paper was applied to the point clouds from the seven test sites from Port-au-Prince and Carrefour. The point clouds were automatically preprocessed and segmented into individual buildings. The 2-D boundaries of the segmented building point clouds were compared with the reference building outlines. Each reference building polygon was labeled a true positive (TP) if it overlapped with the boundary of a building segmented by the algorithm, or a false negative (FN) if there was no overlap. Buildings segmented by the algorithm that did not overlap any of the reference polygons were classified as a false positive (FP). True negatives (TNs) were not considered, because all segmented objects are assumed to be buildings. The TP, FP, and FN counts were used to characterize the performance in terms of completeness, correctness, and quality. Completeness reflects the percentage of validation buildings that were detected by the algorithm [Eq. (10)]. Correctness reflects the percentage of segmented buildings that were true buildings [Eq. (11)]. Quality is a measure of overall performance that takes into account both the completeness and the correctness of the results [Eq. (12)]²⁸

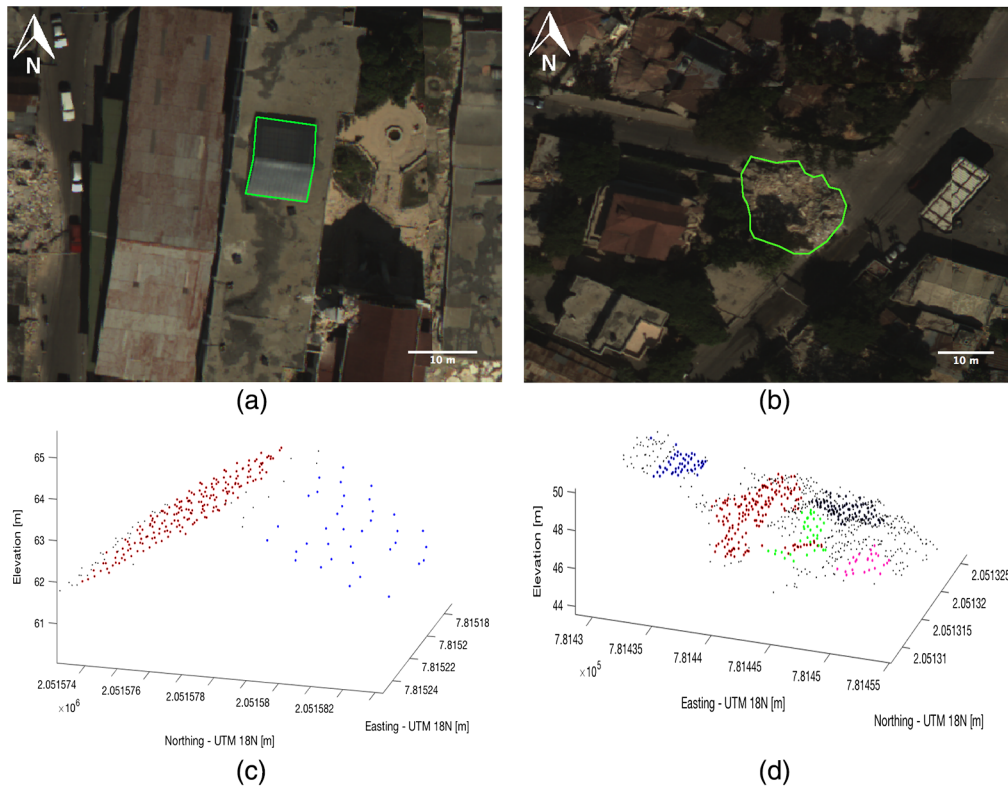


Fig. 7 Segmentation of planes in undamaged and damaged buildings. Points segmented into planes are colored based on the planar segment, and black points are points that were unsegmented: (a) a WASP image of the undamaged building; (b) a WASP image of the damaged building; (c) an undamaged building is segmented into two planes and very few points are left unsegmented; and (d) a damaged building is segmented into five planes and almost half of the points remain unsegmented.









	I: Insignificant or No Damage	III: Moderate Damage	IV: Substantial/ Heavy Damage	V: Destroyed
Masonry				
Concrete				

Fig. 8 Illustrations of the damage grades used for masonry and concrete buildings in the Haiti building damage assessment. Figure adapted from the EMS98 document.²⁵

$$\text{Completeness} = \frac{TP}{TP + FN} \times 100, \quad (10)$$

$$\text{Correctness} = \frac{TP}{TP + FP} \times 100, \quad (11)$$

$$\text{Quality} = \frac{TP}{TP + FP + FN} \times 100. \quad (12)$$

4.4.3 Building damage assessment

In order to accurately evaluate the building damage assessment algorithm, it was necessary to decouple it from the effects of building segmentation. Rather than using the building point clouds from building segmentation, building point clouds were segmented for each building by extracting all of the points from the preprocessed point cloud that fell within the validation building boundaries. The building damage assessment algorithm was applied to each building, resulting in a classification of damaged or undamaged. To match the binary output of the algorithm, the reference damage data were grouped from four severity grades into two classes. Buildings that were labeled grade I (no visible damage) were considered undamaged, and buildings that were classified as grades III to V (from moderately damaged to destroyed) were clustered into the damaged group. If a damaged reference building was correctly classified as damaged by the algorithm, it was labeled a TP. If an undamaged reference building was correctly classified as undamaged by the algorithm, it was labeled a TN. If a damaged reference building was misclassified as undamaged by the algorithm, it was labeled an FN. If an undamaged reference building was misclassified as damaged by the algorithm, it was labeled an FP.

5 Results and Discussion

5.1 Building Segmentation Results

Table 1 presents the segmentation results for all seven scenes in terms of completeness, correctness, and quality. A visual representation of the segmentation results for scene one is shown in Fig. 9. The results are overlaid on a WASP image of the scene in Port-au-Prince. The building outlines drawn in green represent TP, those in red are FP, and those in blue are FN.

The algorithm achieved high accuracy for building segmentation, obtaining completeness, and correctness scores of over 96%, and an overall quality of 93.46%. Out of 1953 validation buildings, 1890 were correctly segmented with only 0.03% errors of commission and 0.03% errors of omission. Although the algorithm showed robust performance during the building segmentation stage, it was still important to investigate the source of the errors of commission and omission. Building segmentation preceded the main objective, damage assessment, and therefore errors during the segmentation stage were propagated throughout the rest of the workflow. The following paragraphs take a closer look at the sources of false positives and negatives during building segmentation.

Overall, obtaining 63 false negatives out of 1953 buildings was a satisfactory result. Analysis of the false negatives revealed that most of them were undamaged buildings. Although it is better to omit an undamaged building than a damaged building in a disaster response scenario, all errors are important to understand. One of the main causes of buildings being omitted was vegetation covering the roof. Although the lidar penetrated the tree canopies and produced returns

Table 1 Building segmentation results for all seven of the scenes.

Site #	# Validation buildings	TP	FP	FN	Completeness (%)	Correctness (%)	Quality (%)
1	479	474	20	5	98.96	95.95	94.99
2	243	238	20	5	97.94	92.25	90.49
3	132	129	1	3	97.73	99.23	96.99
4	232	223	4	9	96.12	98.24	94.49
5	326	303	6	23	92.94	98.06	91.27
6	199	191	4	8	95.98	97.95	94.09
7	342	242	8	10	96.03	96.80	93.08
Total	1953	1890	63	63	96.77	96.77	93.75



Fig. 9 Building segmentation results for a scene in Port-au-Prince. True positives are outlined in green, false positives are outlined in red, and false negatives are outlined in blue.

below the canopy in some cases, the number of returns on rooftops obscured by vegetation was much lower than those with no obscurations. In cases where the vegetation covered a significant portion of the rooftop, buildings sometimes did not meet the 100 point criteria required for segmentation. Many of the other false negatives were simply a result of the buildings not having enough lidar points and therefore being rejected. Shanty housing represented a majority of the false negatives caused by size. The best way to overcome this problem would be to collect lidar data at a higher point density. The vegetation removal step may have also contributed to the errors of omission. Both vegetation and building debris exhibited increased surface normal variation and high curvature. Typically, normal variation and curvature were significantly higher for vegetation than for damaged buildings, but in some extreme cases, portions of damaged buildings were misclassified as vegetation and removed. Future work will investigate augmenting the vegetation identification step to prevent accidental debris removal. One potential solution would be to use the number of returns for each point as an added feature to classify vegetation.

In total, 63 errors of commission were observed when evaluating the proposed algorithm. False positives were troublesome when passed into the building damage assessment stage, because the planarity assumptions used for rooftops did not hold for other objects, especially when they were not manmade objects. The largest cause of false positives was vegetation. In some cases, the centers of large, dense canopies were not properly classified during the vegetation removal stage. Multiple iterations of graph cuts with adjusted parameters could potentially be used to improve vegetation removal, but the need for parameter adjustment would result in a loss of automation and therefore was not implemented in our workflow. Relief tents were another source of errors of commission. The canopy tents were distributed during the emergency response for people displaced by the earthquake. Often times, many of the tents were placed directly next to each other, and as a result, several tents were segmented together as a single building. In some cases, groups of two or three tall vehicles in close proximity to each other were mistaken as a building. Despite these sources of errors of commission, a correctness

Table 2 Confusion matrix for damage classification of all scenes.

Classification		Reference		
		Undamaged	Damaged	User's accuracy (%)
Algorithm	Undamaged	922	193	82.69
	Damaged	219	619	73.87
	Producer's accuracy (%)	80.81	76.23	
	Overall accuracy (%)	78.90		
	Kappa coefficient	0.57		

of almost 97% suggests that our algorithm can reliably be used for building segmentation even at relatively low point densities (4.2 pts/m² on average).

5.2 Building Damage Assessment Results

A total of 1953 buildings (812 damaged and 1141 undamaged in the reference data) across seven scenes were classified as damaged or undamaged by the algorithm. The combined confusion matrix for all seven scenes is presented in Table 2. An overall accuracy of 78.9% and a Kappa coefficient of 0.57 suggest reasonable performance in damage classification, given the wide range of building types and damage patterns in the Haiti scenes. Both errors of commission and omission were caused by several factors, which are discussed in the following paragraphs. A visualization of the results for a scene in Port-au-Prince is shown in Fig. 10.

In total, 24% errors of omission were observed across the seven scenes for the damage class. Most of these false negatives were a result of building damage not adequately being represented in the lidar data, which occurred most frequently for buildings that sustained only minor or moderate damage. For example, in some cases, small portions of a roof would break off and fall to the ground. Although the building damage was visible in imagery, the portions on the ground were low enough to be omitted during preprocessing, and the remaining intact portions of the roof resembled an intact roof. The building damage assessment considered only the intact portion of the roof, and therefore omitted the damage. The airborne lidar data did not obtain returns from building walls, so any damage sustained to the walls was not detected. In some cases buildings that were labeled damaged in the reference set were classified as undamaged by the algorithm, because there was simply no discernible damage in the point cloud or image. Rathje conducted field surveys to assess the accuracy of the UNITAR-WB team and found that the assessments were around 77% accurate.¹⁹ The authors even suggested that buildings in grade III should be grouped as undamaged, because of how difficult they are to identify from satellite and airborne imagery, but we chose not to do so because many of the buildings assigned a grade of III were clearly damaged. Despite these issues, 619 out of the 812 damaged buildings in the reference set were correctly classified. At this level of accuracy, the damage maps created by the algorithm would be vital tools for helping direct emergency responders to areas of heavy damage, and dictate regions of interest for detailed damage analysis by ground crews. Feedback from emergency response teams was that they would rather have a rapid damage assessment product than a marginally more accurate, but delayed damage map.

Errors of commission for the damage class were slightly more prevalent than errors of omission, with a total of 26% buildings misclassified as damaged. One of the main causes of false positives was vegetation that was not fully removed during preprocessing. All points that fell within the validation building outlines were used for damage assessment, including vegetation. During the actual workflow of the algorithm, vegetation that remains after preprocessing would typically be removed during the region growing stage of building segmentation due to the merging criteria. As a result of the vegetation remaining in the scene, the buildings were sometimes mistaken as damaged by the planarity rule. Another source of errors of commission were oddly shaped roofs or roofs that contained many small structures. Due to the relatively low point

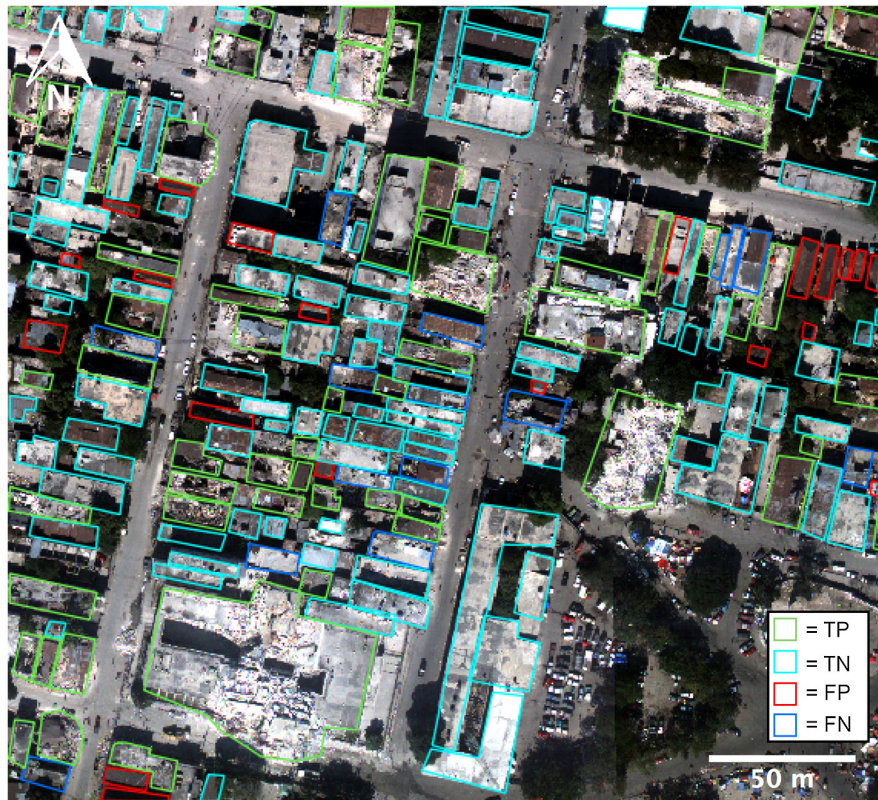


Fig. 10 Building damage assessment results for a scene in Port-au-Prince. True positives are outlined in green, true negatives are outlined in cyan, false positives are outlined in red, and false negatives are outlined in blue.

density (4.2 pts/m^2), these rooftop structures were often undersampled and resembled a damaged rooftop that was not planar. If the point clouds had been collected at higher point density, these errors would likely occur far less often. The algorithm still achieved producer's and user's accuracies of 80.8% and 82.7%, respectively, in classifying undamaged buildings, which was reasonable for a preliminary rapid-delivery damage map in such a diverse environment.

6 Conclusions

In this paper, we proposed and evaluated an automated technique for assessing building damage via airborne lidar point clouds. Local surface properties of lidar points were used, along with region growing to cluster the points into individual buildings, and then features such as surface normal angle, planarity, and height above ground were used to make a classification of damaged or undamaged. The building segmentation method was tested on seven point clouds from the 2010 Haiti earthquake, and achieved a detection accuracy of 93.75% for 1953 validation buildings. The building damage assessment algorithm was tested on the same 1953 buildings and obtained an overall damage classification accuracy of 78.9% and a Kappa coefficient of 0.57. The main factors that affected the quality of the building damage assessment were vegetation that was not successfully removed during preprocessing, and undersampling of complex rooftops due to relatively low lidar point densities (4.2 pts/m^2). Future efforts will focus on improved vegetation removal using iterative application of the graph cuts technique with data-driven parameter calculation. The results obtained from this research suggest that automated building damage assessment can be used in lieu of the traditional manual interpretation of imagery with similar levels of accuracy. Automated damage assessment could significantly reduce the time needed to produce damage maps, ultimately leading to faster and more efficient search and rescue missions and prioritization of resources.

Acknowledgments

This research was funded by the United States Department of Transportation (USDOT) under Cooperative Agreement No. OASRTRS-14-H-RPI. The views, opinions, findings, and conclusions reflected in this publication are solely those of the authors and do not represent the official policy or position of the USDOT/OST-R, or any State or other entity. USDOT/OST-R does not endorse any third party products or services that may be included in this publication or associated materials.

References

1. R. T. Eguchi et al., "The application of remote sensing technologies for disaster management," in *14th World Conf. on Earthquake Engineering* (2008).
2. L. Dong and J. Shan, "A comprehensive review of earthquake-induced building damage detection with remote sensing techniques," *ISPRS J. Photogramm. Remote Sens.* **84**, 85–99 (2013).
3. S. Voigt et al., "Rapid damage assessment and situation mapping: learning from the 2010 Haiti earthquake," *Photogramm. Eng. Remote Sens.* **77**(9), 923–931 (2011).
4. M. Rehor et al., "Contribution of two plane detection algorithms to recognition of intact and damaged buildings in lidar data," *Photogramm. Rec.* **23**(124), 441–456 (2008).
5. E. P. Baltsavias, "A comparison between photogrammetry and laser scanning," *ISPRS J. Photogramm. Remote Sens.* **54**(2), 83–94 (1999).
6. Y. Haiyang, C. Gang, and G. Xiaosan, "Earthquake-collapsed building extraction from lidar and aerophotograph based on OBIA," in *2nd Int. Conf. on Information Science and Engineering*, pp. 2034–2037, IEEE (2010).
7. S. Ural et al., "Building extraction and rubble mapping for city Port-au-Prince post-2010 earthquake with geoeye-1 imagery and lidar data," *Photogramm. Eng. Remote Sens.* **77**(10), 1011–1023 (2011).
8. C. Schweier, M. Markus, and E. Steinle, "Simulation of earthquake caused building damages for the development of fast reconnaissance techniques," *Nat. Hazards Earth Syst. Sci.* **4**(2), 285–293 (2004).
9. M. Rehor, "Classification of building damage based on laser scanning data," *Photogramm. J. Finl.* **20**(2), 54–63 (2007).
10. H. Rastiveis, E. Hosseini-Ziradoo, and F. Eslamizade, "Automatic blocked roads assessment after earthquake using high resolution satellite imagery," *Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci.* **XL-1-W5**, 601–605 (2015).
11. S. O. Elberink et al., "Detection of collapsed buildings by classifying segmented airborne laser scanner data," *Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci.* **XXXVIII-5/W12**, 307–312 (2011).
12. K. Khoshelham, S. O. Elberink, and S. Xu, "Segment-based classification of damaged building roofs in aerial laser scanning data," *IEEE Geosci. Remote Sens. Lett.* **10**(5), 1258–1262 (2013).
13. R. C. Labiak et al., "Automated method for detection and quantification of building damage and debris using post-disaster lidar data," *Proc. SPIE* **8037**, 80370F (2011).
14. R. Labiak, "A method for detection and quantification of building damage using post-disaster lidar data," Master's Thesis, Rochester Institute of Technology (2011).
15. D. Aixia et al., "Building damage extraction from post-earthquake airborne lidar data," *Acta Geol. Sin.* **90**(4), 1481–1489 (2016).
16. S. Yonglin, W. Lixin, and W. Zhi, "Identification of inclined buildings from aerial lidar data for disaster management," in *18th Int. Conf. on Geoinformatics*, pp. 1–5, IEEE (2010).
17. M. He et al., "A 3D shape descriptor based on contour clusters for damaged roof detection using airborne lidar point clouds," *Remote Sens.* **8**(3), 189 (2016).
18. G. Vosselman et al., "Recognising structure in laser scanner point clouds," *Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci.* **46**(8), 33–38 (2004).
19. E. M. Rathje et al., "Damage patterns in Port-au-Prince during the 2010 Haiti earthquake," *Earthquake Spectra* **27**(S1), S117–S136 (2011).

20. D. W. Messinger et al., "High-resolution and lidar imaging support to the Haiti earthquake relief effort," *Proc. SPIE* **7812**, 78120L (2010).
21. R. B. Rusu, "Semantic 3D object maps for everyday manipulation in human living environments," PhD Thesis, KI-Künstliche Intelligenz (2010).
22. K. Zhang et al., "A progressive morphological filter for removing nonground measurements from airborne lidar data," *IEEE Trans. Geosci. Remote Sens.* **41**(4), 872–882 (2003).
23. S. Sun, "Automatic 3D building detection and modeling from airborne LiDAR Point clouds," PhD Thesis, Rochester Institute of Technology (2013).
24. T. Rabbani, F. Van Den Heuvel, and G. Vosselmann, "Segmentation of point clouds using smoothness constraint," *Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci.* **35**(5), 248–253 (2006).
25. G. Grünthal, "European macroseismic scale 1998 (ems-98). European seismological commission, sub commission on engineering seismology, working group macroseismic scales. Conseil de l'Europe," *Cah. Cent. Eur. Géodynamique Séismol.* **15**, 15–16 (1998).
26. T.-T.-H. Pham et al., "Towards a rapid automatic detection of building damage using remote sensing for disaster management: the 2010 Haiti earthquake," *Disaster Prev. Manage.* **23**(1), 53–66 (2014).
27. UNITAR/UNOSAT, "Joint remote sensing damage assessment database," Technical Report, UNITAR/UNOSAT, World Bank, and EC Joint Research Centre (2010).
28. C. Heipke et al., "Evaluation of automatic road extraction," *Int. Arch. Photogramm. Remote Sens.* **32**(3), 151–160 (1997).

Colin Axel is a geospatial scientist at PrecisionHawk. He received his BS and PhD degrees in imaging science from Rochester Institute of Technology in 2013 and 2017, respectively. His research interests include remote sensing for disaster response and precision agriculture.

Jan van Aardt received his BSc and BSc (Hons.) degrees from the University of Stellenbosch, Stellenbosch, South Africa, in 1996 and 1998, respectively, his MS degree in forestry from Virginia Tech, Blacksburg, Virginia, USA, in 2000, and his PhD in forestry from Virginia Tech, Blacksburg, Virginia, USA, in 2004, where he focused on remote sensing (imaging spectroscopy and lidar research). He is currently a professor at the Chester F. Carlson Center for Imaging Science, Rochester Institute of Technology, Rochester, New York, USA. His research interests include remote sensing for canopy structure assessment, precision agriculture, and disaster response.