BUILDING EDGE DETECTION USING SMALL-FOOTPRINT AIRBORNE
FULL-WAVEFORM LIDAR DATA

Jean-Christophe Michelin, Clément Mallet, Nicolas David
IGN, MATIS, 73 avenue de Paris, 94160 Saint-Mandé, France; Université Paris-Est
firstname.lastname@ign.fr
Commission III - WG III/2

KEY WORDS: Building, edge, detection, lidar, waveform, line, segmentation, diagnostic.

ABSTRACT:

The full-waveform lidar technology allows a complete access to the information related to the emitted and backscattered laser signals. Although most of the common applications of full-waveform lidar are currently dedicated to the study of forested areas, some recent studies have shown that airborne full-waveform data is relevant for urban area analysis. We extend the field to pattern recognition with a focus on retrieval. Our proposed approach combines two steps. In a first time, building edges are coarsely extracted. Then, a physical analysis has barely investigated the field of pattern recognition (Jutzi and Stilla, 2005b). Building edges, and more generally 3D lines, are patterns of high interest for many applications such as strip face Models (DSM) or normalized DSM (Rutzinger et al., 2009). Classification can also be carried out by calculating the local curvature or variance of the normal vectors for each area of interest (Arefi, 2009). Furthermore, the intensity or amplitude value may also provide discriminant images for building edge detection. However, such information requires a calibration step (Höfle and Pfeifer, 2007), and has not yet been optimally used. Visible or Infra-red geospatial images are therefore often preferred as complementary information (Rottensteiner et al., 2005; Matikainen et al., 2007). The second possible approach is based on the geometrical study of the 3D point cloud. Linear features or significant vertical discontinuities can be directly detected in the point cloud (Sampath and Shan, 2007). Such analysis can be preceded and facilitated by a building focusing step, traditionally 3D point classification as in (Zhang et al., 2006; Poullis and You, 2011). Mesh-based methods such Zhou and Neumann (2010) may allow to compute accurate edges while losing the semantic information. Only few specific methods are dedicated to the lidar signal analysis. The aim is to estimate a more accurate position of the edges than the size of the lidar footprint. Jutzi and Stilla (2005a) have initiated the field with airborne simulated data, and managed to improve edge location to one to seven tenth of a pixel according to the measurement noise. Their method is based on the fact that the building edge orientation cannot be estimated with a single waveform. For each waveform, the distance to the building edge is estimated. The 3D point positioned on the roof can be anywhere on a circle centered on the waveform line of sight. Several possible building edges are extracted and the ambiguity problem is solved thanks to the power of the first backscattered echo. 3D point posi-
tion improvement has only been tackled with real data for terrestrial datasets (Neilsen, 2011). In the proposed method, the lidar point cloud is corrected from the surface response, and particularly the 3D position is refined with the backscattered waveform. Unfortunately, there are major differences between aerial and terrestrial lidar issues. This method assumes all surface to be homogeneous within the lidar footprint, which may be inaccurate with airborne large scale data.

The paper is organised as follows. The proposed strategy will be first presented as well as the available dataset. Then, each step of the workflow will be described and commented. Finally, results will be presented, and conclusions will be drawn.

2 OVERALL STRATEGY

2.1 Methodology

The proposed methodology is based on the assumption that buildings can be approximated by polygons, i.e., fragmented in segments. Given a set of lidar measurements, our method is decomposed into three steps:

- **Step 1: Segmentation of boundary regions.** Firstly, the location of building areas is coarsely retrieved, in order to decompose the problem and offer parallelization opportunities. Main vegetated areas are removed and not examined in the following steps. Secondly, for each focus area, lidar waveforms are segmented in “boundary region” or not. The process is not limited to waveforms with multiple echoes.

- **Step 2: Initial boundary extraction and diagnostic.** A coarse determination of the 3D locations of the segments is first performed. 3D point clouds are first extracted from the waveforms. Then, 3D boundary segments are estimated with the standard RANSAC algorithm, adapted to deal with remaining vegetated areas and with focus areas composed of several buildings. Finally, segments are qualified, and vegetation points are filtered. Only the most reliable segments are fed into the adjustment process.

- **Step 3: Boundary adjustment.** Based on a physical model, the positions of the 3D segments and support 3D points are improved, up to twice more accurately than the lidar footprint.

2.2 Data

Full-waveform data has been acquired over the city of Amiens, France, in February 2008, under leaf-off conditions (OPTECH 3100-EA sensor). The point density is 2 points/m². We will focus our study on the city center, which corresponds to 5 million backscattered waveforms covering 0.7 km². For each lidar pulse, both emitted (pulse width = 4.8 ns) and backscattered waveforms (1 GHz digitization rate, GPS time, position and attitude of the sensor are available. Therefore, waveforms can be georeferenced, as well as 3D subsequent points (see Section 4.1).

The French national 2D topographic database is available in order to qualify the results. The building layer has a planimetric accuracy of around 1 m. A 0.5 m orthoimage is also used to visually assess the extraction process.

Due to the scan pattern, the 3D genuine spatial sampling of the data is not regular, which does not facilitate the local analysis. The paper is organised as follows. The proposed strategy will be first presented as well as the available dataset. Then, each step of the workflow will be described and commented. Finally, results will be presented, and conclusions will be drawn.

2.3 Acquisition of building edges

Depending on the angle of incidence of the laser beam and the position of the sensor, edges of a building will not be all similarly acquired. Two cases can be distinguished:

- "Single echo" (SE): Edges corresponding to walls illuminated by the laser beam will create waveforms with a single echo (ground, facade or roof).
- "Multiple echo" (ME): The backscattered waveform is composed of two echoes: one on the roof, and one on the ground.

Our strategy mainly relies on the detection of "multiple echo" waveforms, but Step 1 is designed to retrieve all the roof waveforms close to the gutter heights. We consider that the other lidar strips may take measures on the building from the reverse point of view, thus offering the adequate complementary information for the full contour description.

3 SEGMENTATION OF BOUNDARY REGIONS

3.1 Focus on boundary areas

The method is based on the computation of vertical discontinuities in 3D waveform data. For each waveform, if the vertical discontinuity is superior to a theoretical minimal building gutter height, $h_{\text{min}}$ (set to 2 m to favour a high detection rate), it may lie on an edge. Computing such discontinuity as only the difference of elevation between the first and the last echoes of one waveform is not adequate, since it may discard many waveforms and provide noisy results. Therefore, a 4-connexity neighborhood $V$ in the sensor topology is used to detect the two illumination cases.

\[
\forall w \in \mathcal{W}, \ w \in B \iff (Z_{\text{first}} - \arg\min_{w \in V} Z_{\text{last}}) \geq h_{\text{min}}.
\]

Where $W$ and $B$ are the sets of waveforms and of coarse boundary waveforms respectively. $Z_{\text{first}}$ (resp. $Z_{\text{last}}$) is the elevation of the first (resp. last) object hit by a given laser pulse. Such values can be retrieved by detecting local maxima within the signal (e.g., Gaussian smoothing + first derivative computation). Once

![Figure 1: Results on the focusing step: one color per area.](image)
3.2 Waveform segmentation

The goal of this part is to select waveforms from the \(p\)th focus area \((w \in F_p)\) which really lying a building edge by favouring a high detection rate. Selecting waveforms with two echoes will result in segmenting both building edges, chimneys, and vegetation. The segment extraction may be highly corrupted by such false detections.

Two assumptions are made for this segmentation: the height \((h_i)\) and the gutter line elevation \((E_i)\) are approximately constant by piece for the \(i^{th}\) building edge. So, for each wall, waveforms have a first echo elevation \(E_i\) and a difference of elevation between their first echoes and one of the echoes of their neighbouring waveforms of \(h_i\). Thus, one of the major difficulties is to compute \(E_i\) and \(h_i\).

An area can contain several buildings, and in case of sloped terrains, may have walls with different heights. Thus, an iterative procedure should be set up, starting with the most prominent wall until finding the lowest one. In one focus area, waveforms with two echoes and largest differences of elevation between the first and the last echoes are likely to be located on building edges. Building height can then be computed among the most important two echoes and largest differences of elevation between the first and the last echoes are likely to be located on building edges. Building height can then be computed among the most important differences of elevation. Since trees can be taller than buildings, the largest differences of elevation may sometimes be discarded.

A rank filter is applied on the candidate waveforms: the differences of elevation. Since trees can be taller than buildings, Building height can then be computed among the most important differences of elevation. Since trees can be taller than buildings, the largest differences of elevation may sometimes be discarded. A rank filter is applied on the candidate waveforms: the 70th percentile \((P_{70})\) is selected, and provides the \(h_i\) value. Thus, for each waveform \(w \in \Gamma_i\) \((\Gamma_i\) is the set of waveforms which lying the \(i^{th}\) building edge), we have:

\[
h_i = P_{70} \left\{ \left\{ Z_{\text{first}}^w - Z_{\text{last}}^w, Z_{\text{last}}^w \neq Z_{\text{last}}^w \right\} \right\}.
\]

Such value allows to select all multiple-echo waveforms that exhibit similar differences. A buffer \(\chi\) around \(h_i\) is introduced to gather all relevant waveforms. \(\chi\) is set to 0.5 m, which corresponds to half of the maximum variation of the wall height for a building. Moreover, the median (Md) altitude of the first echo of these waveforms is a good approximation of the elevation of the tallest gutter line \((E_i)\).

\[
E_i = \text{Md} \left\{ \left\{ |Z_{\text{last}}^w - Z_{\text{first}}^w| = h_i \pm \chi \right\} \right\}.
\]

Wall height and gutter line altitude of the tallest building in the focus area are then known. Waveforms lying on this building can now be selected. A waveform \(w\) is labelled as "building edge" lying on \(\Gamma_i\) if:

- Its first echo altitude is in the buffer \(\gamma\) around the gutter line elevation \(E_i\); \(\gamma\) is set to 0.2 m, which is the addition of the spatial resolution of waveforms (0.15 m), and an altimetric tolerance on the gutter line elevation (0.05 m);
- The difference of elevation between its first (or single) echo and one echo of the neighbouring waveforms (in the sensor topology, 4-connexity) is close to the wall height \(h_i\) \((\pm \chi)\).

\[
\forall w \in W, w \in \Gamma_i \iff \begin{cases} Z_{\text{first}}^w - \arg\min_{w \in \Gamma_i} Z_{\text{last}}^w = h_i \pm \chi \\ Z_{\text{first}}^w = E_i \pm \gamma \end{cases}
\]

In practice, the procedure starts with the highest building. The corresponding waveforms are then removed, and the process iterates until no significant differences of elevations are found in one area.

4 INITIAL BOUNDARY EXTRACTION AND SEGMENT DIAGNOSTIC

4.1 3D point extraction

Once waveforms lying on building edges have been segmented, 3D segments can be computed. Two strategies are conceivable: processing the spatio-temporal data volume of the georeferenced waveforms as proposed in [Jutzi and Stilla, 2005b] or the 3D points extracted from the waveforms. In order to deal with large scale issues, the second possibility is selected. For that purpose, the Gaussian decomposition is performed on the waveform (Mallet and Bretar, 2009), and only the first detected echo is preserved. Figure 2 shows the location of the extracted 3D points for an area of interest. One can notice that, building edges are well described but also that many points remain on trees. Since the following steps will deal with this problem, the segmentation step is considered as satisfactory: a high true positive rate is achieved for "building edge" points. These "building edge" points are evaluated with the available ground truth. Around 50% of the extracted 3D points are located on building edges (Figure 2).

Figure 2: First echoes extracted from waveforms labelled as "building edge" in Section 3.2. Yellow (resp. red) points are located on building (resp. vegetation) areas, with respect to the ground truth.

4.2 Segment extraction

In order to fit segments on the 3D points, a robust algorithm is required. These 3D points are positioned in the center of the 0.8 m lidar footprint diameter. Furthermore, for various configurations and objects (balconies, antennas etc.), multiple echoes are outliers and may appear. The iterative RANdom Sample Consensus (RANSAC) algorithm is appropriate to achieve this step (Fischler and Bolles, 1981). However, it has to be adapted to cope with two main issues: points located in vegetation, and separated but aligned buildings within a single focus area. For the latter case, a single segment would be extracted for two distinct buildings. Consequently, two constraints are introduced. 3D points have to fulfill both criteria to yet be considered as inliers. (cf. Figure 3, red segments).

- Orientation consistency of lines. For a given building edge, the roof is always in the same side i.e., the vertical discontinuity has always the same orientation (Poullis and You, 2011). The orientation of the building edge is thus constrained to be in the same direction along the segment. A tolerance of \(\pm 45^\circ\) is sufficient to remove false detections in vegetation. (false positive rate: 59% \(\to\) 34%).
- A maximum horizontal range between two successive points projected on the extracted line. It is set to 2 m to prevent the extraction of a unique segment on both sides of a street.
4.3 Boundary diagnostic

The final adjustment procedure requires a correct initialization. Furthermore, some remaining segments are still lying on vegetated areas, and should be removed. Consequently, the boundary diagnostic is decomposed into two steps. Firstly, segments in vegetated areas are removed, and then filtered segments are diagnosed (geometrical accuracy and reliability). The segments in vegetation areas are filtered by using two criteria:

- **Planimetric scattering**: Vegetated areas are known to generate backscattered lidar signals with a larger width than the width of the emitted pulse (here 4.8 ns). Indeed, several objects closer than the sensor range resolution will contribute to the same echo. Such echo will be wider because of the spatial extent of these targets. This is particularly true for the first echo which corresponds to the tree canopy with many material. Conversely, if the waveform hits a building edge, a single object will contribute to the first echo but the proportion of roof within the footprint is not sufficient to notice a big influence on the echo width (even if the roof is inclined). Thus, roof echoes are narrower. Consequently, segments where RANSAC inliers were extracted from waveform which have an average width superior to a threshold of 5.5 ns. Such value has been retrieved using a study of waveform widths in vegetation areas.

- **Volumetric scattering**: This criterion is point density-based. Vegetation generates more 3D points, that are in addition spatially scattered. A segment is labelled as "volumetric" whether:

\[
\frac{N_{3q} - N_{q}}{N_{q}} > 0.4 .
\]

(5)

Where \(N_{3q}\) and \(N_{q}\) are the numbers of 3D points located in a buffer respectively of 3q and q meters around the segment. This criterion is similar to (Höfle et al., 2012). For building areas, point density does not vary much while considering increasing areas of interest: \(N_{3q} \simeq N_{q}\). In this paper \(q=1\) m so 3q is lesser than street width. The 40% increase reflects the ability of lidar beams to penetrate vegetated areas, and thus generating multiple scatterings. The buffer size has to be empirically adapted to the lidar footprint and the increasing rate needs to be adjusted to the point cloud density.

<table>
<thead>
<tr>
<th></th>
<th>TRUE</th>
<th>FALSE</th>
</tr>
</thead>
<tbody>
<tr>
<td>TRUE</td>
<td>88</td>
<td>12</td>
</tr>
<tr>
<td>FALSE</td>
<td>7</td>
<td>93</td>
</tr>
</tbody>
</table>

Table 1: Confusion matrix for segments corresponding to building edges.

A segment is considered as "planar" if it fulfills these two criterions. Non-planar segments are removed. As illustrated wit the Table 1 and on Figure 3, the global filtering step permits to efficiently remove vegetated segments. The ratio of segments in vegetation areas decreases from 34 % to 7%. Few isolated residual false detections are still noticeable, but we consider that an optional regularization step that turns segments to polygonal areas, posterior to our workflow, is likely to remove these errors.

The final quality of the extracted segments can be estimated with two criterion :

- **Geometrical accuracy**: The Root Mean Square (RMS) value is used. An segment is "accurate" when: RMS < 0.5 \( d \), where \( d \) is the lidar footprint.

![Figure 3: Vegetation filtering. Red segments are discarded using orientation criterion (RANSAC procedure), blue segments with planar criterion. Yellow segments are can be diagnosed.](image)

- **Reliability**: A segment with a point density higher than 0.5 inliner/m is considered as "reliable".

The segment adjustment step is only processing accurate and reliable segments.

5 BOUNDARY ADJUSTMENT

5.1 Principle

This step is inspired from the theoretical inversion on simulated data of Jutzi and Stilla (2005a). 3D extracted points have the same order of accuracy than the lidar footprint radius. The Gaussian decomposition permits to retrieve a correct vertical accuracy but their horizontal position need to be adjusted. For that purpose, the lidar equation (Wagner, 2010) is used and inverted to exactly find their position. A Lambertian assumption allows a clear simplification of this equation but is only valid for echoes lying on building roofs:

\[
P_r = \int \int_{\xi \in \xi} \frac{K'}{R^2} \cos(\alpha) \frac{p_m}{C_{cal}} P_c(t) d(t) d(\tau),
\]

(6)

with \(C_{cal}\) is a calibration constant, \(R\) is the range between the sensor and the target, \(\alpha\) is the angle between the local normal of the roof and the direction of incidence of the laser beam, \(\xi\) is the footprint area, \(P_c\) and \(P_r\) are the emitted and received power. \(K'\) gathers other constant values. One can notes that area of the target can’t be approximated by a disk for waveforms which lying the building edge, as illustrated on Figure 5.

Roof are locally planar and radiometrically homogeneous inside the lidar footprint, but this is not the case for ground and vegetation echoes. Only first echoes are considered to adjust a segment. Our proposal method consists to iteratively:

- Use the knowledge of segment orientation to adjust its points (inliers).
- Estimate a new segment (and therefore a new orientation).

The proposed approach is a numerical Expectation-Maximisation algorithm, as illustrated in Figure 4.

5.2 Initialisation

For the \(j^{th}\) accurate and reliable segment (Section 4.3), the initial coarse orientation (\(\theta^0_j\)) and position (\(\rho^0_j\)) are already estimated in Section 4.2. The normal \(\vec{n}\) of each part of the roof edge can be extracted from the 3D point cloud using RANSAC. The first set
the roof. If the direction of propagation. A Kronecker test is performed in order to have a regular mesh in the orthogonal direction of waveforms. The local reflectance of points $\Theta_j$ is extracted from multiple-echo waveforms and located on roof edges. The goal of this step is to adjust the horizontal position of points $\Theta_j$ with a least-square algorithm ($\Theta_j$). Then, the new orientation $\theta_j^{k+1}$ and position $\gamma_j^{k+1}$ are derived. Such value can be used to iterate on the 3D point positions. The algorithm stops when the change of the orientation is negligible ($\eta = 5\sigma$).

As illustrated on Figure 6, the adjusted segment has a better accuracy (in average the RMS decrease by 0.15 m).

\section{5.4 Maximisation}

Adjusted segment can be computed on the set of adjusted points with a least-square algorithm ($\Theta_j$). Then, the new orientation $\theta_j^{k+1}$ and position $\gamma_j^{k+1}$ are derived. Such value can be used to iterate on the 3D point positions. The algorithm stops when the change of the orientation is negligible ($\eta = 5\sigma$).

As illustrated on Figure 6, the adjusted segment has a better accuracy (in average the RMS decrease by 0.15 m).

\section{5.3 Expectation}

The goal of this step is to adjust the horizontal position of points extracted from multiple-echo waveforms and located on roof edges. In a first time, the ellipsis resulting from the intersection between the laser beam and the roof plane is computed. This requires the knowledge of $N$, the beam divergence $\beta$, the position and orientation of the sensor, and the standard deviation of the beam $\sigma_B$. $\sigma_B$, in case of a spatial Gaussian beam, is the distance between the center of the line of sight and the roof, where the amplitude is divided by $e^2$ ($\sigma_B = R/(2\ln 2)^{1/2}$). The ellipsis is then spatially sampled ($A(\tau) = 1$ cm$^2$). An angular decomposition is performed in order to have a regular mesh in the orthogonal direction of propagation. A Kronecker test $\delta(\tau_j)$, based on orientation, is run for each cell $\tau_j$ in order to know whether it belongs to the roof.

\begin{figure}[h]
\centering
\includegraphics[width=0.5\textwidth]{segment_adjustment_algorithm}
\caption{Illustration of the iterative segment adjustment algorithm.}
\end{figure}

In a second time, the laser beam is temporally sampled (1 GHz). For a fixed time $t$, the intersection between each discretized ray and the roof is computed. Then, the associated transmitted incident power for each cell can be calculated:

$$P_r(t, \tau) = P_r(t) c(R) A(\tau) \exp\left(\frac{\alpha^2}{2\sigma_B^2}\right) \delta(\tau),$$

where $d$ is the orthogonal distance between the laser beam axis and the intersection between the discretized ray and the roof, $c(R)$ is the amplitude diminution factor due to the distance $R$ between the sensor and the roof. For each discretized ray integration, the associated backscattered power $P_{\text{rb}}$ can be calculated by using the simplified lidar equation (Mallet and Bretar, 2009):

$$P_{\text{rb}} = \int K'' \cos(\alpha) R^2 \frac{cm}{\text{Cd}} P_r(t) dt,$$

where $K''$ is a constant which depends on the wavelength and the aperture of the laser. By an integration over time, the backscattered power of each laser beam is known. With the fixed orientation, different positions of the building edge are simulated for the given ellipsis position. Calculated backscattered powers $P_{\text{rb}}$ are compared to the measured backscattered power $P_r$. The position of the good simulated building edge is obtained when the sign of ($P_{\text{rb}} - P_r$) change. The adjusted point is positioned in the orthogonal position to the laser beam axis on the good simulated position of the building edge.

\begin{figure}[h]
\centering
\includegraphics[width=0.5\textwidth]{segment_adjustment_results}
\caption{Adjustment results. Extracted points before adjustment (red), after adjustment (green). Segment RMS decreases from 0.37 to 0.27 m.}
\end{figure}
vertical distance between each side of these edges is not important enough.
Segment adjustment results are less exhaustive (about 5%). The major parts of segments cannot be adjusted because the one material planar roof is a too restrictive assumption. However, when the procedure is possible, the extraction is improved: the RMS increases of 0.15 m, which is satisfactory with respect to the footprint radius (0.4 m).

Figure 7 provides an overview of the full process applied to the whole dataset. One can notice that once results from different strips are merged, building are correctly outlined.

![Figure 7: Extracted feature before vegetation filtering: one colour per strip.](image)

7 CONCLUSION

Building edge detection using airborne full-waveform lidar is a problem which have never been addressed with real data before. The approach described in this paper proposes a solution providing reliable boundaries segments with very low false positive rate. We have tried to benefit from the particular specification of the acquisition process with 0.8 m footprint size, delivering a large number of waveforms with multiple and well separated echoes. The algorithm is decomposed into several independent parts that can be easily tuned to fit to the specification of other surveys. The main concern of the paper was to provide fast methods to be able to efficiently process large amount of data.

The first limitation of the proposed approach is that only segments are considered. However, other primitives (such as circles, curves) may be inserted for specific landscapes. One just has to add a model selection step within the RANSAC procedure. Furthermore, the adjustment step is limited because of the too strict roof modelling. Future work focusing therefore in improving the roof reflection model, by providing a more complete mathematical formulation. This would increase the rate of adjusted primitives, and being quantitatively assessed with simulated datasets. Finally, it would be interest to add a final step that would turn the set of adjusted primitives into polygons to obtain geometric forms similar to cadastral data.

References


