

Extraction of Buildings Footprint from LiDAR Altimetry Data with the Hermite Transform

José Luis Silván-Cárdenas¹ and Le Wang^{2,*}

¹ Centro de Investigación en Geografía y Geomática “Ing. Jorge L. Tamayo” A.C.
Contoy 137, Lomas de Padierna, Tlalpan, Mexico D.F. 14240

jlsilvan@centrogeo.org.mx

<http://www.centrogeo.org.mx>

² Department of Geography, The State University of New York
105 Wilkeson Quad, Buffalo, NY 14261

lewang@buffalo.edu

<http://www.buffalo.edu>

Abstract. Building footprint geometry is a basic layer of information required by government institutions for a number of land management operations and research. LiDAR (light detection and ranging) is a laser-based altimetry measurement instrument that is flown over relatively wide land areas in order to produce digital surface models. Although high spatial resolution LiDAR measurements (of around 1 m horizontally) are suitable to detect aboveground features through elevation discrimination, the automatic extraction of buildings in many cases, such as in residential areas with complex terrain forms, has proved a difficult task. In this study, we developed a method for detecting building footprint from LiDAR altimetry data and tested its performance over four sites located in Austin, TX. Compared to another standard method, the proposed method had comparable accuracy and better efficiency.

Keywords: Building footprint, Hermite Transform, Local Orientation.

1 Introduction

Remotely sensed data has become a primary source of information for a number of land management and research activities. The increased spatial resolution of remote sensing data has made it possible to produce detailed inventories of above ground features, such as buildings. Unfortunately, the production of such inventories still relies on much on-screen visual interpretation by human experts. The automatization of such processes is of great value for large-scale applications.

Airborne light detection and ranging (LiDAR) is a technology used routinely for producing high-spatial resolution digital terrain models. LiDAR systems deliver irregularly spaced 3-D points of ground and nonground surfaces. LiDAR measurements have several advantages over traditional aerial photographs and

* This study was partly supported by NSF grants (BCS-0822489 and SEB-0810933) and by CentroGeo.

satellite images because they are not influenced by sun shadow and relief displacement [10]. However, datasets tend to be voluminous and not suitable for automated extraction of building information, mainly because many raster image processing techniques cannot be directly applied to irregularly spaced points. To circumvent such limitation, elevation values are usually rasterized. Once LIDAR measurements are in raster format, the problem consists in discriminating building cells. Such a processing serves as a precursor to form building footprint polygons that can be incorporated in a geographic information system.

Two approaches are often utilized to detect building cells from gridded elevation measurements. One is to apply a classification method to separate the ground, buildings, trees, and other features simultaneously [2]. The more popular way is to separate the ground from nonground LIDAR measurements first and then identify the building points from nonground measurements [8,10]. The proposed method followed the latter approach, building upon a prior work on ground filtering [5]. The rest of the paper presents the theoretical background (Section 2), the method description (Section 3), some results from building detection tests (Section 4) and conclusions (Section 5).

2 Background

This section summarizes basic theoretical results that are relevant for the proposed building detection method, as well as a standard method used for comparison. The reader is referred to the original sources for further details.

2.1 Discrete Hermite Transform

The DHT of a two-dimensional signal $z : \mathcal{G} \mapsto \mathbb{R}$ defined on a grid $\mathcal{G} \subseteq \mathbb{Z}^2$, is comprised by filtered and downsampled versions of the original signal, i.e.,

$$z_{n,m}(p, q) = \sum_{(x,y) \in \mathcal{G}} z(x, y) b_n(x - 2p) b_n(y - 2q) \tag{1}$$

for $n, m = 0, \dots, N$, where the analysis functions

$$b_n(x) = 2^{-N} \sqrt{C_N^n} \Delta^n \left\{ C_{N-n}^{x+N/2} \right\} \tag{2}$$

for $x = -N/2, \dots, N/2$, correspond to discrete approximations of Gaussian derivatives. The discrete counterpart of the derivative operator corresponds to the forward difference denoted by Δ .

The signal is reconstructed from the DHT representation above as follows:

$$z(x, y) = \sum_{n,m=0}^N \sum_{(p,q) \in \mathcal{G}} z_{n,m}(p, q) \tilde{b}_n(2p - x) \tilde{b}_n(2q - y) \tag{3}$$

where the synthesis function are given by $\tilde{b}_n(x) = 2b_n(-x)$.

The coefficient $z_{n,m}(p, q)$ approximates (up to a normalization factor) the partial derivative of order n with respect to x and order m with respect to y of a Gaussian-smoothed version of the signal z at the location $(p, q) \in \mathcal{G}$. The degree of smoothness is controlled through the standard deviation of the Gaussian function, given by $\sigma = \sqrt{N}/2$. In practice, the parameter N takes the values of 2, 4, 6 or 8, whereas a larger degree of smoothing can be achieved through a multiscale DHT, as described in [4].

Rotated DHT. The rotated DHT [3,4] is defined in terms of derivatives with respect to a coordinate system (u, v) , that has been rotated by an angle θ with respect to the original coordinate system (x, y) . The rotated coefficients at the sampling location (p, q) , here denoted by $z_{n,m}^{(\theta)}(p, q)$, are expressed as a linear combinations of the original (non-rotated) coefficients $z_{n,m}$. More specifically,

$$\frac{z_{n-m,m}^{(\theta)}(p, q)}{\sqrt{C_n^m}} = \sum_{k=0}^n A_{m,k} \frac{z_{k,n-k}(p, q)}{\sqrt{C_n^k}} \tag{4}$$

where the coefficients $A_{m,k}$ correspond to the generalized binomial filters (GBF), a family of discrete sequences with parameters n and θ , which are given by

$$A_{m,k} = s^k c^{-k} \Delta^m \{ C_{n-m}^{k-m} c^{2k-m} s^{n-2k+m} \} \tag{5}$$

for $m, k = 0, \dots, n$, and $c = \cos(\theta)$ and $s = \sin(\theta)$. The first few GBF can be expressed using the matrix notation $\mathbf{A}_n = [A_{m,k}]_{m,k=0,\dots,n}$ as

$$\begin{pmatrix} s & c \\ c & -s \end{pmatrix}, \begin{pmatrix} s^2 & 2sc & c^2 \\ sc & c^2 - s^2 & -sc \\ c^2 & -2sc & s^2 \end{pmatrix}, \begin{pmatrix} s^3 & 3s^2c & 3sc^2 & c^3 \\ s^2c & -s^3 + 2sc^2 & -2s^2c + c^3 & -sc^2 \\ sc^2 & -2s^2c + c^3 & s^3 - 2sc^2 & s^2c \\ c^3 & -3sc^2 & 3s^2c & -s^3 \end{pmatrix}$$

$$\begin{pmatrix} s^4 & 4s^3c & 6s^2c^2 & 4sc^3 & c^4 \\ s^3c & -s^4 + 3s^2c^2 & -3s^3c + 3sc^3 & -3s^2c^2 + c^4 & -sc^3 \\ s^2c^2 & -2s^3c + 2sc^3 & s^4 - 4s^2c^2 + c^4 & 2s^3c - 2sc^3 & s^2c^2 \\ sc^3 & -3s^2c^2 + c^4 & 3s^3c - 3sc^3 & -s^4 + 3s^2c^2 & -s^3c \\ c^4 & -4sc^3 & 6s^2c^4 & -4s^3c & s^4 \end{pmatrix}$$

for $n = 1, 2, 3$ and 4 respectively.

In all the examples presented here, the rotation was set to $\theta = \arctan(z_{0,1}/z_{1,0})$, which makes the rotated coefficients $z_{0,1}^{(\theta)} = 0$ and $z_{1,0}^{(\theta)} = g$, where $g = \sqrt{z_{1,0}^2 + z_{0,1}^2}$ is proportional to the gradient magnitude.

2.2 Region Growing Segmentation

The region-growing segmentation (RGS) method, as applied for building detection in [10], was used as a benchmark for the proposed building detection

method. This RGS is an iterative method that applies a plane-fitting technique to grow regions from seeds.

The RGS algorithm requires a non-ground mask of grided elevation values. Then, for each nonground measurement area, inside and boundary cells are identified. If at least one of the eight neighbors of a cell is a ground measurement, the cell is defined as a boundary cell. Otherwise, the cell is an inside cell. The following residual is calculated for each inside cell $p_0(x_0, y_0, z_0)$ and its eight neighbors:

$$R = \sum_{k \in M} [a(x_k - x_0) + b(y_k - y_0) + c - z_k]^2 \quad (6)$$

where M is a set for the inside cell and its neighbors, and a, b and c are plane parameters estimated through least square. The cell with the minimum residual R is labelled and selected as the first seed cell for region growing. All neighbors of a seed cell are labelled as belonging to the same segment if the deviation between its height and the plane height is under a threshold. A threshold of 0.1 m was used in our implementation. The plane parameters are then updated including the new labelled cells. The neighbors of the grown area are examined further, and the process is continued until no additional cells can be added into the segment. Then, the unlabelled cell with the minimum R is selected as the next seed. The process is repeated until all nonground cells are labelled.

After the RGS algorithm was run (and following [10]), small segments (with less than five pixels) were removed, holes were filled and contiguous segments were merged to form building footprints.

3 Building Detection

This section provides the details of the proposed building detection method, which can be summarized in two steps. In the first step, a height map for above-ground features is produced. This step requires subtracting the terrain component $t(x, y)$ from the elevation surface $z(x, y)$. The second step consists on separating building cells from vegetation cell on the height map. A description of each step is presented bellow.

3.1 Aboveground Features Height

The method used here for generating the terrain component is based on the so-called ground filtering method introduced in [5], which in turn was based on a multiscale implementation of the DHT [4]. Let $\{z_{n,m}^{(k)}\}$ denote the multiscale DHT coefficients of the grided DSM $z(x, y)$ at a generic lattice point, and let $g^{(k)}$ denote the gradient magnitude at the scale index k , for $k = 0, \dots, K - 1$, where K is the number of pyramid layers of the multiscale DHT. It can be shown that the coefficients

$$\hat{t}_{n,m}^{(k)} = \begin{cases} z_{n,m}^{(k)} & \text{if } g^{(k)} \leq T^{(k)} \\ z_{0,0}^{(k)} - g^{(k)} / \sqrt{2} & \text{if } g^{(k)} > T^{(k)} \text{ and } n, m = 0 \\ 0 & \text{if } g^{(k)} > T^{(k)} \text{ and } n, m > 0 \end{cases} \quad (7)$$

reconstruct a signal $\hat{t}(x, y)$ that approximates the terrain elevation surface, provided that the multiscale gradient thresholds are selected according to

$$T^{(k)} = \frac{2^k m_{max}}{\sqrt{2 + 2\pi(2^k m_{max}/\Delta_{max})^2}} \quad (8)$$

where m_{max} and Δ_{max} are the maximum terrain slope and maximum terrain elevation difference in the site.

In all the tests presented bellow, the number of pyramid layers K of the multiscale DHT decomposition was determined as

$$K = \lceil \log_2(L_{max}/2\delta) \rceil \quad (9)$$

where L_{max} denote the maximum length of aboveground features and δ is the cell size of the gridded elevation values. This number of layers assured that large aboveground features were effectively removed.

In the ground mask, a cell (x, y) is assumed nonground if $z(x, y) - \hat{t}(x, y) > \epsilon$, where ϵ was set to 0.1 in all tests performed here. Once the ground mask is build, elevation values of detected nonground cells can be interpolated from elevation values of surrounding ground cells to produce a more accurate terrain component $t(x, y)$. Finally, the feature height map is computed as:

$$h(x, y) = z(x, y) - t(x, y) \quad (10)$$

3.2 The Planar Roof Model

The rationale here is that most roofs are composed of strongly oriented, mainly planar, surfaces, whereas forested areas are not. Hence, the rotated DHT coefficients along the local gradient is essentially distinct for building roofs and trees.

Let $h_{n,m}^{(\theta)}$ denote the rotated DHT coefficients along the local gradient of the feature height map. Then, building cells can be separated from vegetation cells by thresholding the residual energy term

$$E = \sum_{i=2}^N \sum_{j=1}^i \{h_{i-j,j}^{(\theta)}\}^2 \quad (11)$$

This residual energy measures the degree to which the local pattern does not conform to a one-dimensional signal (such as a planar surface). As it turns out, E is insensitive to planar roofs because the scale-space derivatives are only sensitive to polynomial variations of the derivation order or above.

In all the tests performed in this study an empirical threshold of 0.15 was used. Building masks so-produced were filtered in a similar fashion as the region-growing segmentation masks, so that small segments were eliminated and holes filled.

4 Results

The datasets used here consisted of grided elevation values from the last return of a laser pulse (Fig. 1, top row) and the actual building footprint from visual interpretation of aerial photography (Fig. 1, 2nd row). The study selected four representative sites of Austin City in Texas. Further descriptions of acquisition and preprocessing of the datasets are provided in [6]. The building detection results for each inset are illustrated in Fig. 1 (3rd and 4rd rows). These error maps were built through comparing the detection mask from each method with the actual building footprint layer in raster format. Errors of omission and commission are colored with blue and red for easy identification.

The overall per-pixel accuracy, the kappa statistics, the detection rate, and the commission error [7] were calculated for each method and inset. These results are provided in Table 1. As observed in these table, there was some accuracy variability across insets. Specifically, inset 1 was the most accurately classified into building and non-building by both methods with a per-pixel accuracy around 95%. This was due to the relatively high and large structure of multifamily buildings. On the other hand, insets 2 and 4 represented the most challenging area due to the relatively small size of single-family buildings and the high chance of occlusions by trees. In this case, the largest accuracy was under 90%. On the other hand, no significant differences in accuracies from the DHT and the RGS methods existed. In any case, differences did not always favored one method consistently, so that in the average both methods performed comparably. However, the most significant advantage of DHT over RGS was the saving in computation time, which was in the order of several hundred of times (data not shown). This is because the plane-fitting technique employed by RGS requires multiple matrix inversions, whereas the DHT method mainly involves convolutions operations which are computed efficiently [4].

Table 1. Two level accuracy assessment of building detection. Statistics were based on four insets, each of which had 145, 476, 427 and 490 buildings, respectively. Values were rounded to the nearest integer.

Method	Dataset	% Pixels		% Objects	
		Overall Acc.	Kappa	Detection Rate	Commission
DHT	Inset 1	95.0	80.9	77.9	18.1
	Inset 2	88.5	51.5	80.7	22.4
	Inset 3	90.8	56.3	90.9	26.3
	Inset 4	88.9	61.0	89.8	18.3
	Average	90.8	62.4	84.8	21.3
RGS	Inset 1	94.9	81.3	81.8	29.8
	Inset 2	89.0	51.6	73.0	12.4
	Inset 3	92.0	62.4	90.1	19.6
	Inset 4	88.8	60.5	88.4	14.6
	Average	91.2	63.9	83.3	19.1

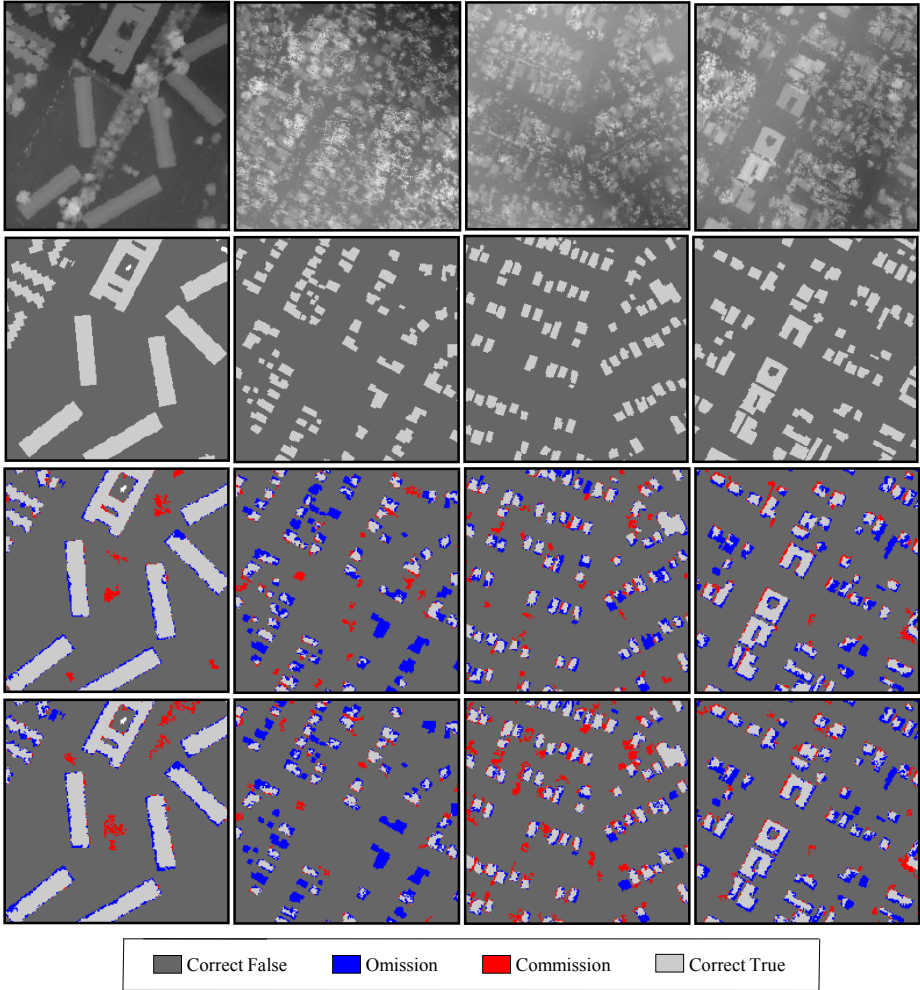


Fig. 1. Building detection results. Rows from top to bottom correspond to original gridded LiDAR data, actual building footprint, error map from RGS, and error maps from DHT, whereas columns from left to right correspond to subsets from inset 1 through inset 4, respectively.

5 Conclusions

This study proposed and tested the DHT as a tool for building extraction from gridded LiDAR data. The proposed building detection method used a multi-resolution ground filtering method based on the multiscale DHT, which is

efficiently computed [5]. The detection of buildings consisted on a simple thresholding of an energy term of the rotated DHT. Results indicated that the DHT building detection method competes with a more traditional method based on plane-fitting region growing segmentation. The appealing advantage of the proposed approach seemed to be its computational efficiency, which is crucial for large scale applications. For instance, this technique can be used for small-area population estimation as in [6]. Further research should explore the optimality of parameter selection for both the ground filtering and the energy thresholding. Also, building occlusions by trees represent a big challenge, which demands alternate approaches. One of such alternatives would be the partial active basis model presented in [9,1], where Gabor wavelet elements could be replaced by Gaussian derivatives.

References

1. Herrera-Domínguez, P., Altamirano-Robles, L.: A Hierarchical Recursive Partial Active Basis Model. *Advances in Pattern Recognition*, 1–10 (2010)
2. Miliareisis, G., Kokkas, N.: Segmentation and object-based classification for the extraction of the building class from LIDAR DEMs. *Computers & Geosciences* 33(8), 1076–1087 (2007)
3. Silván-Cárdenas, J.L., Escalante-Ramírez, B.: Image coding with a directional-oriented discrete hermite transform on a hexagonal sampling lattice. In: Tescher, A. (ed.) *Applications of Digital Image Processing XXIV*, vol. 4472, pp. 528–536. SPIE, San Diego (2001)
4. Silván-Cárdenas, J.L., Escalante-Ramírez, B.: The multiscale Hermite transform for local orientation analysis. *IEEE Transactions on Image Processing* 15(5), 1236–1253 (2006)
5. Silván-Cárdenas, J.L., Wang, L.: A multi-resolution approach for filtering LiDAR altimetry data. *ISPRS Journal of Photogrammetry and Remote Sensing* 61(1), 11–22 (2006)
6. Silván-Cárdenas, J., Wang, L., Rogerson, P., Wu, C., Feng, T., Kamphaus, B.: Assessing fine-spatial-resolution remote sensing for small-area population estimation. *International Journal of Remote Sensing* 31(21), 5605–5634 (2010)
7. Song, W., Haithcoat, T.: Development of comprehensive accuracy assessment indexes for building footprint extraction. *IEEE Transactions on Geoscience and Remote Sensing* 43(2), 402–404 (2005)
8. Weidner, U., Förstner, W.: Towards automatic building extraction from high-resolution digital elevation models. *ISPRS Journal of Photogrammetry and Remote Sensing* 50(4), 38–49 (1995)
9. Wu, Y., Si, Z., Gong, H., Zhu, S.: Learning active basis model for object detection and recognition. *International Journal of Computer Vision*, 1–38 (2009)
10. Zhang, K., Yan, J., Chen, S.: Automatic construction of building footprints from airborne LIDAR data. *IEEE Transactions on Geoscience and Remote Sensing* 44(9), 2523–2533 (2006)