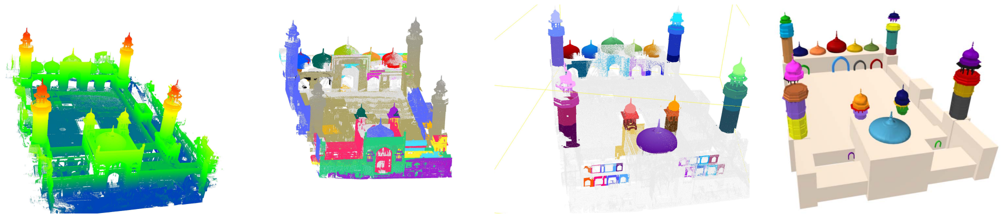
Volume 36 (2017), Number 7

Outdoor scene segmentation and reconstruction using LiDAR data

O. Hassaan, A. Shamail, Z. Butt, M. Taj

Department of Computer Science, Syed Babar Ali School of Science and Engineering
Lahore University of Management Sciences, Pakistan



(a) (b) (c) (d)

Figure 1: *Example showing detailed segmentation and reconstruction. (a) Raw point cloud (b) Coarse segmentation (c) Hierarchical segmentation (d) Generated 3D CAD-based Model.*

Abstract

Recent advancements in 3D scanning technologies have paved way for generation of highly accurate 3D scenes in the form of point cloud data. For the segmentation and reconstruction of such scenes, a number of techniques have been introduced in literature. Our approach is a hybrid technique for the segmentation and 3D reconstruction of LiDAR point cloud data, primarily focusing on 3D outdoor scenes. Our model, the hierarchical tree, iteratively divides the point cloud into segments, simultaneously using a novel energy function and a 3D convolutional neural network, HollowNets to classify the segments. We have tested the efficacy of our proposed approach using real data of two sites obtained from Leica Scan Station P20 Terrestrial Laser Scanner and found an accuracy of 95% *for domes and minarets.*

1. Introduction

Recent advancements in 3D scanning technologies have paved way for generation of highly accurate point cloud data. The increasing use of LiDAR along with photogrammetry has enabled large-scale reconstruction, such as an entire housing block [LGZ*13] and a coarse model of a city [LM12]. A number of techniques, such as plane based [MMBM15], segmentation by model fitting, machine learning based segmentation [A*16], and Manhattan world assumptions [BGT16], exist for the segmentation and classification. Many of these approaches, however, are confined to symmetric and planner surfaces and therefore cannot be generalized.

We focus on 3D outdoor scenes consisting of planar as well as non-planar structures. Our model is a hierarchical tree that iteratively divides the point cloud into smaller components using a greedy approach which uses an automated energy function that estimates the correctness of segmentation. For the purpose of segmentation and detection, the energy function uses a Convolution Neural Network (CNN) called HollowNets which is applied on surface sampled points (voxels).

2. Methodology

We perform initial segmentation by using Schnabel's algorithm [SWK07]. To resolve the problem of over & under segmentation, we first apply spatial clustering and then, using the tree methodology, three different projection sequences are applied on each segment. Each projection sequence converts input point cloud into a low-dimensional signal using four projections; namely vertical ($\upsilon_1 \& \upsilon_2$), horizontal (h), circular profile (p) and circular unwarping (u). Figure 2 shows a flow diagram of our proposed approach.

Sequence S_1: $V_i^1 = \rho(\upsilon_2(p(\psi(\upsilon_1(v_{i-1,k}^m)))))$
Sequence S_2: $V_i^2 = \rho(\upsilon_2(h(v_{i-1,k}^m)))$
Sequence S_3: $V_i^3 = \rho(\upsilon_2(u(\psi(\upsilon_1(v_{i-1,k}^m)))))$

In sequences S_1 & S_3, the projection υ_1 eliminates the vertical dimension. Circular and n-gonal RANSAC $\psi(.)$ are then applied on the projected data to recover the location, position and orientation. Similarly, in sequence S_2 the projection h eliminates one of the vertical dimensions. Finally, peak finding $\rho(.)$ is performed to obtain the segments. To estimate their correctness, we propose an

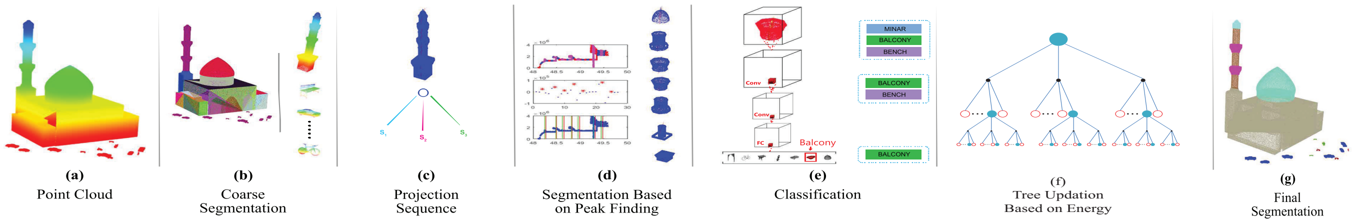


Figure 2: *Flow diagram of the proposed approach. (a) Input point cloud. (b) Coarse segments obtained using RANSAC [SWK07] and spatial clustering. (c) Projection sequence. (d) Peak finding on profile curve of minar showing 7 peaks as asterix and vertical line. (e) Classification of one of the segments. (f) Illustration of tree being updated as more segments being added. (g) Obtained segments.*

energy function based on the following 5 criteria:

$$\text{uniformity of segmentation } e_1 = 1 - \frac{\sigma([N_1, N_2, \ldots, N_s])}{\sigma([1, N_p - 1])}, \quad (1a)$$

$$\text{parent population } e_2 = \min(N_p, N_{min})/N_{min}, \quad (1b)$$

$$\text{number of segments } e_3 = 1 - \sim \mathcal{N}(s|\mu, \sigma^2), \quad (1c)$$

$$\text{prior probability } e_4 = W(r_{id}(v^m_{i-1,k}), seq.id), \quad (1d)$$

$$\text{classification score } e_5 = \frac{1}{s} \sum r_{scr}(v^n_{i,j}),\ j \in \{1, 2, \ldots, s\}, \quad (1e)$$

where, N is the no. of points, r_{scr}, r_{id} are classification score and id respectively, s is the no. of segments and W is 6×3 matrix of prior probabilities.

The projection sequences thus obtained are used as edges (E) of a hierarchical tree to iteratively divide the point cloud into segments (V). We recursively build our tree $G\{V, E\}$, which is a top-down model of hierarchical segmentation rather than a bottom-up hierarchy of planar patches [LGZ*13]. The weight of each of these edges is computed using an objective function $\xi(.)$ defined as:

$$\xi(v^n_{i,j}, v^m_{i-1,k}) = \omega^T \varepsilon, \quad (2)$$

where $\omega = \{\omega_1, \cdots, \omega_5\}$ are the weights, obtained using simple linear regression, for each of the five energy terms $\varepsilon = \{e_1 \cdots, e_5\}$. The node $v^n_{i,j}$ is the j^{th} segment of the i^{th} iteration (tree depth) obtained via n^{th} projection sequence such that $v^n_{i,j} \subset v^m_{i-1,k}$ and $v^m_{i-1,k}$ is the parent node of v_i. Also $v^m_{i-1,k} = \{\cup v^n_{i,j}\}$ where, for all segments s, $j \in \{1, 2, \ldots, s\}$, $n \in \{1, 2, 3\}$. The set V^n_i contains all segments of $v^m_{i-1,k}$ obtained via n^{th} projection sequence.

Instead of commonly used voxel representation as input to CNN, we propose a *hollow* voxel where each (x, y, z) point is mapped to (i, j, k) index of a 3D regular grid (see Fig. 3). For our problem, we have chosen 6 classes each having 300 and 50 training and testing samples, respectively collected from 3D Warehouse and ModelNet10 [W*15]. Instead of binary voxels, we scale the values between $[-1, 5]$, thus allowing the network to learn more from positive integer values.

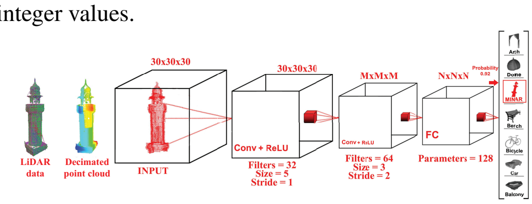


Figure 3: *Hollow voxel & Layered Architecture of HollowNet. Accuracy=99.03%, learning rate $\eta = 0.001$, EPOCHS=1000.*

Site	Dimensions	Points	Arches			Domes			Minartes		
	$L \times W \times H\ m^3$	in Bn	GT	AG	Acc	GT	AG	Acc	GT	AG	Acc
MWK	$91 \times 53 \times 33$	0.288	46	19	41.30	12	12	100	6	6	100
MKA	$60 \times 36 \times 16$	0.548	12	7	58.33	21	19	90.4	0	0	100

Table 1: *Results showing comparison of automatically generated (AG) primitives with ground truth (GT) and accuracy (Acc%).*

3. Results

We evaluated our algorithm on real data which was obtained from Leica Scan Station P20 Terrestrial Laser Scanner. Two sites were scanned using P20 namely Masjid Wazir Khan (MWK) in Lahore, and Masjid Khudabad (MKA) in Dadu. The details of these sites and obtained results are shown in Tab. 1. The higher accuracy in case of domes and minarets is due to their rotational symmetry. Each segment is substituted using an architectural primitive based on the class labels and its size to create a 3D CAD model.

4. Conclusions and Future work

We propose a hierarchical tree model for outdoor scene segmentation and reconstruction. We have tested our approach using two real scenes and have found an accumulative accuracy of 95% for the dome and minaret features. This work is applicable to geometry that exhibits structural regularity. We aim to evaluate our algorithm on different scenes including Derawar Fort and Shiva temple.

References

[A*16] ARMENI I., ET AL.: 3D semantic parsing of large-scale indoor spaces. In *Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition* (Las Vegas, NV, US, Jun 2016). 1

[BGT16] BAJWA R., GILANI S. R., TAJ M.: 3D architectural modeling: Coarse-to-fine model fitting on point cloud. In *Computer Graphics Int.* (Heraklion, GR, Jun 2016), pp. 65–68. 1

[LGZ*13] LIN H., GAO J., ZHOU Y., LU G., YE M., ZHANG C., LIU L., YANG R.: Semantic decomposition and reconstruction of residential scenes from lidar data. *ACM Trans. Graph. 32*, 4 (Jul 2013), 1–10. 1, 2

[LM12] LAFARGE F., MALLET C.: Creating large-scale city models from 3D-point clouds: a robust approach with hybrid representation. *Int. Journal of Comp. Vision 99*, 1 (2012), 69–85. 1

[MMBM15] MONSZPART A., MELLADO N., BROSTOW G. J., MITRA N. J.: RAPter: Rebuilding man-made scenes with regular arrangements of planes. *ACM Trans. Graph. 34*, 4 (Jul 2015), 1–12. 1

[SWK07] SCHNABEL R., WAHL R., KLEIN R.: Efficient RANSAC for point-cloud shape detection. *Computer Graphics Forum 26*, 2 (Jun 2007), 214–226. 1, 2

[W*15] WU Z., ET AL.: 3D ShapeNets: A deep representation for volumetric shapes. In *IEEE Conf. on Computer Vision and Pattern Recognition* (Boston, MA, USA, Jun 2015), pp. 1912–1920. 2