ROBUST AND ACCURATE PLANE SEGMENTATION FROM POINT CLOUDS OF STRUCTURED SCENES

Pingbo Hu [1], Yanan Liu [1,*], Mao Tian [2], Miaole Hou[1]

[1] School of Geomatics and Urban Spatial, Beijing University of Civil Engineering and Architecture, 102616 Beijing, China - (hupingbo, liuyanan, houmiaole)@bucea.edu.cn
[2] State Key Laboratory of Information Engineering in Surveying, Mapping and Remote Sensing, Wuhan University, China - (mtian)@whu.edu.cn

Commission II, WG II/3

KEY WORDS: Point Cloud, Plane, Supervoxel, Segmentation, Graph Cut, Structured

ABSTRACT:

Plane segmentation from the point cloud is an important step in various types of geo-information related to human activities. In this paper, we present a new approach to accurate segment planar primitives simultaneously by transforming it into the best matching issue between the over-segmented super-voxels and the 3D plane models. The super-voxels and its adjacent topological graph are firstly derived from the input point cloud as over-segmented small patches. Such initial 3D plane models are then enriched by fitting centroids of randomly sampled super-voxels, and translating these grouped planar super-voxels by structured scene prior (e.g. orthogonality, parallelism), while the generated adjacent graph will be updated along with planar clustering. To achieve the final super-voxels to planes assignment problem, an energy minimization framework is constructed using the productions of candidate planes, initial super-voxels, and the improved adjacent graph, and optimized to segment multiple consistent planar surfaces in the scenes simultaneously. The proposed algorithms are implemented, and three types of point clouds differing in feature characteristics (e.g. point density, complexity) are mainly tested to validate the efficiency and effectiveness of our segmentation method.

1. INTRODUCTION

Detecting planar surfaces from LiDAR and photogrammetry point cloud, due to its vast applications in many areas, has been an active topic in many research communities (Brook et al., 2013). The segmented planes can be applied to classification, scene understanding, navigation, and the building information model (BIM) reconstruction, but a poor segmentation can make these tasks fail. During past decades, many algorithms and systems have been proposed to the plane segmentation based on the type of input data and objects, making the production of segmentation faster and better. Even though much progress has been successfully achieved, the robust and accurate 3D plane segmentation from the point cloud remains to be a challenging issue, especially for the complex scenes with noise, outliers, and occlusions. Besides, such a process on the acquired massive point clouds can be quite a time consuming, and information of surfaces, boundaries, scene priors (e.g. orthogonality, parallelism) are not preserved or even extracted. Thus, this paper proposes a robust and efficient unsupervised method to the segmentation of point cloud acquired from structural scenes.

2. RELATED WORK

The issue of plane segmentation has received considerable attention in the area of photogrammetry, computer vision, and autonomous vehicles. Within these large bodies of work in this broad topic, the research involving the plane extraction or segmentation referred to the scope of this paper is reviewed.

2.1 Supervised Methods

The supervised 3D plane segmentation, especially using joint segmentation and recognition, has aroused great interest along with the machine learning and deep learning. Similar to the 2D semantic labeling technology, the 3D method learns a classification model from the training data to predict the semantic category of each 3D element (e.g. 3D point, patches). The graphical model like Conditional Random Fields (CRF) is always employed to capture scene features and different categories (Pham et al., 2015; Vosselman et al., 2017). This encoded 3D contextual information hinders its wide application to construct and optimize such a complex graphical model. Recently, deep learning-based approaches (Kong et al., 2019; Milioto et al., 2019) can directly achieve the semantic information without feature calculation and can obtain a state-of-the-art result. However, the main limitations are the huge training samples and weak network migration capabilities between different layered architectures.

2.2 Unsupervised Methods

An extensive literature has been proposed to improve the robustness and efficiency of planar segmentation, which can be roughly categorized into four categories: model fitting-based methods (Chen et al., 2014; Schnabel et al., 2007), region growing-based methods (Deschaud et al., 2010; Vo et al., 2015; Vosselman et al., 2017; Yang and Dong, 2013), based feature clustering-based methods (Kim et al., 2016; Zhou et al., 2016), and energy optimization-based methods (Dong et al., 2018; Pham et al., 2016; Yan et al., 2014).

Model fitting-based methods The early model fitting-based methods approaches, Hough Transform (HT) by Duda and Hart (1972) and Random Sample Consensus (RANSAC) proposed by Fischler and Bolles (1981), are widely employed and have been proven to successfully extract 2D and 3D elements (Schnabel et al., 2007). Although these approaches and

---

* Corresponding author

improvements have achieved satisfying 3D plane segmentation results, it always fails as the sensitive model parameters with noise and outliers.

Region growing-based methods These methods for 3D plane extraction is an iteration process by progressively merging adjacent points or patches with similar feature characteristics. It starts with potential seeds and then expands to its neighboring points. Nevertheless, it is susceptible to the seeds selection, and difficult to terminate when the transitions between the two regions are smooth (Sampath and Shan, 2010).

Feature clustering-based methods The statistical method classifies the point clouds into primitives based on fixed pre-calculated local feature properties (e.g. saliency feature). The clustering (Vo et al., 2015) in the feature space excluded the boundary points, thus refinement was needed to test whether the points were within the same cluster space (Zhou et al., 2016). Despite the popularity and efficiency of this approach, it suffers the difficulty in neighbourhood definition and is sensitive to noise and outliers.

Energy optimization-based methods The widely used energy minimization approach is a global optimization solution by constructing a stable plane energy model. It aims at fidelity data, continuity of feature values, and compactness of segment boundaries (Kim and Shan, 2011). The widespread applications of energy-based methods in the field of 2D image process can be found in (Dong et al., 2018; Hossam Isack, 2012; Pham et al., 2014). These methods are robust, and can produce spatially coherent plane models, and improve the quality of plane extraction. However, the energy optimization methods are computationally expensive for the huge point cloud, and are greatly affected by the adequacy and reliably of initial inputs.

Even though these proposed methods can generally provide satisfactory extraction results, there still exist limitations to extract primitives from point clouds, especially for the complex structural objects with occlusion and bias. To overcome these problems, this paper develops a simple segmentation strategy that is to transform the plane segmentation issue into the best matching issue between the over-segmented super-voxels and the 3D plane models, improving the robustness to noise and the efficiency of global optimization.

The remainder of the paper is structured as follows. In Section 3, the details for plane segmentation is carried out, and results including assessment and discussion are presented in Section 4. Section 5 consists of the concluding remarks on the introduced method and future effort.

3. METHODOLOGY

The proposed approach, as illustrated in Figure 1, aims to extract planes from LiDAR or photogrammetric point clouds of a structured scene by best matching the generated super-voxels to the potential planes. It encompasses three key components, namely super-voxel segmentation (**Preprocessing**), candidate planes and geometry relationship generation (**Initialization**), and the super-voxels to planes assignment using the graph-based optimization framework (**Optimization**).
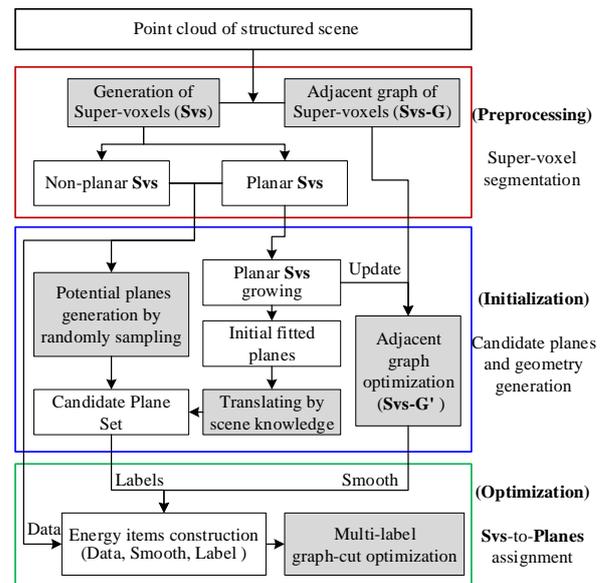

Figure 1. The illustration of the proposed method

The structured scenes are a dominant element of three-dimensional modelling, which exists lots of planar surfaces and valuable scene priors as orthogonality, parallelism. Taking the point cloud of a structured scene as input, an over-segmentation preprocessing (Section 3.1.1) is introduced to segment the point clouds into super-voxels (Svs) and its adjacent geometric relations graph (Svs-G), then voxels can be classified as planar or non-planar segments with its geometrical characteristic (Section 3.1.2). To enrich the potential candidate planes, a randomly sampling strategy (Section 3.2.1) and a scene prior-based translating of fitted grouped planar Svs (Section 3.2.2) is adopted, which can fill some issues like occlusion, noise, and outliers. The final super-voxels to planes assignment (planes segmentation) can be optimized by a multi-label graph-cut framework (Section 3.3), where these energy items are constructed by these candidate planes, original super-voxels, and its adjacent graph (**Svs-G'**).

3.1 Preprocessing for Supervoxels Segmentation

In this section, we introduce an over-segmentation approach to handle massive points and then extract the planar super-voxels using its geometrical characteristics.

3.1.1 Super-voxel Generation

Plane segmentation from the original point cloud (neither LiDAR or photogrammetric point cloud) is time-consuming, and the direct handle will increase the computational cost. Thus, we represent these several millions of points by a collection of small patches, named super-voxels (**Svs**), to reduces the processing complexity. To obtain the super-voxels, a planar over-segmentation approach (Papon et al., 2013) is adopted, producing a set of small patches marked as $Svs = \{sv_i\}$. Each super-voxel has similar geometric features and can be formulated by centroid c_i, curvature f_i, and normal vector n_i. In addition, an adjacent connected graph (**Svs-G**) for the super-voxels is constructed between super-voxels. The graph can ensure super-voxels do not flow across object boundaries and can be efficiently used for further searching. A vertex (sv_i) in the graph (**Svs-G**) is an individual super-voxels, and a connected edge (e_i) is linked with two adjacent vertexes. Thus,

the input point cloud can be finally recorded as **Svs-G** = {sv_i, e_i}. In some cases, the colour is always missed or uncorrected, thus we focus on geometric features, and take the spatial distance, and normal vector deviation to generate **Svs-G**.

3.1.2 Geometric Features Calculation

Each generated super-voxel (**Svs**) is most likely to be part of a plane. To achieve a precise plane from super-voxels, the proposed method firstly calculates the saliency geometric features (Yang and Dong, 2013) of each super-voxel based on the formula as follows:

$$\alpha_1 = \frac{\sqrt{\lambda_1 - \lambda_2}}{\lambda_1}, \ \alpha_2 = \frac{\sqrt{\lambda_2 - \lambda_3}}{\lambda_1}, \ \alpha_3 = \frac{\sqrt{\lambda_3}}{\lambda_1} \quad (1)$$

Where the value of λ_i is the singular values in descending order. Then each **Svs** can be classified into planar surface based on the following formula Eq. (2).

$$Svs = (\alpha_2 \geq s_\alpha \cdot \alpha_1) \bigcap (\alpha_2 \geq s_\alpha \cdot \alpha_3) \quad (2)$$

Where s_α is scaling factors defining relative tolerances for the acceptable amount off-plane displacement, and can empirically be set to (0.75-1.3) for the extraction of non-planar and planar super-voxels.

3.2 Generation of Candidate Planes and Geometry

Accurate and reliable candidate planes and its corresponding geometric connection (Monszpart et al., 2015) are the key factors for plane extraction. In this section, we will introduce the proposed approach for candidate planes enrichment, including the fitted planes $pSet_{sam}$ by randomly sampling on subsets of the centroids from the over-segmented super-voxels, and potential planes translating ($pSet_{trans}$) from the combined the grouped planar super-voxels and structured scene prior (e.g. orthogonality, parallelism).

3.2.1 Plane Candidates Fitted by Random Sampling

A fast and simple method to generate candidate planes is to fit randomly sampled minimum subsets from the original points. However, sampling a large number of minimum point subsets is time-consuming, which can ensure the adequacy of the plane hypothesis set. Instead, we propose to generate candidate plane models by randomly sampling the over-segmented super-voxels, which can provide centroids and normal vectors for the further RANSAC or Least-Squares plane fitting.

3.2.2 Candidate Planes Translation using Scene Priors

For such structured scenes, there are always a large number of planar surfaces and valuable scene priors (e.g. orthogonality, parallelism), which can be effectively used to enrich potential planes especially in the case of noisy, incomplete, outlier-ridden data. Here a region growing approach (Rabbani et al., 2006) based on curve smooth is applied to group planar super-voxels from the over-segmented **Svs**. It merges two adjacent planar super-voxels (shared a valid edge e_i,) with similar features ({c_i, f_i, n_i,}). The optimal features of the referred super-voxels, as well as the local part of the adjacent graph (**Svs-G**), will be synchronous update, which can avoid the points in the same plane being scattered on different planes as noisy. What's more,

this merging update is local as only edges with the related super-voxels adjacent to the two planar merged are processed. resulting in an updated adjacent graph (**Svs-G'**).

With these grouped planar super-voxels, we first calculate the 3D plane models by least-square estimation. We prefer these planes to be orthogonal or parallel (depends on the structural prior knowledge), thus the potential planes translating will be performed. For each pair of nearly orthogonal or parallel planes (P_A, P_B), we can rotate one plane (P_A) by a fixed angle (e.g. 90°) and translate it to the other (P_B). This translating, which can generate the potential planes, is consistent with the fact that points of P_A can be better explained by plane model P_B.

3.3 Graph-based Optimization for (Svs-to-planes)

The generated candidate planes ($pSet = pSet_{sam} + pSet_{trans}$), super-voxels (**Svs**), and updated adjacent graph (**Svs-G'**) will be used to be optimized by the global energy minimization (Delong et al., 2012) solution as follows:

$$E(L) = \overbrace{\sum_{p \in Svs} D(p)}^{DataItem} + \overbrace{\sum_{\substack{(p,q) \\ \in Svs-G'}} \omega_{pq} \cdot \delta(p,q)}^{SmoothItem} + \overbrace{\sum_{L_i \in pSet} |\zeta_L|}^{LabelItem} \quad (3)$$

Where p and q are the symbolic super-voxel (**Svs**), and L_p is the 3D plane model from the candidates (**pSet**) expressed by:

$$L_p = ax + by + cz + d = 0 \cap (a^2 + b^2 + c^2 = 1) \quad (4)$$

The optimization problem can be resolved via a popular graph cut like the extended α-expansion algorithm (Delong et al., 2012; Isack and Boykov, 2012), which achieves a good balance of data item cost (geometric errors), smooth item cost (spatial coherence), label item cost (number of planes).

The data item $\sum_{p \in Svs} D(p)$ is used to measure the sum of geometric errors using a quadratic perpendicular deviation between super-voxel and plane label L_p as Eq. (5), and the construct quadratic distance is equivalent to the Gaussian distribution of assumed errors.

$$D(p) = \|p - L_p\| = (a \cdot x_p + b \cdot y_p + c \cdot z_p + d)^2 \quad (5)$$

The second smoothness prior term in Eq. (3) assumes some specific neighborhood system edge for the adjacent super-voxels along with the updated adjacent graph (**Svs-G'**). In this paper, the Potts model (Delong et al., 2012) is adopted for the indicator function $\delta(\cdot)$, written by:

$$\delta(L_p \neq L_q) = \begin{cases} 1, \ L_p = L_q \\ 0, otherwise \end{cases} \quad (6)$$

A closer super-voxel is a priori more likely to fit the same plane, thus the weight ω_{pq} is set inversely proportional to the distance of adjacent the super-voxel p and q, as Eq. (7).

$$\omega_{pq} = e^{-\|p-q\|^2} \quad (7)$$

It is encouraging to express structural scenes with fewer planes, resulting in a brief description. Thus, the label item is built by the number of super-voxels for a plane, written by:

$$\left|\zeta_L\right| = e^{-|L_i|} \qquad (8)$$

The final optimization framework for assigning super-voxels to planes can be organized as follows:

$$E(L) = \sum_{p \in Svs} \left\|p - L_p\right\| + \sum_{(p,q) \in Svs - G'} e^{-\|p-q\|^2} \cdot \delta\left(L_p \neq L_q\right) + \sum_{L_i \in pSet} e^{-|L_i|} \quad (9)$$

The proposed energy optimization problem used for plane segmentation can be solved by α-expansion algorithm (Delong et al., 2012; Hossam Isack, 2012). When the global energy is no longer reduced, and the iterative will be terminated, and then 3D plane models can be reconstructed from the optimized labels.

4. RESULTS

The proposed approach was implemented and applied to three data sets (**S1**, **S2**, and **S3**) that differ in density and feature characteristics, and the results of qualitative and quantitative analysis for plane segmentation are explained as follows. The **S1** is a stand-alone building with noises and outliers, while **S2** is a complex roof with different primitive types. Besides, the processed airborne LiDAR point cloud **S3** is obtained from the NYU dataset which is a high-density ALS data for urban areas and contains a complex set of roof types such as multi-layered.
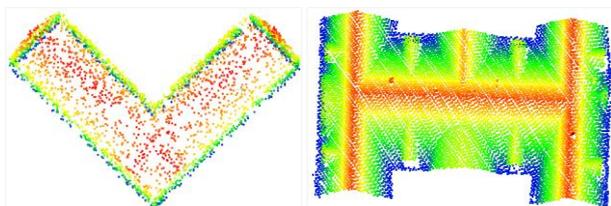

(a) **S1** with noises and outliers   (b) **S2** with different elements
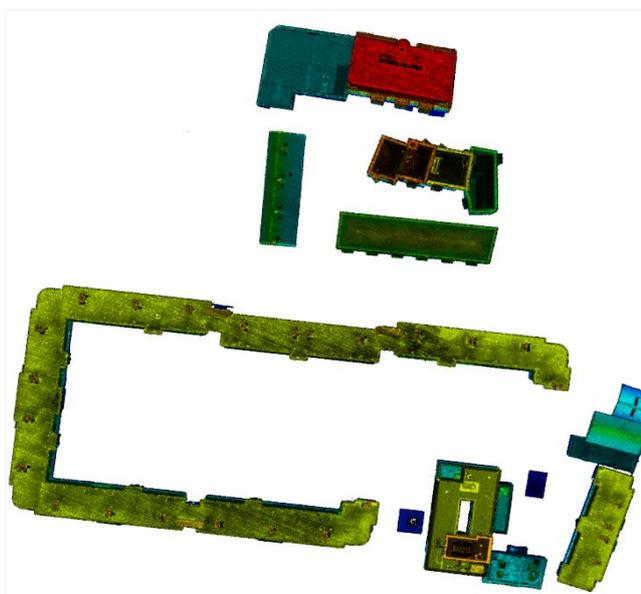

(c) NYU data (**S3**) with complex roof types

Figure 2. Overview of three tested point clouds

Datasets **S1** and **S2** are tested by the proposed approach and compared with the RANSAC. The qualitative results are illustrated in Figure 3, where RANSC can discover the main planes, but failed to with smaller structures and large planes with noises, while the proposed has sucessfully recovered the small patches, espcaily the edges and transition areas.
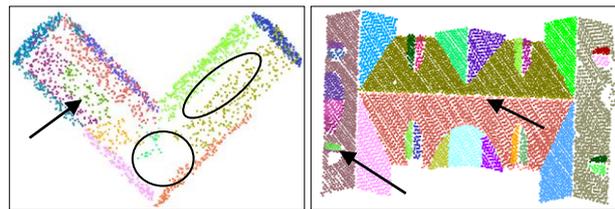

(a) Plane extraction by RANSAC
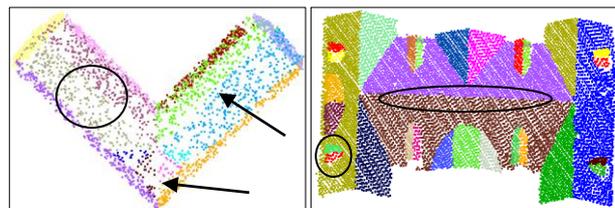
(b) The proposed plane segments

Figure 3. Comparison of plane segmentation from S1 and S2.

An evaluation with a visual inspection for the 3D reconstructed models (**S3**) is shown in Figure 4. The basic planar primitives are well reconstructed including the narrow planes covering multi-layered, overhanging, and multi-layered with flat roofs.
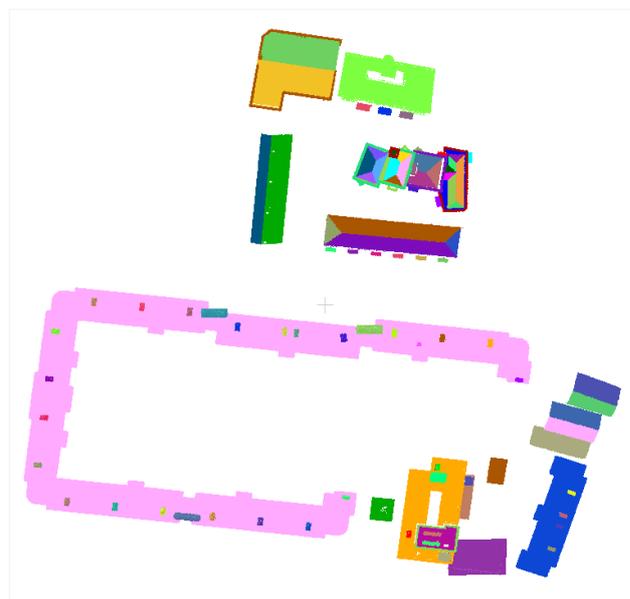
(a) Plane segments by the proposed approach
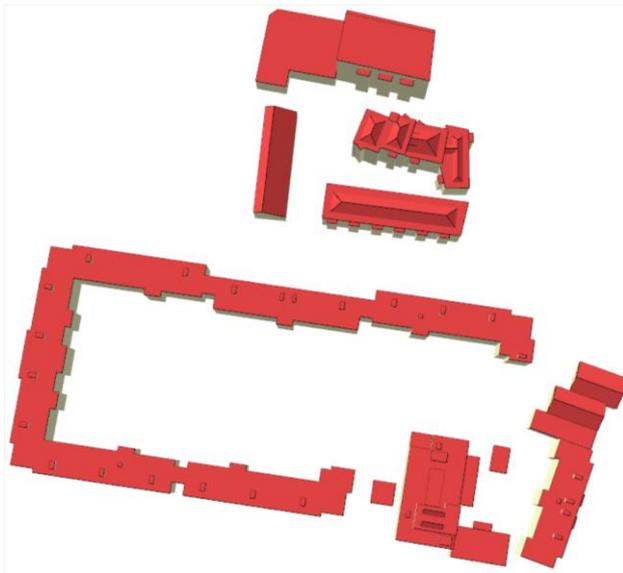
(b) The reconstructed 3D models with planes

Figure 4. The plane segments and 3D models from S3.

In addition, further quantitative evaluation is performed by the average distance between a point to the reconstructed 3D plane model, which is an internal quality measure. The assessment results for the tested point cloud are 0.68cm (**S1**), 1.3cm (**S2**), and 2.6 cm (**S3**), respectively. Moreover, over-segmentation is encouraged in the first data pre-processing, which can achieve more valid planar planes and can be further optimized in the proposed graph-based energy model.

5. CONCLUSION

A robust and accurate segmentation scheme for extracting a set of planar elements has been proposed. The main contribution of this paper is to transform the plane extraction problem into the best matching issue between the over-segmented super-voxels and the 3D plane models using an energy minimization framework. To get robust and reliable plane models, we first divide the input point cloud into over-segmented super-voxels, and cluster planar one to generate planes, then a random sampling strategy and a scene prior-based translating are adopted to enrich these plane models. The final super-voxels to planes assignment (planes segmentation) problem has achieved by these candidate planes, original super-voxels, and its adjacent graph. The qualitative and quantitative results of three types of point clouds with different point density and feature characteristics have proven the effectiveness of the proposed approach. It will be interesting to extract and optimize the freeform surface in the near future.

ACKNOWLEDGEMENTS

REFERENCES

Brook, A., Ben-Dor, E., Richter, R., 2013. Modelling and monitoring urban built environment via multi-source integrated and fused remote sensing data. International Journal of Image and Data Fusion 4, 2-32.

Chen, D., Zhang, L., Mathiopoulos, P.T., Huang, X., 2014. A methodology for automated segmentation and reconstruction of urban 3-d buildings from als point clouds. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing 7, 4199-4217.

Delong, A., Osokin, A., Isack, H.N., Boykov, Y., 2012. Fast Approximate Energy Minimization with Label Costs. International Journal of Computer Vision 98, 1-27.

Deschaud, J.-E.F., amp, ois, 2010. A Fast and Accurate Plane Detection Algorithm for Large Noisy Point Clouds Using Filtered Normals and Voxel Growing.

Dong, Z., Yang, B., Hu, P., Scherer, S., 2018. An efficient global energy optimization approach for robust 3D plane segmentation of point clouds. ISPRS Journal of Photogrammetry and Remote Sensing 137, 112-133.

Duda, R.O., Hart, P.E., 1972. Use of the Hough transformation to detect lines and curves in pictures. Commun. ACM 15, 11-15. Fischler, M.A., Bolles, R.C., 1981. Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography. Commun. ACM 24, 381-395.

Hossam Isack, Y.B., 2012. Energy-Based Geometric Multi-model Fitting. International Journal of Computer Vision 97, 123-147.

Isack, H., Boykov, Y., 2012. Energy-Based Geometric Multi-model Fitting. International Journal of Computer Vision 97, 123-147.

Kim, C., Habib, A., Pyeon, M., Kwon, G.-r., Jung, J., Heo, J., 2016. Segmentation of Planar Surfaces from Laser Scanning Data Using the Magnitude of Normal Position Vector for Adaptive Neighborhoods. Sensors 16, 140.

Kim, K., Shan, J., 2011. Building roof modeling from airborne laser scanning data based on level set approach. ISPRS Journal of Photogrammetry and Remote Sensing 66, 484-497.

Kong, X., Zhai, G., Zhong, B., Liu, Y., 2019. PASS3D: Precise and Accelerated Semantic Segmentation for 3D Point Cloud. arXiv preprint arXiv:1909.01643.

Milioto, A., Vizzo, I., Behley, J., Stachniss, C., 2019. Rangenet++: Fast and accurate lidar semantic segmentation, Proc. of the IEEE/RSJ Intl. Conf. on Intelligent Robots and Systems (IROS).

Monszpart, A., Mellado, N., Brostow, G.J., Mitra, N.J., 2015. RAPter: rebuilding man-made scenes with regular arrangements of planes. ACM Trans. Graph. 34, 1-12.

Papon, J., Abramov, A., Schoeler, M., Worgotter, F., 2013. Voxel cloud connectivity segmentation-supervoxels for point clouds, Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 2027-2034.

Pham, T.T., Chin, T.-J., Schindler, K., Suter, D., 2014. Interacting geometric priors for robust multimodel fitting. IEEE transactions on image processing 23, 4601-4610.

Pham, T.T., Eich, M., Reid, I., Wyeth, G., 2016. Geometrically consistent plane extraction for dense indoor 3D maps segmentation, 2016 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), pp. 4199-4204.

Pham, T.T., Reid, I., Latif, Y., Gould, S., 2015. Hierarchical higher-order regression forest fields: An application to 3d indoor scene labelling, Proceedings of the IEEE international conference on computer vision, pp. 2246-2254.

Rabbani, T., Van Den Heuvel, F., Vosselmann, G., 2006. Segmentation of point clouds using smoothness constraint. International archives of photogrammetry, remote sensing and spatial information sciences 36, 248-253.

Sampath, A., Shan, J., 2010. Segmentation and Reconstruction of Polyhedral Building Roofs From Aerial Lidar Point Clouds. IEEE Transactions on Geoscience and Remote Sensing 48, 1554-1567.

Schnabel, R., Wahl, R., Klein, R., 2007. Efficient RANSAC for point‐cloud shape detection, Computer graphics forum. Wiley Online Library, pp. 214-226.

Vo, A.-V., Truong-Hong, L., Laefer, D.F., Bertolotto, M., 2015. Octree-based region growing for point cloud segmentation. ISPRS Journal of Photogrammetry and Remote Sensing 104, 88-100.

Vosselman, G., Coenen, M., Rottensteiner, F., 2017. Contextual segment-based classification of airborne laser scanner data. ISPRS Journal of Photogrammetry and Remote Sensing 128, 354-371.

Yan, J., Shan, J., Jiang, W., 2014. A global optimization approach to roof segmentation from airborne lidar point clouds. ISPRS Journal of Photogrammetry and Remote Sensing 94, 183-193.

Yang, B., Dong, Z., 2013. A shape-based segmentation method for mobile laser scanning point clouds. ISPRS Journal of Photogrammetry and Remote Sensing 81, 19-30.

Zhou, G., Cao, S., Zhou, J., 2016. Planar Segmentation Using Range Images From Terrestrial Laser Scanning. IEEE Geoscience and Remote Sensing Letters 13, 257-261.