

## 3D-CNN BASED TREE SPECIES CLASSIFICATION USING MOBILE LIDAR DATA

Haiyan Guan <sup>1,\*</sup>, Yongtao Yu <sup>2</sup>, Wanqian Yan <sup>3</sup>, Dilong Li <sup>4</sup> and Jonathan Li <sup>5</sup>

<sup>1</sup> School of Remote Sensing and Geomatics Engineering, Nanjing University of Information Science & Technology, Nanjing, 210044, China — \*guanhy.nj@nuit.edu.cn

<sup>2</sup> Faculty of Computer and Software Engineering, Huaiyin Institute of Technology, Huaian, 223003, China — allennessy.yu@gmail.com

<sup>3</sup> School of Geographical Science, Nanjing University of Information Science & Technology, Nanjing, 210044, China — ywq\_nuist@163.com

<sup>4</sup> State Key laboratory of Information Engineering in Surveying, Mapping, and Remote Sensing, Wuhan University, Wuhan, China — scholar.dll@gmail.com

<sup>5</sup> Department of Geography and Environmental Management, University of Waterloo, Waterloo, ON N2L 3G1, Canada - junli@uwaterloo.ca

**KEY WORDS:** Mobile LiDAR Data, Tree Segmentation, Voxelization, NCut, Supervoxels, 3D-CNN

### ABSTRACT:

Our work addresses the problem of classifying tree species from mobile LiDAR data. The work is a two step-wise strategy, including tree segmentation and tree species classification. In the tree segmentation step, a voxel-based upward growing filtering is proposed to remove terrain points from the mobile laser scanning data. Then, individual trees are segmented via a Euclidean distance clustering approach and Voxel-based Normalized Cut (VNCut) segmentation approach. In the tree species classification, a voxel-based 3D convolutional neural network (3D-CNN) model is developed based on intensity information. A road section data acquired by a RIEGL VMX-450 system are selected for evaluating the proposed tree classification method. Qualitative analysis shows that our algorithm achieves a good performance.

### 1. INTRODUCTION

In the case of urban areas, tree species classification is gaining increasing attention for safety studies, noise modelling, and environmental and ecological analysis because trees play a critical role in urban ecosystems for the maintenance of environmental quality, aesthetic beauty of urban landscape, and social service for inhabitants. As cities grow rapidly, urban forests are increasingly displaced by infrastructure. Therefore, municipal governments desire to control development near greenbelt areas by using land cover maps. Moreover, biodiversity parameters such as tree species and age and height distributions are primarily chosen for ecosystem analysis.

In the last two decades, light detection and ranging (LiDAR) technology has been developed dramatically. Various systems have been designed for different applications and purposes, such as mobile, terrestrial, and aerial LiDAR systems, as well as Unmanned Aerial Vehicles (UAV) LiDAR systems. Mobile LiDAR, a widely used technology since year 2003 when the first mobile LiDAR system was developed, has attracted much attention for mainly transportation-related surveys (Jacobs, 2005; Toth, 2009). It is a data revolution. With a mobile LiDAR system, mobile mapping engineers can drive on a highway, rural road, and railroad, or along the shoreline of a river or lake. Along the way, the system captures trees, bridges, streetlights, buildings, power lines, other street-scene small objects (e.g. cracks, road markings). The collected data are a totally immersive three-dimensional (3D) view of the objects and surroundings (Rybka, 2011).

Due to the superior capabilities of LiDAR systems in collecting three-dimensional (3D) georeferenced scene data, algorithms and techniques for tree detection and classification using point cloud data have been intensively studied in the literature. Li et

al. (2016) developed a dual growing method for extracting individual trees from mobile LiDAR data. Both the trunks and crowns were effectively delineated through the dual growing process followed by a refining process. 3D segmentation methods, such as 3D Hough transform (Rutzinger et al. 2010), minimum spanning tree (Shen et al. 2008), and stepwise voxel-based marked neighborhood searching (Wu et al. 2013), were developed for identifying street trees from scattered mobile LiDAR points. In addition, algorithms combining mobile LiDAR data with digital images or videos captured by onboard digital camera(s) or video camera(s) were also exploited for tree detection (Zhong et al. 2013).

As for tree species classification in mobile LiDAR data, some deep learning methods were used recently. Guan et al. (2014) proposed a Deep Boltzmann Machines (DBMs) based tree classification method, which classify ten tree species from the tree waveform representation, reflecting tree geometric structures in mobile LiDAR data. Zou et al. (2017) proposed a deep belief network (DBN) model based tree classification method, in which the generated high-level features were used in a softmax classifier in the tree species classification step. However, these methods were applied deep learning algorithms to generate low-level or high-level features, which were input into a machine learning classifier later. This operation just utilizes these deep learning methods as a feature generator, leading to not fully reflecting their advantages in object recognition and classification.

In this paper, we develop an effective processing workflow for individual tree detection and species classification using mobile LiDAR data. The remainder of this paper is organized as follows. Section 2 describes the proposed tree species classification method. Section 3 demonstrates the experimental results and discussion. Finally, concluding remarks are given.

## 2. METHODOLOGY

The proposed method is a three step-wise strategy: (1) tree segmentation, which includes the separation of terrain points and non-terrain points via a voxel-based upward growing filtering, individual tree segmentation based on a Euclidean distance clustering approach and a voxel-based Normalized Cut approach, and (2) tree species classification based on 3D-CNN.

### 2.1 Tree Segmentation

In order to reduce the amount of data to be processed, first, we filter out the ground points from the entire scene by using the voxel-based upward growing method (Yu et al., 2015). For this method, first, the entire scene is vertically partitioned into a group of data blocks in the XY plane with a block size of  $w_b$ . These data blocks are processed separately for ground point filtering. Then, each data block is voxelized into an octree structure with a spacing of  $w_v$ , in which a voxel is connected with 26 neighbors. Finally, the upward growing scheme, which only grows upward to the nine neighbors above a voxel, is applied to each of these voxels to label it into the ground or the non-ground according to a pre-defined ground threshold,  $H_{\text{thresh}}$ . The voxels labeled as non-ground are retained, whereas the others are filtered out to narrow the searching regions.

To segment individual trees from the non-ground points, first, we adopt a fast Euclidean distance clustering approach, which groups unordered points based on their Euclidean distances to their neighbors. Specifically, for an unlabeled point, it will be grouped into a specific cluster if and only if its shortest Euclidean distance to the points within this cluster lies below a clustering threshold,  $d_t$ . After Euclidean distance clustering, the isolated individual trees located with a certain spatial distance from other trees are well segmented. However, the overlapping trees are falsely segmented into the same cluster. Such overlapping trees should be further separated. In this paper, we propose a voxel-based normalized cut segmentation method to perform individual tree segmentation from such clusters.

First, a cluster is divided into a voxel structure using the octree partition strategy with a spacing of  $w_s$ . Then, all the voxels are connected and constructed into a weighted graph  $G=\{V, E\}$ , where the vertices  $V$  are composed of the voxels, and the edges  $E$  are linked between each pair of voxels. The weights on the edges evaluate the similarity between the two connected voxels. To provide promising segmentation performance and take full advantage of the properties of mobile LiDAR data, the weight  $w_{ij}$  on the edge connecting voxels  $i$  and  $j$  is formulated as follows:

$$w_{ij} = \begin{cases} e^{-\frac{D_{ij}^{XY}}{\sigma^{XY}}} \times e^{-\frac{D_{ij}^Z}{\sigma^Z}} \times e^{-\frac{D_{ij}^S}{\sigma^S}} & \text{if } D_{ij}^{XY} < \Gamma_R \\ 0 & \text{otherwise} \end{cases} \quad (1)$$

Where  $D_{ij}^{XY}$ ,  $D_{ij}^Z$  and  $D_{ij}^S$  represent the horizontal, vertical, and shortest distance, respectively, between nodes  $i$  and  $j$ .  $\sigma^{XY}$ ,  $\sigma^Z$  and  $\sigma^S$  are coefficients, set to be 0.05 times of the maximum of  $D_{ij}^{XY}$ ,  $D_{ij}^Z$  and  $D_{ij}^S$ , for controlling the sensitivity of the impact factors, respectively.  $\Gamma_R$  represents the maximum horizontal distance threshold between nodes.  $w_{ij} = 0$ , if the horizontal distance between nodes  $\{i,j\}$  exceeds the threshold,  $\Gamma_R$ . Ncut aims to divide the graph  $G$  into two disjoint voxel groups  $A$  and  $B$  by maximizing the similarity within each voxel group and

minimizing the similarity between two voxel groups  $A$  and  $B$  (Shi and Malik 2000).

### 2.2 Tree Species Classification

To effectively conduct tree species classification on the segmented individual trees, we develop a 3D convolutional neural network (3D-CNN) model. To this end, first, for each segmented tree, we uniformly segment it into a voxel representation with a set of  $W \times W \times H$  voxels along the horizontal and vertical directions. Then, for each voxel, the intensities of the points in this voxel are interpolated to from the voxel's spectral information. In this paper, the inverse distance weighted interpolation method (Yu et al., 2014) is adopted to interpolate intensities. Finally, such voxel representations of trees with interpolated voxel spectral information are fed into the 3D-CNN for tree species classification.

As shown in Figure 1, the 3D-CNN model consists of four 3D convolutional layers (Conv1 to Conv4), a max-pooling layer, and three fully connected layers (FC1 to FC3). The four 3D convolutional layers function to extract different levels of local or partial features of trees through 3D convolution operations. The Conv1 layers is performed on the voxel representations of trees and only considers the voxel spectral properties as features for convolution operations. To perform 3D convolution operations, the convolution kernel is designed to be a 3D cubic form (e.g.,  $K \times K \times K \times D$ , where  $K$  is the kernel length and  $D$  is the number of feature maps) sliding across the 3D feature maps. The 3D convolution operations provide a horizontal-rotation-invariant representation. That is, the horizontal rotations of trees have no effect on the resultant feature representations. Similar to conventional 2D CNNs, the max-pooling layer is used for feature selection and reduction. Through max-pooling operations, the most salient and representative features are selected and the other features are suppressed. In addition, the number of network parameters is dramatically reduced after max-pooling, which helps to build a lightweight network.

By taking into account and fusing the local features extracted by the convolutional layers, the FC1 and FC2 layers are designed to extract global features of trees. The features extracted by the FC2 layer provide a tree-level feature representation. Such features are used for distinguishing different tree species. The FC3 layer is a softmax layer for tree species classification, where each neuron represents a specific tree species. For an input tree, the species label of this tree is determined according to the neuron with the maximum output of predication in the FC3 layer. It can be deduced as follows:

$$T^* = \arg \max_k h_k \quad (2)$$

where  $T^*$  is the predicted species and  $h_k$  is the output of the  $k$ -th neuron in the FC3 layer. Specifically, for all the convolutional layers and fully connected layers, the widely used ReLU is used as the activation function to make nonlinear transformations on the neuron outputs. In addition, the cross entropy loss is used as the loss function for conducting the backpropagation process when training the 3D-CNN. After applying the 3D-CNN model to the voxelized trees, each tree is assigned to a specific tree species.

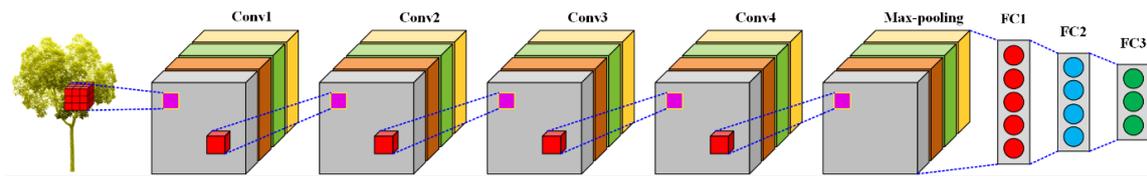


Figure 1. Architecture of the 3D-CNN model.

### 3. RESULTS AND DISCUSSION

In this study, the data were acquired by a RIEGL VMX-450 mobile LiDAR system, which can achieve a maximum effective measurement rate of 1.1 million points per second and a scan speed of 400 lines per second, according to its specification. This survey, at a driving speed of approximately 30 - 50 km/h, was carried out along Huandao Road from Xiamen University to International Conference and Exhibition Center (ICEC) in Xiamen, China. The surveyed area is in a subtropical urban environment with high buildings, dense vegetation, and traffic signposts on both sides of the road. The point density for mobile LiDAR data strongly relies on the nominal distance to the target where the point spacing is measured as well as the incidence angle. Accordingly, at the vehicle speed of 30 km/hour, the values of point density are estimated as 286.44 points/ m<sup>2</sup>.

For urban landscaping, a total number of ten different species of trees are planted along both sides of the road corridor. At training stage, 50,000 tree samples from ten different tree species, each of which includes 5,000 tree samples, were selected for training the 3D-CNN. At test stage, from the collected data, a road segment of approximately four kilometers, containing 2013 trees from ten species, was selected for evaluating the proposed 3D-CNN-based tree species classification algorithm.

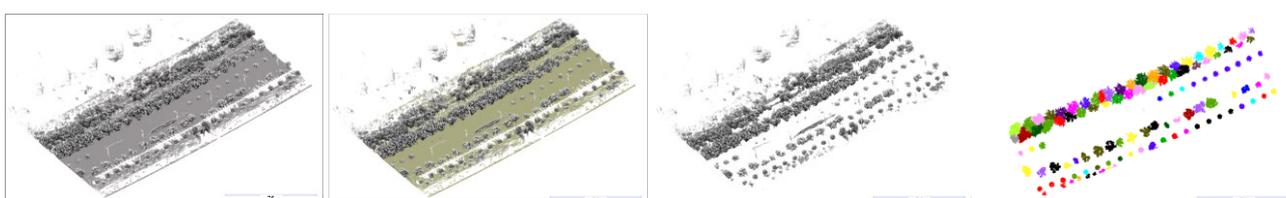
#### 3.1 Single-tree Segmentation

In the single-tree segmentation, five parameters were used: block size ( $w_b$ ), voxel size ( $w_v$ ), terrain threshold ( $H_{\text{thresh}}$ ), clustering threshold ( $d_t$ ), and NCut voxel size ( $w_s$ ). The block size ( $w_b$ ) and voxel size ( $w_v$ ) control the efficiency of data processing, and thus were set to 3 m and 5 cm, respectively, according to the data to be used in this study. The survey terrain is relatively flat, and thus was set to 0.3 m for removing terrain points. After the removal of terrain points, the tree segmentation was performed on these non-terrain points via the proposed Euclidean distance clustering and voxel-based normalized cut. The clustering threshold ( $d_t$ ) was set to 0.15 m and the voxel size ( $w_s$ ) was set to 5 cm. The thresholds of these parameters were empirically determined based on the point density of the collected mobile LiDAR data. However, some clusters included several trees and other objects such as light poles because the trees, particularly planted along the road sides, were very dense

and even severely overlapped together. Thus, the proposed voxel-based normalized cut was performed to obtain the individual trees. As shown in Figure 2(b), road points (rendered by olive colour) are separated well from non-terrain points. Figure 2(c) shows the classified non-terrain points only. The segmented trees are shown in Figure 2(d). The visual inspection shows that the majority of trees individually segmented and the tree segmentation results are satisfactory.

#### 3.2 Tree Classification

To evaluate the performance of the proposed method, ten different tree species were selected from the surveyed area. Figure 3 shows some examples of the ten tree species, which were termed as T1, T2 ... T10. In the selected dataset, there are a total number of 2013 trees from ten different tree species. The tree species classification results are illustrated in Table 1 by using a confusion matrix. In Table 1, each row corresponds to the ground truth of each tree species and the classification results with regard to different tree species. Each column denotes the trees that were classified into a specific tree species. The main diagonal reflects the trees that were correctly classified as their associated tree species. To quantitatively evaluate the tree species classification results, we adopted the following two metrics: overall accuracy and kappa coefficient. As reflected in Table 1, the overall tree classification accuracy is 96.4% and the kappa coefficient is 0.96. To quantitatively analyse the classification results of each tree species, we also computed and recorded the producer's and user's accuracies. The user's accuracy reflects the proportion of correctly classified trees with regard to the ground truth. The producer's accuracy reflects the proportion of correctly classified trees with regard to trees classified as a specific species. Thus, the higher these two accuracies, the better the classification performance. Specifically, the producer's accuracies for all trees are greater than 93.4%, and user's accuracies for all trees are greater than 94.7%. The misclassification errors were mainly caused by the geometry similarities of some trees among different species; thus, these trees were falsely classified into other species. For example, tree species T6 and T8 are quite similar in geometry structures. Therefore, some trees of species T6 were falsely classified as species T8. In addition, some trees of species T8 were falsely classified as species T6. However, for all tree species, the misclassification rate is quite low and acceptable.



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

Figure 2. A test data selected from Figure 1. (a) The raw point cloud, (b) the classified terrain points and non-terrain point, (c) the classified non-terrain points, and (d) the individual tree segmentation results.

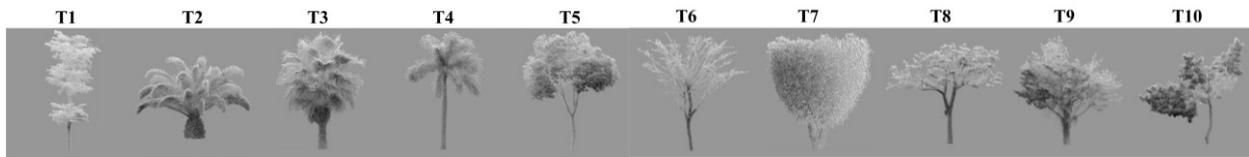


Figure 3. Illustration of 10 different tree species (termed as T1, T2, ..... , T10) in the study area.

Species	Validation trees										Total	Accuracy UA(%)
	T1	T2	T3	T4	T5	T6	T7	T8	T9	T10		
T1	201	0	0	1	1	4	0	2	0	1	210	95.7
T2	0	183	3	1	0	0	3	0	0	0	190	96.3
T3	1	1	182	2	0	0	2	0	1	0	189	96.3
T4	2	0	1	203	2	1	0	1	0	0	210	96.7
T5	0	0	1	0	198	0	0	2	2	0	203	97.5
T6	0	0	0	1	2	200	0	5	2	1	211	94.8
T7	0	0	3	0	1	0	189	0	0	0	193	97.9
T8	0	0	0	0	3	5	0	197	2	1	208	94.7
T9	0	0	1	0	2	0	0	1	182	0	186	97.8
T10	0	0	0	0	3	0	0	3	2	205	213	96.2
Total	204	184	191	208	212	210	194	211	191	208	2013	
Accuracy	PA(%)	98.5	99.5	95.3	97.6	93.4	95.2	97.4	93.4	95.3	98.6	

OA(%) = 96.4; Kappa = 0.96

Table 1. Overall results of classification accuracies of ten tree species by using confusion matrix, showing the numbers of classified vs. validation trees. PA – producer’s accuracy; UA – user’s accuracy; OA – overall accuracy; Kappa – Kappa coefficient

### 3.3 Comparative Experiments

In this study, the 3D-CNN model was used to perform tree species classification on the segmented individual trees. A promising classification performance was achieved on the test LiDAR data. Thus, to further demonstrate the superior performance of the 3D-CNN model, we compared it with the following two methods: DBM-based method (Guan et al. 2014) and DBN-based method (Zou et al. 2017). For the DBM-based method, first, a tree point cloud is converted into a waveform representation, which well reflects the geometrical properties of a tree. Then, DBM is performed on the waveform representation to conduct tree species classification. For the DBN-based method, first, a tree point cloud is projected and rasterized into a set of 2D images from different views. Then, DBN is performed on these 2D images to carry out tree species classification. Table 2 exhibits the results of the comparison. Likewise, to quantitatively compare their performances, we computed and recorded the overall accuracy and kappa coefficient of each method, as well as the user’s and producer’s accuracies. In table 2, only the minimum and maximum user’s and producer’s accuracies and presented. Comparatively, The DBN-based method performed better than the DBM-based method with an overall accuracy and kappa coefficient of 95.6 and 0.9, respectively. However, by using the 3D-CNN model, our proposed method achieved the overall accuracy and kappa coefficient of 0.8% and 0.06 higher than the DBN-based method, and 10.3% and 0.16 higher than the DBM-based method, respectively. On the whole, the proposed 3D-CNN model worked efficiently and obtained promising performance in tree species classification from mobile LiDAR data.

Method	OA (%)	Kappa coefficient	UA (%)		PA (%)	
			MIN	MAX	MIN	MAX
DBM-based	86.1	0.8	78.2	96.2	81.3	90.7
DBN-based	95.6	0.9	93.3	97.1	94.5	96.2
Proposed	96.4	0.96	94.7	97.9	93.4	99.5

Table 2. Comparative results of classification accuracies of three methods.

## 4. CONCLUSION

This paper has presented a two step-wise tree classification method from mobile LiDAR data. To efficiently identify trees, in the tree segmentation, the voxel-based upward growing filtering method was first presented to classify mobile LiDAR data into terrain points and non-terrain points. Next, based on the classified non-terrain points, the Euclidean distance clustering was performed to obtain a set of clusters and then the voxel-based normalized cut method was used to obtain individual trees from these clusters. In the tree classification, voxel-based 3D-CNN was performed on the segmented individual trees to classify them into different tree species. To evaluate the performance of the proposed tree classification approach, a road section was selected from the mobile LiDAR data acquired by a RIEGL VMX-450 system. The road section

contains 2013 trees from ten species. Through quantitative evaluations, the proposed method achieved an overall tree classification accuracy and a kappa coefficient of 96.4% and 0.96, respectively. Comparative tests also demonstrated the feasibility and high performance of the proposed method in tree species classification using mobile LiDAR data.

#### ACKNOWLEDGEMENTS

This work was supported in part by the National Natural Science Foundation of China under Grant 41671454 and Grant 61603146, in part by the Natural Science Foundation of Jiangsu Province under Grant BK20160427, and in part by the Natural Science Research in Colleges and Universities of Jiangsu Province under Grant 16KJB520006. Contact corresponding author via [guanhy.nj@nuist.edu.cn](mailto:guanhy.nj@nuist.edu.cn) for more information about the software tools used in this study.

#### REFERENCES

- Aval, J., Demuynck, J., Zenou, E., et al., 2018. Detection of Individual Trees in Urban Alignment from Airborne Data and Contextual Information: A Marked Point Process Approach. *ISPRS Journal of Photogrammetry and Remote Sensing*, 146, pp.197-210.
- Guan, H., Yu, Y., Ji, Z., Li, J., Zhang, Q., 2015. Deep Learning Based Tree Classification Using Mobile Lidar Data. *Remote Sensing Letters*, 6(11), pp.864-873.
- Huang, H., Li, X., Chen, C., 2018. Individual Tree Crown Detection and Delineation from Very-High-Resolution UAV Images Based on Bias Field and Marker-Controlled Watershed Segmentation Algorithms. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 11(7), pp.2253-2262.
- Jacobs, G., 2005. Uses in Transportation in High-Definition Surveying: 3D Laser Scanning, *Professional Surveyor Magazine*, vol. April.
- Li, L., Li, D., Zhu, H., Li, Y., 2016. A Dual Growing Method for The Automatic Extraction of Individual Trees from Mobile Laser Scanning Data. *ISPRS Journal of Photogrammetry and Remote Sensing*, 120, pp.37-52.
- Rutzinger, M., Pratihast, A., Oude Elberink, S., and Vosselman, G., 2010. Detection and Modelling of 3D Trees from Mobile Laser Scanning Data. In: *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, Newcastle upon Tyne, UK, Vol.38, Part 5, pp. 520–525.
- Rybka, R., 2011. Autodesk and Bentley Systems Talk about Mobile LiDAR. *LiDAR*, 1(2), pp.41-44.
- Shen, Y., Sheng, Y., Zhang, K., Tang, Z., and Yan, S., 2008. Feature Extraction from vehicle-Borne Laser Scanning Data. In: *The International Conference on Earth Observation Data Processing and Analysis (ICEODPA)*, Wuhan, China, Vol.7285.
- Shi, J., Malik, J. 2000. Normalized Cuts and Image Segmentation, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22(8), pp.888–905.
- Toth, C., 2009. R&D of Mobile Lidar Mapping and Future Trends. ASPRS Annual Conference, Baltimore, Maryland, USA.
- Wagner, F.H., Ferreira, M.P., Sanchez, A., et al., 2018. Individual Tree Crown Delineation in Highly Diverse Tropical Forest Using Very High Resolution Satellite Images. *ISPRS Journal of Photogrammetry and Remote Sensing*, 145, 362-377.
- Wu, B., Yu, B., Yue, W., Shu, S., Tan, W., Hu, C., Huang, Y., Wu, J., and Liu, H., 2013. A Voxel-Based Method for Automated Identification and Morphological Parameters Estimation of Individual Street Trees from Mobile Laser Scanning Data. *Remote Sensing*, 5, pp.584–611.
- Yu, Y., Li, J., Guan, H., Wang, C., and Yu, J., 2014. Automated Detection of Road Manhole and Sewer Well Covers from Mobile Lidar Point Clouds. *IEEE Geoscience and Remote Sensing Letters*, 11(9), pp.1549-1553.
- Yu, Y., Li, J., Guan, H., Wang, C., and Yu, J., 2015. Semiautomated Extraction of Street Light Poles from Mobile LiDAR Point-Clouds. *IEEE Transactions on Geosciences & Remote Sensing*, 53(3), pp.1374–1386.
- Yu, Y., Li, J., Guan, H., Wang, C., 2015. Automated Extraction of Urban Road Facilities Using Mobile Laser Scanning Data. *IEEE Transactions on Intelligent Transportation Systems*, 16(4), pp.2167-2181.
- Zhong, R., Wei, J., Su, W., and Chen, Y., 2013. A Method for Extracting Trees from Vehicle-Borne Laser Scanning Data, *Mathematical and Computer Modelling*, 58, pp.733-742.
- Zou, X., Cheng, M., Wang, C., Xia, Y., and Li, J., 2017. Tree Classification in Complex Forest Point Clouds Based on Deep Learning, *IEEE Geoscience and Remote Sensing Letters*, 14, pp.2360–2364.