

Article

Outdoor Scene Understanding Based on Multi-Scale PBA Image Features and Point Cloud Features

Yisha Liu ¹, Yufeng Gu ², Fei Yan ^{2,*} and Yan Zhuang ²

¹ Information Science and Technology College, Dalian Maritime University, Dalian 116026, China; liuyisha@dmlu.edu.cn

² School of Control Science and Engineering, Dalian University of Technology, Dalian 116024, China; guyufeng@mail.dlut.edu.cn (Y.G.); zhuang@dlut.edu.cn (Y.Z.)

* Correspondence: fyan@dlut.edu.cn; Tel.: +86-411-84707581

Received: 8 September 2019; Accepted: 17 October 2019; Published: 19 October 2019



Abstract: Outdoor scene understanding based on the results of point cloud classification plays an important role in mobile robots and autonomous vehicles equipped with a light detection and ranging (LiDAR) system. In this paper, a novel model named Panoramic Bearing Angle (PBA) images is proposed which is generated from 3D point clouds. In a PBA model, laser point clouds are projected onto the spherical surface to establish the correspondence relationship between the laser ranging point and the image pixels, and then we use the relative location relationship of the laser point in the 3D space to calculate the gray value of the corresponding pixel. To extract robust features from 3D laser point clouds, both image pyramid model and point cloud pyramid model are utilized to extract multiple-scale features from PBA images and original point clouds, respectively. A Random Forest classifier is used to accomplish feature screening on extracted high-dimensional features to obtain the initial classification results. Moreover, reclassification is carried out to correct the misclassification points by remapping the classification results into the PBA images and using superpixel segmentation, which makes full use of the contextual information between laser points. Within each superpixel block, the reclassification is carried out again based on the results of the initial classification results, so as to correct some misclassification points and improve the classification accuracy. Two datasets published by ETH Zurich and MINES ParisTech are used to test the classification performance, and the results show the precision and recall rate of the proposed algorithms.

Keywords: 3D point cloud; outdoor scene understanding; mobile laser scanning

1. Introduction

Outdoor scene understanding based on mobile laser scanning (MLS) point cloud data are a fundamental ability for unmanned vehicles and autonomous mobile robots navigating in urban environments. Recently, a variety of laser point cloud processing methods have been presented to recognize the main elements of road environment [1], to accomplish robust place recognition [2], to extract parameters of trees [3], and so on. Moreover, the point clouds obtained from a laser scanner can also be utilized to accomplish real-time shape acquisition [4], outdoor 3D laser data classification [5], and outdoor scene understanding [6]. A state-of-the-art review for object recognition, segmentation, and classification of MLS Point Clouds was also given in [7].

In order to reduce the computational complexity of feature extraction and classification, some scholars have converted 3D laser point clouds into 2D images and used image processing methods to process 3D point clouds, such as range image [8], reflectance image [9], and bearing angle image (BA image) [10]. The BA image was originally used to solve the calibration problem between camera and laser scanner [10]. Since the BA image has clearer texture details than range image and

reflectance image, it is also used to solve the laser point cloud classification problem in outdoor or indoor scenes. Zhuang et al. used a 2D BA image to represent the outdoor 3D point cloud [11]. By extracting texture features from the BA image, the scene understanding of the 3D point cloud was realized. Zhang et al. studied the problem of 3D object detection in a cluttered indoor environment and transformed the 3D laser point cloud into a 2D BA image, which enabled the robot to complete the task of scene understanding at a lower computational cost [12]. However, the quality of the BA image depends on the selection of the viewpoint position. If the viewpoint is not selected properly, the image will be indistinct. In addition, the BA image also has the problem of grayscale change for the same object.

It is very important to determine the neighborhood range of laser points for feature extraction [13]. There are two kinds of neighborhood selection methods: fixed-scale neighborhood selection and multi-scale neighborhood selection. Fixed-scale neighborhood selection depends heavily on experience. If the scale is too small, the whole view cannot be seen clearly. If the scale is too large, the details can easily be ignored [14]. Therefore, the multi-scale feature extraction of 3D point clouds is a good choice, which is very helpful for improving classification performance.

Many classifiers can be used for point cloud classification, such as Support Vector Machine classifier, Nearest Neighbor classifier, K Nearest Neighbor classifier, and Naive Bayesian classifier [15]. When the selected feature dimension is very high (over 100 dimensions) and there are a lot of redundant features, the Random Forest classifier can process the high-dimensional data and complete the feature screening [16]. In addition, some classification methods based on contextual information of the point cloud are also widely used. Munoz et al. proposed a point cloud classification framework based on the Markov Random Field. Based on the contextual information, the classification of vehicular laser point cloud was realized [17]. Najafi et al. introduced a non-associative higher-order Markov Random Field to address the problem of semantic 3D point classification, which took into account the non-associative geometric context between different classes [18].

In this paper, a novel image model named PBA image is firstly proposed to represent the MLS point cloud data, which shows superior performance to display a large-scale scene with a panoramic view. Compared with the traditional BA model, the PBA model can still transform the unordered laser scanning data to a 2D image without fixed scan sequence relationships. To improve the accuracy and robustness of scene understanding results, multiple-scale features are extracted not only from the PBA images but also from the corresponding original LiDAR point clouds. In our work, the Random Forest algorithm is adopted to build the classifier which can complete feature screening and improve the generalization ability of classification. After the initial classification, superpixel segmentation is performed on the PBA images, which considers the contextual information between laser points in 2D images. Within each superpixel block, reclassification algorithm is performed based on the results of the initial classification, so as to correct partial misclassification points. A series of experimental results from both ETH Zurich and MINES ParisTech datasets are given to test the validity and robustness of the proposed approach.

2. Panoramic Bearing Angle Images Generating from 3D Laser Point Clouds

Generally speaking, there are two ways to obtain 3D laser point clouds to represent large-scale outdoor scenes in the field of mobile robotics. The first one is to install a 2D laser scanner on a mobile robot to perform on-the-fly scanning or fixed-point scanning. As shown in Figure 1, the left one is a driverless car using two lateral SICK LMS 511 laser ranger finders (produced by SICK AG, Waldkirch, Germany) to capture groups of sequenced 2D laser points in the on-the-fly scanning mode, while the right one is a mobile robot using a pitch rotating SICK LMS 511 to obtain 3D point clouds in fixed-point scanning mode. In these cases, the 3D point clouds are composed of sequential 2D laser scanning points and can be represented by a matrix, which can be transformed to grayscale images by using the Bearing Angle (BA) image model.

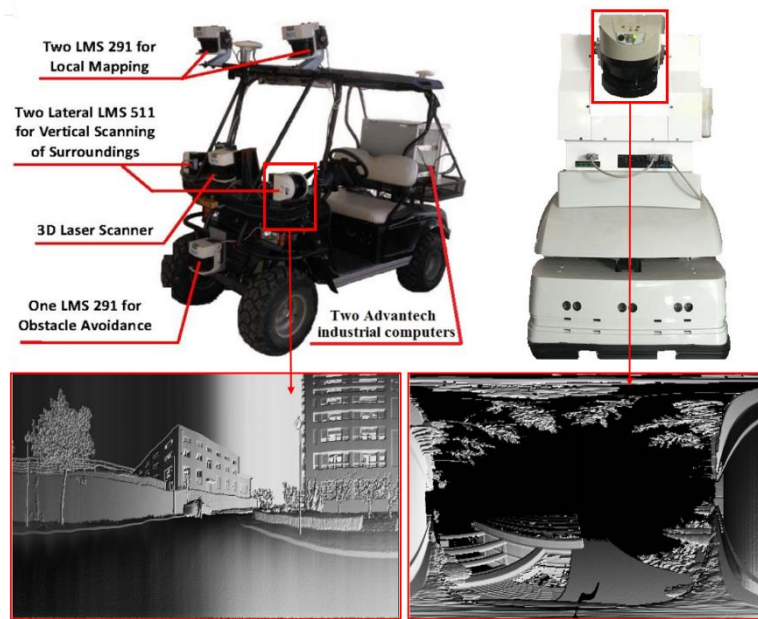


Figure 1. Grayscale images generated from laser scanning points obtained by the 2D laser scanners equipped on mobile robots.

The other method is to use the 3D laser scanner to obtain the 3D point clouds directly. However, these 3D point clouds are composed of several groups of scanning data and are always unordered when stored, so they cannot be represented by a matrix. In addition, in most public laser scanning datasets, there are no scan sequence relationships stored between different laser scans. To solve this problem, a novel Panoramic Bearing Angle (PBA) image model is proposed in this paper and introduced as follows.

2.1. Projection of 3D Laser Point Cloud to Pixel Plane

Viewpoint selection is a crucial step for 2D images generating from 3D laser point clouds. For fixed-point scanning, the location of the rotating 2D laser range finder is selected as the viewpoint. For on-the-fly scanning, the viewpoint is usually selected on the trajectory of the moving laser range finder. Suppose that a selected viewpoint of a 3D point cloud is $V(x_v, y_v, z_v)$, a laser point in the cloud is $P_i(x_i, y_i, z_i)$, and the matrix size for the 2D image to be generated is $M \times N$. As shown in Figure 2, a spherical coordinate system is established in which the viewpoint V is the center of the sphere. It should be noted that the size of the panoramic image is only related to the resolution of the image (the size of the image matrix) regardless of the size of the projection surface. According to (1), the original 3D laser point $P_i(x_i, y_i, z_i)$ is converted from the global coordinate system to the spherical coordinate system with the viewpoint V as the center of the sphere. The point in the spherical coordinate system is $P_i(r_i, \theta_i, \varphi_i)$.

$$\begin{cases} r_i = \sqrt{(x_i - x_v)^2 + (y_i - y_v)^2 + (z_i - z_v)^2} \\ \theta_i = \arccos\left(\frac{z_i - z_v}{r_i}\right) \\ \varphi_i = \arctan\left(\frac{y_i - y_v}{x_i - x_v}\right) \end{cases}, \quad (1)$$

where $\theta_i \in [0, \pi]$ and $\varphi_i \in [0, 2\pi]$.

According to (2), M warps l_m and $N + 1$ wefts l_p are drawn which can divide the sphere into $M \times N$ independent grids. The left image in Figure 3 is a spherical coordinate system which is divided into 64 grids by eight warps and nine wefts (two poles are included).

$$\begin{cases} l_m = m \times \frac{2\pi}{M}, m \in [0, M - 1] \\ l_p = p \times \frac{\pi}{N}, p \in [0, N] \end{cases} \quad (2)$$

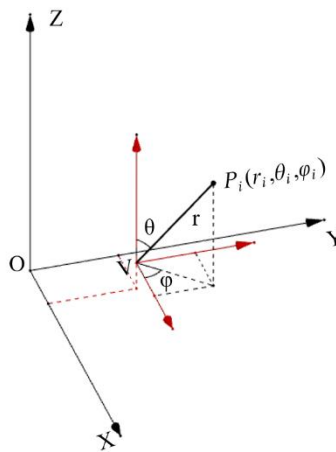


Figure 2. An illustration of coordinate transformation.

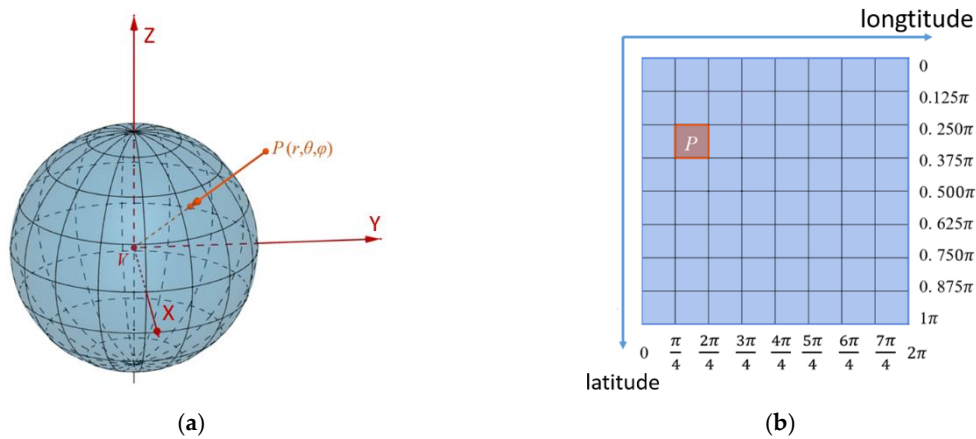


Figure 3. An illustration of laser ranging point projection. (a) A spherical coordinate system; (b) 2D matrix of the PBA.

Take the center of the sphere V as the starting point and make a ray through each laser scanning point $P_i(r_i, \theta_i, \varphi_i)$, so that the laser point can be projected to a grid of the sphere. If there are more than one projections of laser points in a grid, the one closest to the center of the sphere is retained. Then cut the spherical surface along the 0-degree warp and spread it to the horizontal plane to obtain the 2D matrix of the PBA (see the right image of Figure 3).

As shown in Figure 4a, a 3D laser point cloud is obtained in the fixed scanning point V , and Figure 4b is the corresponding panoramic image, which is displayed in the binary value. The white pixel indicates that there is a laser scanning point corresponding to it, while the black pixel indicates that no laser point corresponds to it.

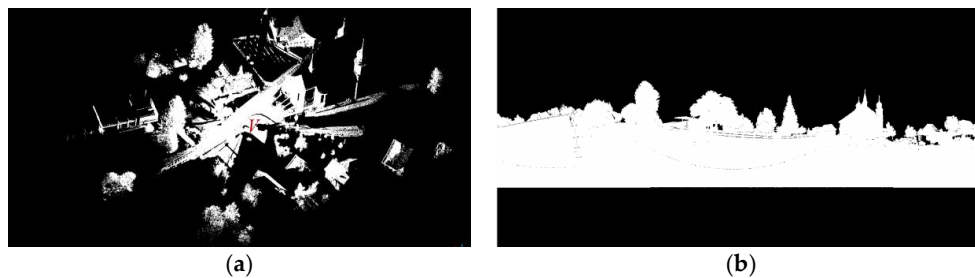


Figure 4. (a) A whole view of a large-scale outdoor scene represented by the original 3D point cloud; (b) The panoramic binarized image of the same scene's point cloud projected to the spherical coordinate system.

2.2. Calculating of Image Gray Value

There are many classical image models to represent laser points stored in the 2D matrix, such as reflectance image, range image and bearing angle (BA) image. However, the reflectance image is less robust and the edge description in range image is not clear enough, especially in large-scale scenes. The quality of the BA image depends on the selection of the viewpoint position. In addition, grayscale change may appear in the BA image. As shown in Figure 5, the gray values for the same railing are inconsistent, which is not beneficial to feature extraction and classification.

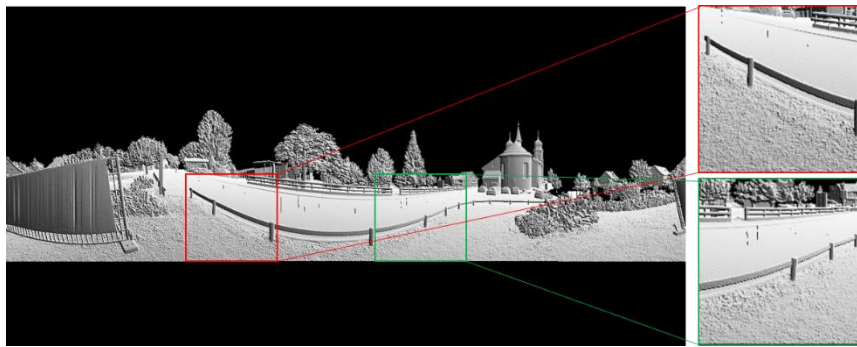


Figure 5. An example of a Bearing Angle (BA) image for an outdoor scene.

In order to overcome the above limitations, a novel PBA image model is proposed in this paper inspired by the BA model, which is not related to the selection of viewpoints. Moreover, the PBA image model can provide stable gray values for the same object and also ensure clear texture and high image contrast with high computational efficiency.

Here we will explain how to calculate the gray value of each pixel in the PBA image. As shown in Figure 6, there are M rows in the image matrix, and the image pixel corresponding to the laser scanning point P is defined as $P_{x,y}$, which is located in row x and column y . Two neighboring laser points P_l and P_r for point P are chosen as:

$$\begin{cases} \Psi(P_l) = P_{x-1,y-1}, \Psi(P_r) = P_{x-1,y+1}, X < M/2 \\ \Psi(P_l) = P_{x+1,y-1}, \Psi(P_r) = P_{x+1,y+1}, X \geq M/2 \end{cases} \quad (3)$$

where $\Psi(\cdot)$ represents the image pixel of a laser point. If the pixel is in the upper part of the image, its upper left and upper right pixels are selected as neighboring pixels; otherwise, the lower left and right lower pixels are selected as neighboring pixels.

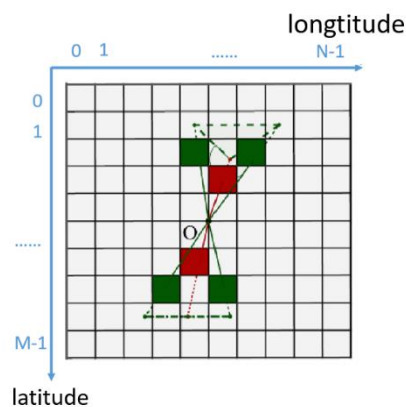


Figure 6. An illustration of the grayscale calculation.

The pixel gray value of $P_{x,y}$ is defined as:

$$G_{X,Y} = 255(1 - \alpha/2\pi) \quad (4)$$

where α is the angle between P and its neighboring laser scanning points P_l and P_r , which can be obtained as follows:

$$\begin{cases} \alpha = \angle P_l P P_r, & |VP| \geq (|VP_l| + |VP_r|)/2 \\ \alpha = 2\pi - \angle P_l P P_r, & |VP| < (|VP_l| + |VP_r|)/2 \end{cases} \quad (5)$$

where VP , VP_l , VP_r represent the distances between the center of the sphere V and laser points P , P_l , P_r , respectively.

An example of a PBA gray image is given in Figure 7. Compared with the BA image in Figure 5, the gray values for the same railing are consistent, and the boundaries of the objects in the scene are clearer.

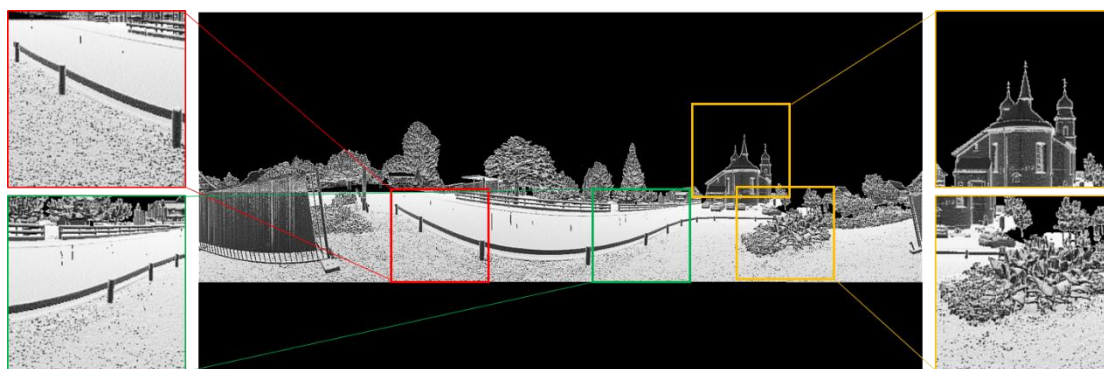


Figure 7. An example of a Panoramic Bearing Angle (PBA) image for the same outdoor scene.

3. Laser Point Clouds Classification Using Multi-Scale PBA Image Features and Point Cloud Features

It is important to select the neighborhood range of the laser points in the feature extraction step. In our work, the image pyramid model is adopted to extract the texture features of PBA images on multiple scales. The point cloud pyramid model is then used to extract the local features of the 3D point cloud on multiple scales.

3.1. Multi-Scale PBA Image Feature Extraction

In our work, feature extraction is accomplished in 2D gray images on multiple scales. When the scale is large, the computational cost is very high. Therefore, the PBA image is downsampled by using the image pyramid model [19]. The image pyramid model for PBA images is given in Figure 8. It should be noted that the image in each layer of the pyramid model is generated directly from the 3D laser point cloud, rather than from the downsampling of the original image.

Local Binary Pattern (LBP) is a kind of image texture feature, which is extracted from multi-resolution PBA images. For the classic LBP feature, eight fixed neighborhood pixels are selected (see Figure 9a). In order to extract multi-scale texture features an improved neighborhood selection method [20] is adopted for LBP feature extraction in our work, in which a circular neighborhood is selected with variable radius r . The pixel coordinate of the neighborhood points (x_p, y_p) can be obtained as follows:

$$\begin{cases} x_p = x_c + r \times \cos(\frac{2\pi P}{8}) \\ y_p = y_c - r \times \sin(\frac{2\pi P}{8}) \end{cases}, P = 0, 1, \dots, 7 \quad (6)$$

where (x_c, y_c) is the pixel coordinate of the center pixel. As shown in Figure 9a,b, r is selected as 1 and 2, respectively.

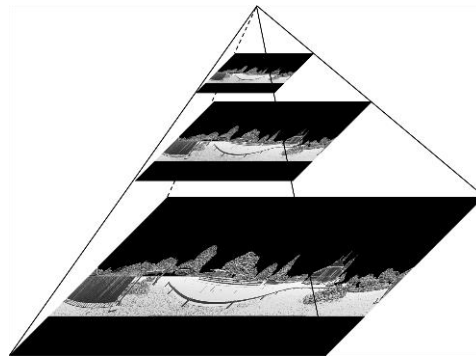


Figure 8. An example of image pyramid model for PBA image.

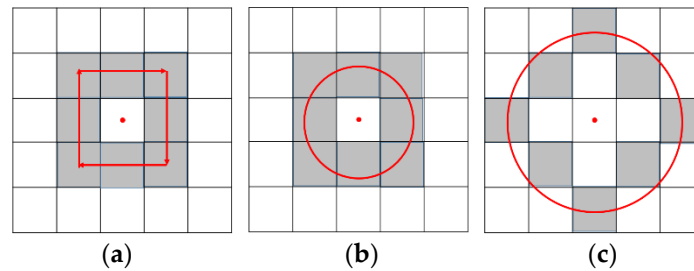


Figure 9. (a) Classic neighborhood selection method for classic LBP features; (b) Improved neighborhood selection method for LBP features ($r = 1$); (c) Improved neighborhood selection method for LBP features ($r = 2$).

Reduce the original 256-level gray PBA image to 3-level and the pixel gray value G_{new} of the simplified 3-level image can be obtained as follows:

$$G_{new} = \begin{cases} 0, & G_{old} = 0 \\ 127, & 1 \leq G_{old} \leq 127 \\ 255, & 128 \leq G_{old} \leq 255 \end{cases} \quad (7)$$

where G_{old} is the pixel gray value of the original PBA image.

Figure 10 shows an example of the simplified 3-level PBA image (black-0; gray-127; white-255), and four categories of typical local scenes also show distinct texture features, which are artificial ground (top left), natural ground (bottom left), buildings (top right), and vegetation (bottom right).

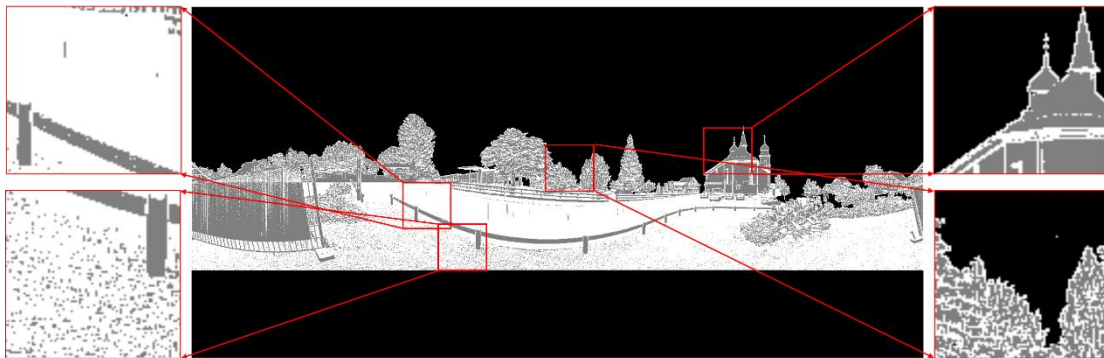


Figure 10. An example of a 3-level gray PBA image.

When feature extraction in different layers of the image pyramid model for the PBA images is completed, these features in different layers need to be fused. Starting from the top layer image of the

pyramid, the image features are upsampled, and then superimposed with the image features of the next layer. These two steps are repeated until the features in all layers are superimposed on the image at the bottom layer of the image pyramid model.

In summary, the $(P + 1)$ layer image pyramid model of PBA images is built from the original laser point cloud, and each layer of PBA images is converted to a 3-level gray image. LBP features are then extracted in each image pixel on m scales. Finally, the features in different layers are superimposed together from the top layer to the bottom layer. Therefore, there are $m \times (P + 1)$ image features for every pixel in the original PBA image.

3.2. Multi-Scale Point Cloud Feature Extraction

In our work, features are extracted from 3D laser point clouds on multiple scales. However, when the neighborhood radius is expanded at a linear rate, the number of neighborhood points of a laser point is approximately increased at a cubic speed, which greatly increases the computational burden. In order to solve this problem, the point cloud pyramid model is derived which is inspired by the image pyramid model in image processing.

Similar to the image pyramid model, the downsampling algorithm is utilized for the original point clouds to build the point cloud pyramid model. The voxel model is used to divide the laser point cloud to be downsampled into different 3D grids. Then the center of gravity of the laser points in each voxel (3D grid) is calculated to represent all the points in the voxel. An illustration of the point cloud pyramid model is shown in Figure 11, in which the bottom layer is the original laser point cloud. Then a fixed number of laser points are selected as neighborhood points in different layers of the point cloud pyramid model.

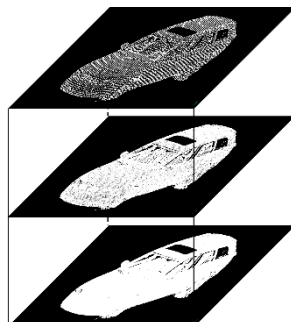


Figure 11. An illustration of point cloud pyramid model.

After determining the neighborhood range of each laser point, feature extraction will be performed which includes statistical features, geometric morphological features, and histogram features.

3.2.1. Statistical Features

Let the total number of laser points in the current neighborhood be $(k + 1)$, and the coordinate of the lowest point in the neighborhood be h_{\min} . In our work, five statistical features are extracted, which are:

- h , the absolute height of the laser point;
- $\Delta h = h - h_{\min}$, the relative height between the laser point and the lowest laser point in the neighborhood;
- $\sigma = \sqrt{\frac{1}{k} \sum_{i=1}^k (h_i - \bar{h})^2}$, the standard deviation of the laser point's height in the neighborhood;
- r , the radius of the maximum bounding sphere of the neighborhood;
- $d = \frac{k+1}{\frac{4}{3}\pi r^3}$, the density of the laser points in the neighborhood.

3.2.2. Morphological Features

According to the summary in [15], a covariance matrix is adopted to describe the 3D laser point cloud in the neighborhood, where p_c is the current query point and p_i is the neighborhood point around the query point. The covariance matrix can be expressed as:

$$\text{COV} = \frac{1}{k} \sum_{i=1}^k (p_i - p_c)(p_i - p_c)^T \quad (8)$$

which is a three-dimensional positive definite matrix. By eigen decomposition, three eigenvalues $\lambda_1, \lambda_2, \lambda_3$ (let $\lambda_1 \geq \lambda_2 \geq \lambda_3 \geq 0$) and three eigenvectors e_1, e_2, e_3 corresponding to $\lambda_1, \lambda_2, \lambda_3$ are obtained, respectively. In our work, nine morphological features are extracted, which are Linearity L_λ , Planarity P_λ , Sphericity S_λ , Omnivariance O_λ , Anisotropy A_λ , Eigenentropy E_λ , Sum Σ_λ , Change of Curvature C_λ and Verticality V_λ . These features can be calculated as follows:

$$\begin{aligned} L_\lambda &= \frac{(\lambda_1 - \lambda_2)}{\lambda_1}, P_\lambda = \frac{(\lambda_2 - \lambda_3)}{\lambda_1}, S_\lambda = \frac{\lambda_3}{\lambda_1}, O_\lambda = (\lambda_1 \cdot \lambda_2 \cdot \lambda_3)^{\frac{1}{3}}, A_\lambda = \frac{(\lambda_1 - \lambda_3)}{\lambda_1}, \\ E_\lambda &= -\sum_{i=1}^3 \lambda_i \ln(\lambda_i), \Sigma_\lambda = \lambda_1 + \lambda_2 + \lambda_3, C_\lambda = \frac{\lambda_3}{\lambda_1 + \lambda_2 + \lambda_3}, V_\lambda = 1 - e_z \end{aligned} \quad (9)$$

3.2.3. Histogram Features

Fast point feature histograms (FPFH) is a set of 33-dimensional histogram features [21]. Compared to morphological features, FPFH can describe the geometric features in the query point's neighborhood in more detail and represent the roughness of the plane effectively, which can be used to distinguish two typical road surfaces (artificial ground and natural ground). As shown in Figure 12a, FPFH consists of two Simplified Point Feature Histograms (SPFH). One is composed of the query point p and its neighborhood point p_k (the points in the red circle) and the other one is composed of each neighborhood point p_k and its neighborhood points (the points in the blue circle). FPFH can be defined as follows:

$$\text{FPFH}(p) = \text{SPFH}(p) + \frac{1}{k} \sum_{i=1}^k \frac{1}{w_k} \cdot \text{SPFH}(p_k) \quad (10)$$

where k stands for the number of neighborhood points around the query point p , w_k stands for distance weight which is used to measure the density between neighborhood points and query points.

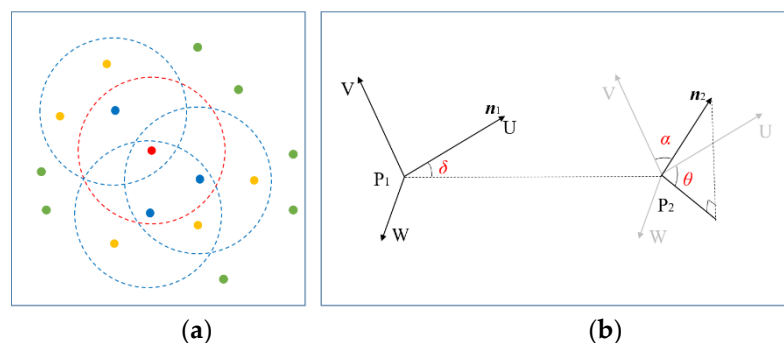


Figure 12. An illustration of FPFH feature extraction. (a) FPFH consists of two Simplified Point Feature Histograms (SPFH); (b) The position relationship between two laser points.

SPFH is composed of Simplified Point Features (SPF). SPF is a three-dimensional angular feature descriptor that represents the position relationship between two laser points. As shown in Figure 12b,

P_2 is a laser point in the neighborhood of P_1 , and n_1 and n_2 are the normal vectors of P_1 and P_2 . According to (12), the UVW coordinate system is established with P_1 as the coordinate origin:

$$\begin{cases} \mathbf{u} = \vec{n}_1 \\ \mathbf{v} = \mathbf{u} \times \frac{(p_2 - p_1)}{\|p_2 - p_1\|_2} \\ \mathbf{w} = \mathbf{u} \times \mathbf{v} \end{cases} \quad (11)$$

The angular parameter δ , α , θ are used to describe the position relationship between two laser points, which can be defined as follows:

$$\begin{cases} \alpha = \mathbf{v} \cdot \vec{n}_2 \\ \delta = \mathbf{u} \cdot \frac{(p_2 - p_1)}{\|p_2 - p_1\|_2} \\ \theta = \arctan(\mathbf{w} \cdot \vec{n}_2, \mathbf{u} \cdot \vec{n}_2) \end{cases} \quad (12)$$

Although FPFH can describe the geometric characteristics of the laser point cloud in more detail, it increases the computational burden significantly. Therefore, we only extract FPFH features for the laser point at the bottom of the point cloud pyramid, while the other 14-point cloud features (five statistical features and nine morphological features) are extracted for each laser point of the point cloud pyramid.

3.3. Classification with Random Forest and Reclassification Based on the Contextual Information

In this paper, the Random Forest classifier is used to perform feature screening on the extracted high-dimensional features, and the initial classification of the 3D laser point clouds is implemented. Since this method does not consider the contextual information between laser points, the credibility of classification results is low for the objects with similar local features (such as eaves and vegetation). In order to make full use of the contextual information between laser points, the classification results are remapped into the PBA images, and superpixel segmentation is performed on the PBA images. Within each superpixel block, the classification is performed again based on the results of the initial classification, so as to correct partial misclassification points and further improve the classification accuracy.

The Random Forest classifier is composed of multiple decision tree classifiers. In the training stage, some training samples are randomly selected to complete the training for each decision tree. In the classification stage, some decision trees are randomly selected and the mode of their output categories is taken as the final classification result.

Figure 13 shows the classification results by using the Random Forest classifier and the ground truth. Seven different colors are used to represent seven different categories: dark gray for artificial ground, yellow for natural ground, dark green for high vegetation, light green for low vegetation, red for buildings, dark brown for railings, and silver for cars. From the classification results, we can see that the main objects, such as buildings, ground, cars, and vegetation, can be effectively classified.



Figure 13. Classification result (a) and the ground truth (b).

By comparing the classification results with the ground truth, we can find that a large number of laser points that do not belong to vegetation are classified into vegetation. This is due to the cluttered distribution of these laser points, and the local features of these laser points are very close to those of the vegetation. Therefore, the laser point clouds will be reclassified by considering the contextual information of the 3D laser point clouds based on the PBA images.

In this paper, the SEEDS-based superpixel segmentation is performed on the PBA images [22]. For each superpixel block, if the pixel proportion of vegetation is less than a threshold, the laser points corresponding to vegetation will be reclassified into the category with the highest pixel proportion in the block. This strategy makes full use of the contextual information of the 3D laser point cloud in 2D images, which can reduce the error rate of the point cloud classification.

As shown in Figure 14, the initial classification result based on the Random Forest classifier is at the top left and the reclassification result is at the top right. The bottom left images and the bottom right images are local details in enlarged images of the initial classification result and reclassification result, respectively. After reclassification, most of the point clouds that were previously misclassified are corrected.

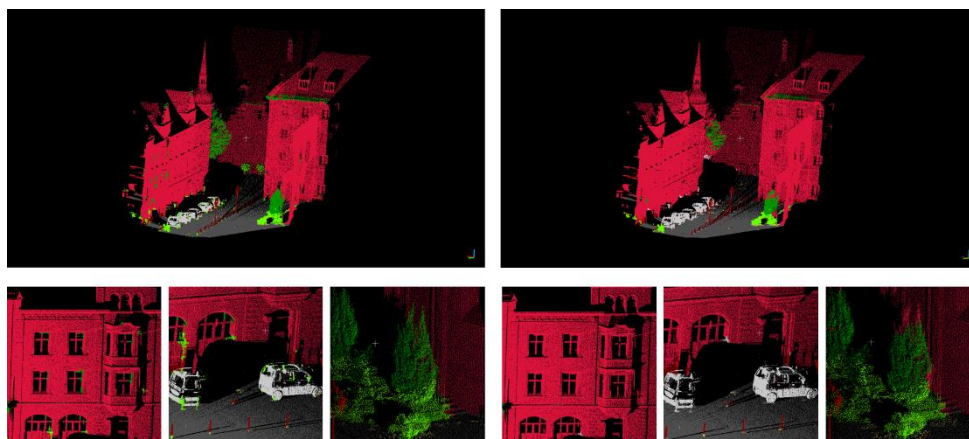


Figure 14. Initial classification result (**top left**) and reclassification result (**top right**); local details of the initial classification result (**bottom left**) and local details of reclassification result (**bottom right**).

4. Experimental Results

4.1. Classification Results of 3D Point Clouds Obtained in Fixed-Point Scanning Mode

In this subsection, a 3D laser point cloud dataset published by ETH Zurich is selected to verify the algorithm. This dataset includes 15 typical scenes. Two typical scenes are selected for testing, and the remaining scenes are used for training. The two testing sets contain seven categories which are represented by seven different colors: dark gray for artificial ground, yellow for natural ground, dark green for high vegetation, light green for low vegetation, red for buildings, dark brown for railings, and silver for cars. The category distribution of the two testing sets is shown in Table 1.

Table 1. The category distribution of the two testing sets.

	Artificial Ground	Natural Ground	High Vegetation	Low Vegetation	Building	Railing	Car
Testing Set A	3932832	0	30402	7685	4129066	492534	89467
Testing Set B	13663073	32956	276317	1326052	8755941	3499737	257352

A four-layer image pyramid model and a six-layer point cloud pyramid model are established. The resolution of the bottom image is 1440×720 and the density of the bottom point cloud is 25^3 points/m³. For the image pyramid model, the texture features are extracted at six scales. For the point cloud pyramid model, the 10 and 20 nearest laser points are selected respectively as neighborhood

points for each laser query point. The 14-point cloud features (five statistical features and nine morphological features) are extracted for each laser query point. Thirty-three FPFH features are only extracted for the laser point at the bottom of the point cloud pyramid. Therefore, for each laser point, 24 (4×6) image texture features and 201 ($6 \times 2 \times 14 + 33$) point cloud features are extracted. The Random Forest classifier consists of 200 decision trees with a depth of 15. The images on the left of Figures 15 and 16 show the initial classification results. The recall rate and precision rate are given in Tables 2 and 3, respectively.

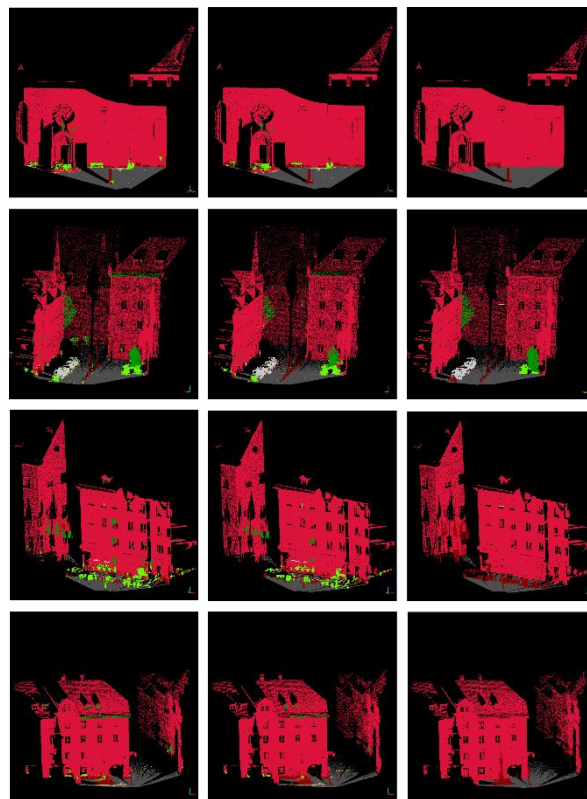


Figure 15. The classification results for Testing Set A (initial classification results (**left**); reclassification results (**middle**); ground truth (**right**)).

Table 2. The classification evaluation metrics of Testing Set A.

	Artificial Ground	Natural Ground	High Vegetation	Low Vegetation	Building	Railing	Car
Precision	97.87%	-	79.08%	94.16%	94.41%	51.51%	89.07%
Recall	99.44%	-	20.72%	2.75%	98.86%	73.43%	91.90%

Table 3. The classification evaluation metrics of Testing Set B.

	Artificial Ground	Natural Ground	High Vegetation	Low Vegetation	Building	Railing	Car
Precision	81.28%	53.83%	76.33%	86.02%	84.49%	61.41%	71.66%
Recall	99.71%	0.7%	23.61%	57.61%	99.58%	87.12	12.57%

According to the initial classification results, it can be seen that the recall rates of vegetation and natural ground are very low. A large number of laser points that belong to cars and buildings are misclassified into vegetation and a large number of laser points that belong to artificial ground are misclassified into natural ground. For misclassified categories (vegetation and natural ground), reclassification will be carried out.

The superpixel segmentation is used for reclassification. In this paper, the PBA image is segmented into 2025 superpixel blocks. For each superpixel block, if the pixel proportion of vegetation is less

than 1/8, the laser points corresponding to vegetation will be reclassified into the category with the highest pixel proportion in the block. If the superpixel block contains both natural ground and artificial ground, the laser points belonging to the category with a small proportion will be reclassified into that with a larger proportion. The images in the middle of Figures 15 and 16 are the reclassification results and the images on the right of Figures 15 and 16 show the ground truth. The recall rate and precision rate are given in Tables 4 and 5.

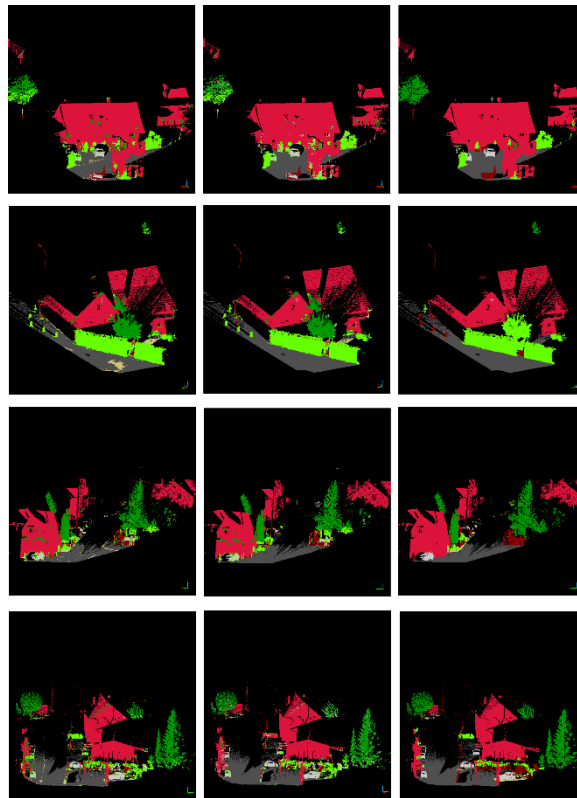


Figure 16. The classification results for Testing Set B (initial classification results (**left**); reclassification results (**middle**); ground truth (**right**)).

Table 4. The reclassification evaluation metrics of Testing Set A.

	Artificial Ground	Natural Ground	High Vegetation	Low Vegetation	Building	Railing	Car
Precision	97.95%	-	62.69%	96.27%	95.99%	71.59%	98.58%
Recall	99.48%	-	45.48%	5.27%	98.21%	77.58%	90.25%

Table 5. The reclassification evaluation metrics of Testing Set B.

	Artificial Ground	Natural Ground	High Vegetation	Low Vegetation	Building	Railing	Car
Precision	99.53%	2.02%	69.79%	84.56%	85.82%	61.41%	75.02%
Recall	99.49%	4.29%	30.75%	66.15%	99.46%	86.58	11.19%

After reclassification, the recall rates of vegetation and natural ground have been improved. However, for Testing Set A, the recall rate of low vegetation is still not high. A large number of laser points belonging to motorcycles are classified into low vegetation. Since motorcycles are not considered as a category, the lower recall rate is acceptable for low vegetation.

In addition, for Testing Set B, the precision rate of natural ground classification declined dramatically due to the disparity in the area between artificial ground and natural ground. After reclassification, some laser points belonging to natural ground are classified into artificial ground. Although this strategy sacrifices the precision rate of natural ground classification, it improves

the precision rate of artificial ground classification greatly and the classification effect of the whole scene is better.

4.2. Classification Results of 3D Point Clouds Obtained in On-the-Fly Scanning Mode

In this subsection, a 3D laser point cloud dataset published by MINES ParisTech is selected to verify the algorithm. Since the data are obtained by on-the-fly scanning, pre-processing is performed to filter out some laser points with large errors. Simple cropping and downsampling are also performed to remove the laser points scanned into the interior of the building. A typical scene is selected for testing and the category distribution is shown in Table 6.

Table 6. The category distribution of the testing set.

Artificial Ground	Building	Railing	Car
886,463	2,136,908	3912	232,714

In on-the-fly scanning mode, multiple PBA images are needed to fully represent the 3D scene. As shown in Figure 17, the red ray approximates the trajectory of the data acquisition vehicle, and the length is about 80 m. The five red triangles are viewpoints selected on the acquisition trajectory. The images on the top and bottom of Figure 17 are the PBA images obtained from the five viewpoints.

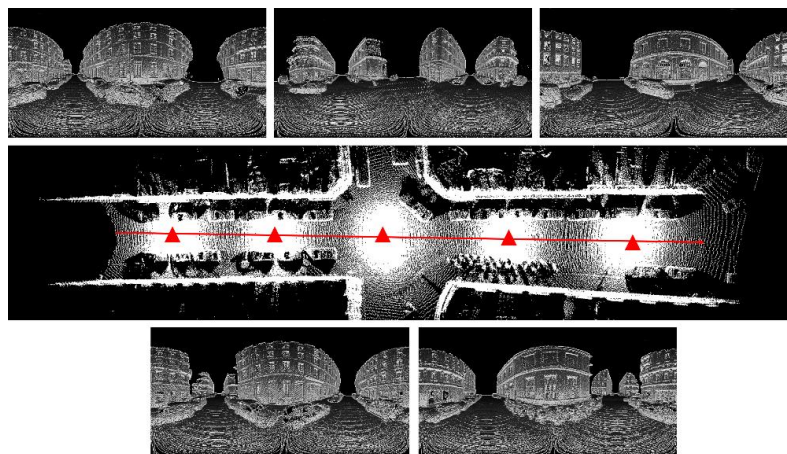


Figure 17. PBA images obtained from different viewpoints (**top** and **bottom**); 3D laser point cloud obtained by inverse mapping of PBA map (**middle**).

Due to the low density of data acquired by on-the-fly scanning, the resolution of the image at the bottom of the image pyramid is selected as 720×360 and FPFH features are not extracted. Since the scene contains only four categories, reclassification is not carried out. The classification results are shown in Figure 18.

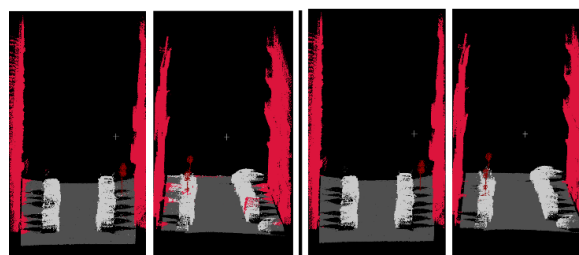


Figure 18. Point cloud classification results for MINES ParisTech Dataset (two images on the **left**) and the ground truth (two images on the **right**).

We also compare the classification results with Weinmann’s work [14]. Weinmann selected a fixed neighborhood scale for point clouds and 21-dimensional features were extracted for each laser point. The comparison of classification results is shown in Table 7. It can be seen that the method proposed in this paper has obvious advantages for the classification of small objects such as railings and cars.

Table 7. The comparison of classification results.

		Artificial Ground	Building	Railing	Car
Our results	Precision	93.80%	99.85%	76.76%	98.9%
	Recall	99.45%	98.32%	76.35%	91.45%
Weinmann’s results	Precision	96.4%	96.2%	5.5%	75.5%
	Recall	90.2%	95.7%	97.4%	60.6%

5. Conclusions

This paper presents an approach of 3D laser point cloud classification to accomplish outdoor scene understanding in urban environments. To improve the performance of point cloud classification, a new transformation model is proposed to transform point clouds to PBA images. Due to the correspondence between the original point cloud and the PBA image, multiple-scale features are extracted from both point clouds and PBA images, and then the Random Forest classifier is adopted to get initial classification results. To correct the misclassification points, reclassification is performed by remapping the classification results into the PBA images and using superpixel segmentation. Finally, we have conducted a series of experiments on two public datasets named ETH Zurich and MINES ParisTech, and testing results demonstrate the validity and the robustness of the proposed method.

Author Contributions: Conceptualization, Y.L., Y.G. and F.Y.; Methodology, Y.L. and Y.G.; Software, Y.G. and F.Y.; Formal analysis, Y.L. and F.Y.; Data curation, Y.G.; Writing—original draft preparation, Y.L. and Y.G.; Writing—review and editing, Y.L. and Y.G.; Supervision, F.Y. and Y.Z.

Funding: This work was supported by the National Natural Science Foundation of China (61973049, U1608253).

Conflicts of Interest: The authors declare no conflict of interest.

References

- Balado, J.; Martínez-Sánchez, J.; Arias, P.; Novo, A. Road Environment Semantic Segmentation with Deep Learning from MLS Point Cloud Data. *Sensors* **2019**, *19*, 3466. [[CrossRef](#)] [[PubMed](#)]
- Cao, F.; Zhuang, Y.; Zhang, H.; Wang, W. Robust place recognition and loop closing in laser-Based SLAM for UGVs in urban environments. *IEEE Sens. J.* **2018**, *18*, 4242–4252. [[CrossRef](#)]
- Zhou, S.; Kang, F.; Li, W.; Kan, J.; Zheng, Y.; He, G. Extracting diameter at breast height with a handheld mobile LiDAR system in an outdoor environment. *Sensors* **2019**, *19*, 3212. [[CrossRef](#)] [[PubMed](#)]
- Niola, V.; Rossi, C.; Savino, S. A new real-time shape acquisition with a laser scanner: First test results. *Robot. Comput.-Integr. Manuf.* **2010**, *26*, 543–550. [[CrossRef](#)]
- Maligo, A.; Lacroix, S. Classification of outdoor 3D LIDAR data based on unsupervised Gaussian mixture models. *IEEE Trans. Autom. Sci. Eng.* **2017**, *14*, 5–16. [[CrossRef](#)]
- Qiu, Z.; Zhuang, Y.; Yan, F.; Hu, H.; Wang, W. RGB-DI images and full convolution neural network based outdoor scene understanding for mobile robots. *IEEE Trans. Instrum. Meas.* **2019**, *68*, 27–37. [[CrossRef](#)]
- Che, E.; Jung, J.; Olsen, M. Object Recognition, Segmentation, and classification of mobile laser scanning point clouds: A state of the art review. *Sensors* **2019**, *19*, 810. [[CrossRef](#)] [[PubMed](#)]
- Kim, S.; Lee, S.; Ho, Y. Three-dimensional natural video system based on layered representation of depth maps. *Consum. Electron. IEEE Trans.* **2006**, *52*, 1035–1042.
- Nüchter, A.; Hertzberg, J. Towards semantic maps for mobile robots. *Robot. Autonom. Systems* **2008**, *56*, 915–926. [[CrossRef](#)]
- Scaramuzza, D.; Harati, A.; Siegwart, R. Extrinsic self calibration of a camera and a 3d laser range finder from natural scenes. In Proceedings of the 2007 IEEE/RSJ International Conference on Intelligent Robots and Systems, San Diego, CA, USA, 29 October–2 November 2007; pp. 4164–4169.

11. Zhuang, Y.; He, G.; Hu, H.; Wu, Z. A novel outdoor scene- understanding framework for unmanned ground vehicles with 3D laser scanners. *Trans. Inst. Meas. Control* **2015**, *37*, 435–445. [[CrossRef](#)]
12. Zhang, X.; Zhuang, Y.; Hu, H.; Wang, W. 3-D Laser-Based Multiclass and Multiview Object Detection in Cluttered Indoor Scenes. *IEEE Trans. Neural Netw. Learn. Syst.* **2017**, *28*, 177–190. [[CrossRef](#)] [[PubMed](#)]
13. Weinmann, M.; Urban, S.; Hinz, S.; Jutzi, B.; Mallet, C. Distinctive 2D and 3D features for automated large-scale scene analysis in urban areas. *Comput. Graph.* **2015**, *49*, 47–57. [[CrossRef](#)]
14. Weinmann, M.; Jutzi, B.; Mallet, C. Semantic 3D scene interpretation: A framework combining optimal neighborhood size selection with relevant features. *ISPRS Ann. Photogramm. Remote Sens. Spat. Inf. Sci.* **2014**, *2*, 181–188. [[CrossRef](#)]
15. Weinmann, M.; Jutzi, B.; Mallet, C. Feature relevance assessment for the semantic interpretation of 3D point cloud data. *ISPRS Ann. Photogramm.* **2013**, *5*, 313–318. [[CrossRef](#)]
16. Breiman, L. Random Forests. *Mach. Learn.* **2001**, *45*, 5–32. [[CrossRef](#)]
17. Munoz, D.; Bagnell, J.A.; Vandapel, N.; Hebert, M. Contextual classification with functional Max-Margin Markov Networks. In Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, Miami, FL, USA, 20–25 June 2009; pp. 975–982.
18. Najafi, M.; Namin S, T.; Salzmann, M.; Petersson, L. Non-associative Higher-Order Markov Networks for Point Cloud Classification. In Proceedings of the European Conference on Computer Vision, Zurich, Switzerland, 5–12 September 2014; pp. 500–515.
19. Burt, P.J.; Adelson, E.H. The Laplacian Pyramid as a Compact Image Code. *Read. Comput. Vis.* **1987**, *31*, 671–679.
20. Ojala, T.; Pietikäinen, M.; Harwood, D. A Comparative Study of Texture Measures with Classification Based on Feature Distributions. *Pattern Recognit.* **1996**, *29*, 51–59. [[CrossRef](#)]
21. Rusu, R.B.; Blodow, N.; Beetz, M. Fast point feature histograms (FPFH) for 3D registration. In Proceedings of the IEEE International Conference on Robotics and Automation, Montreal, Canada, 20–24 May 2009; pp. 1848–1853.
22. Bergh, M.V.D.; Boix, X.; Roig, G.; Gool, L. SEEDS: Superpixels extracted via energy-driven sampling. *Int. J. Comput. Vis.* **2015**, *111*, 298–314. [[CrossRef](#)]



© 2019 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).