

三维点云场景数据获取及其场景理解关键技术综述

李勇¹, 佟国峰^{1*}, 杨景超², 张立强^{3**}, 彭浩¹, 高华帅¹

¹ 东北大学 信息科学与工程学院, 辽宁 沈阳 110819

² 河北交通职业技术学院 电气与信息工程系, 河北 石家庄 050091

³ 北京师范大学 遥感科学国家重点实验室, 北京 100875

摘要 场景理解是信息科学里的重要研究内容, 而三维数据相比于二维数据有着众多优势, 目前点云的获取有多方式, 且不同获取方式的点云具有不同的特点, 此外, 基于点云的三维场景理解中的关键技术研究还没有完整、系统的综述。本文总结了不同方式的点云获取方法, 并对不同的点云数据及相关数据库进行对比分析。然后, 基于目前三维场景理解的研究进展, 针对三维场景理解中的点云滤波、特征提取与点云分割和点云语义分割等技术进行了对比分析与总结。通过对近些年国内外文献的总结梳理, 凝练出三维场景理解关键技术中存在的问题, 并对三维场景理解问题的发展趋势做了展望。基于点云的三维场景理解因其数据的丰富性而被广泛应用在众多领域中, 但是目前基于三维点云的场景理解效果还有众多内容有待深入研究, 尤其是针对具有颜色信息的激光点云的场景理解。

关键词 三维点云; 场景理解; 语义分割; 研究进展

中图分类号 TN958.98 **文献标识码** A

Summary of 3D point cloud scene data acquisition and the key technologies for scene understanding

LI Yong¹, TONG Guo-feng^{1*}, YANG Jing-chao², ZHANG Li-qiang^{3*}, PENG Hao¹,
Gao Hua-shuai¹

¹College of Information Science and Engineering, Northeastern University, Shenyang, Liaoning, 110819, China

²Department of electrical and Information Engineering, Hebei Jiaotong Vocational and Technical College,
Shijiazhuang, Hebei, 050091, China

³The State Key Laboratory of Remote Sensing Science, Beijing Normal University, Beijing 100875, China

Abstract Scene understanding is an important research content in information science. Compared with two-dimensional data, three-dimensional data has many advantages. At present, there are many ways to acquire point clouds, and point clouds with different acquisition methods have different characteristics. In addition, there is lacking of complete and systematic research review on the key techniques for understanding 3D scenes. This paper summarizes the different methods of point cloud acquisition. Then, we compare and analyze different point cloud data and related databases. Next, based on the current research progress of 3D scene understanding, the techniques of point cloud filtering, feature extraction, point cloud segmentation and point cloud semantic segmentation in 3D scene understanding are compared and summarized. By summarizing the references at home and abroad in recent years, the problems in the key technologies for understanding the three-dimensional scene are concise. And the development trend of the 3D scene understanding problems have been forecasted. The 3D scene understanding

based on point cloud is widely used in many fields because of its richness information of data. However, there are still many contents to be studied in depth based on the scene understanding effect of 3D point cloud, especially for the scene understanding of laser point cloud with color information.

Keywords 3D point cloud; scene understanding; semantic segmentation; research progress

OCIS codes 150.1135; 150.6910; 280.3640

基金项目:国家自然科学基金(61175031); 国家 863 计划资助项目(2012AA041402)。

收稿日期: ; **改回日期:**

作者简介: 李勇 (1991-), 男, 博士研究生 (在读东北大学模式识别与智能系统专业), 主要研究方向: 计算机视觉、点云处理、深度学习等。E-mail: 842384077@qq.com.

*佟国峰(1973-), 男, 教授, 博士生导师, 2001 年于东北大学获得博士学位, 现为东北大学信息科学与工程学院教授, 主要从事 SLAM、计算机视觉、点云处理等算法研究。E-mail: tongguofeng@ise.neu.edu.cn.

**张立强(1976-), 男, 教授, 博士生导师, 2004 年于中国科学院遥感应用研究所获得博士学位, 现为北京师范大学地理科学学部教授, 主要从事激光点云处理、遥感影像识别和三维重建等算法研究。E-mail: zhanglq@bnu.edu.cn.

0 引 言

场景理解是在空间感知的基础上, 对场景进行认知和推断的过程。然而, 目前场景理解还没有严格统一的定义。根据目前国内外相关的研究, 场景理解主要包含目标检测与识别、物体间的关系以及语义分割等科学内容^[1-3]。场景理解是人工智能的前沿研究方向之一, 在机器人、虚拟现实、自动驾驶和激光遥感测量等领域具有广泛的应用^[4-7], 例如, 美国 NHSTA (国家高速路安全管理局) 定义的 L5 级自动驾驶、百度高德等的三维语义地图以及现代战场环境下的单兵作战机器人等都需要高精度、高性能的环境感知和场景理解能力。

基于二维图像数据(如 RGB 图像, 遥感图像)的场景理解已被广泛的研究与应用。二维图像提供了物体的颜色、纹理及光谱等关键特征。然而, 二维图像所复现的场景可能会存在光照不均匀、目标遮挡等现象, 并且在生成过程中易受周围环境的影响。同时, 二维图像不包含深度信息。因此, 基于二维图像数据的场景理解鲁棒性差, 且难以精确地提取目标轮廓、空间位置等关键信息。

基于三维建模数据(如点云、CAD 模型等)的场景理解能够感知场景的三维空间信息。近年来, 传感器的快速发展, 使得三维数据的获取越来越容易。此外, 随着计算机视觉、人工智能技术以及机器人技术的发展, 关于各种空间物体的识别和理解的研究已经引起了越来越多的关注。目前, 越来越多的学者针对三维场景理解的关键技术展开了大量研究^[5,8-11]。近五年来, 针对室内与室外环境的三维点云进行场景理解的方法逐年递增, 为了对现有方法进行系统的研究与阐述, 本文检索了近年来关于场景理解的相关文献, 通过整理、阅读和研究发现, 目前的关于场景理解的综述文章都集中在基于二维图像的场景理解中, 还没有对于三

维场景理解的相对完整、系统的综述，也缺乏二维图像和三维数据的对比研究。因此，本文主要针对三维场景理解的关键技术，通过对二维图像和三维数据对比的方式，对三维场景数据的获取、现有三维场景理解数据集以及现有场景理解的方法展开了论述，最后，对于三维场景理解需解决的关键问题进行了展望。

1 三维点云场景数据获取

1.1 三维重建数据获取的传感器

常用于获取三维重建数据的传感器有：普通摄像头、全景摄像机系统、深度相机和激光雷达等。相比较于摄像头获取的二维图像，三维点云可以提供更多几何结构等空间信息，同时，三维点云获取设备(如：Kinect、TOF、激光雷达)的快速发展，使得相关领域的研究人员能够更容易以越来越低的成本获得精度越来越高的三维点云数据。特别是激光扫描技术获取空间信息数据具有快速、精度高和便于处理等优点，利用该技术可以深入到复杂的现场环境及空间中进行扫描操作，直接得到各种大型的、复杂的、不规则、标准或非标准的实体或实景的三维数字几何信息。

1.2 三维数据重建方法

目前，常用的三维数据获取方法主要分为：基于立体视觉、激光和视觉与激光联合重建的方法。基于立体视觉的三维重建方法进行三维重建本质上均是模仿人眼的“视差原理”，这种使用视觉方法进行三维重建的鲁棒性很差。虽然在比较理想的情况下，目前好的多目视觉方案精度可以达到厘米级别，但是当外界光线由强变弱时，或者物体表面纹理信息缺乏时，容易产生误匹配，精度会大打折扣。因此，基于立体视觉进行三维重建，其精度会随着周围环境的变化而受到影响，不但不具有良好的实时性反而具有相当大的局限性，限制了此方法在三维重建领域的广泛应用。而基于 Kinect 和 TOF 的三维重建，由于摄像头拍摄视野范围比较小，只能探测几米的范围，所以一般只能应用于室内场景。基于激光的三维重建方法，虽然激光的探测距离相对较远，但是得到的只有场景的几何结构信息，缺乏场景的彩色信息，因此不利于后期基于场景的语义分割、场景理解等上层研究。

此外，基于雷达 SAR 成像和高清影像可以生成地面场景的彩色点云数据，例如 Seasat SAR, Almaz SAR, JERS-1 SAR 等。这类数据虽然具有颜色信息且为大尺度场景，但是地面底部信息精度不高，会有一定的遮挡和信息缺失。然而，利用相机和激光点云的联合三维重建不但可以得到大尺度场景的几何结构信息和场景的彩色信息，而且能够适应地面上更多的环境场景，从而可以获得精度较高、信息丰富、场景多元化的三维重建效果。现有的联合重建方法中三维点云数据获取通常是如图 1(a)所示的基于车载的移动测绘系统，如中国专利号为 CN202782968U 的基于激光扫描与全景影像的车载测量集成系统，它提供了一个运用高

精度全球定位技术、激光扫描直接测量技术和实景全景图像测量技术相结合的综合性技术。但现阶段移动测量系统均需要 GNSS 和惯导，只能用于室外环境，由于室内没有 GNSS 信号，现有的移动测量系统无法正常工作。虽然，如图 1(b)所示的固定式激光扫描可以用于室内室外的环境，但是作业方式效率低下，复杂场景需要大量换站，然后进行点云拼接。而且，即时定位与地图构建（SLAM）的方法在移动的三维重建问题上具有较好的应用，如欧洲专利号为 EP2913796A1 的基于 SLAM 技术的推车，使用了二维的 SLAM 技术，但该方法只能在水平面上使用，如图 1(c)给出了三种手推车式采集系统。相比于上述采集系统，背负式激光雷达可移动性强，受空间限制最小，获取的点云数据更为完整。如图 1(d)所示的背负式采集系统，不仅可以用于室内场景，也适用于非水平面采集场景，采集的数据可以满足测量等精细作业方面的要求。基于背负式全景影像和激光点云数据采集系统针对场景通过性好，实现了场景的点云和全景影像全覆盖，是新一代的数据采集平台。



图 1 三维数据采集系统

Fig.1 Three-dimensional data acquisition system. (a) Vehicle systems (b) Fixed scanners (c) Trolley type acquisition system (d) Backpack type acquisition system

2 三维场景数据类型对比及数据库

2.1 场景理解不同数据的对比

不同的数据类型进行场景理解时具有不同的效果。图 2(a)为基于二维道路 RGB 图像的场景语义分割示例。通过对图 2(a)中道路图像进行语义分割，实现对场景中植物、道路、标识线、车辆以及建筑等物体的感知。但当场景内目标被遮挡或受光照影响时，如图 3 所示，基于二维图像数据的场景语义理解不能精准地感知目标的轮廓信息。而对于三维点云数据，如图 2(b)为基于激光点云的目标检测示例，首先通过激光点云复现真实场景中的空间结构、物体形状等信息，进而实现对场景中汽车、行人等目标的精确检测。与图 1(b)相比，图 1(c)中三维场景模型具有颜色信息，更有利于场景的细节理解。

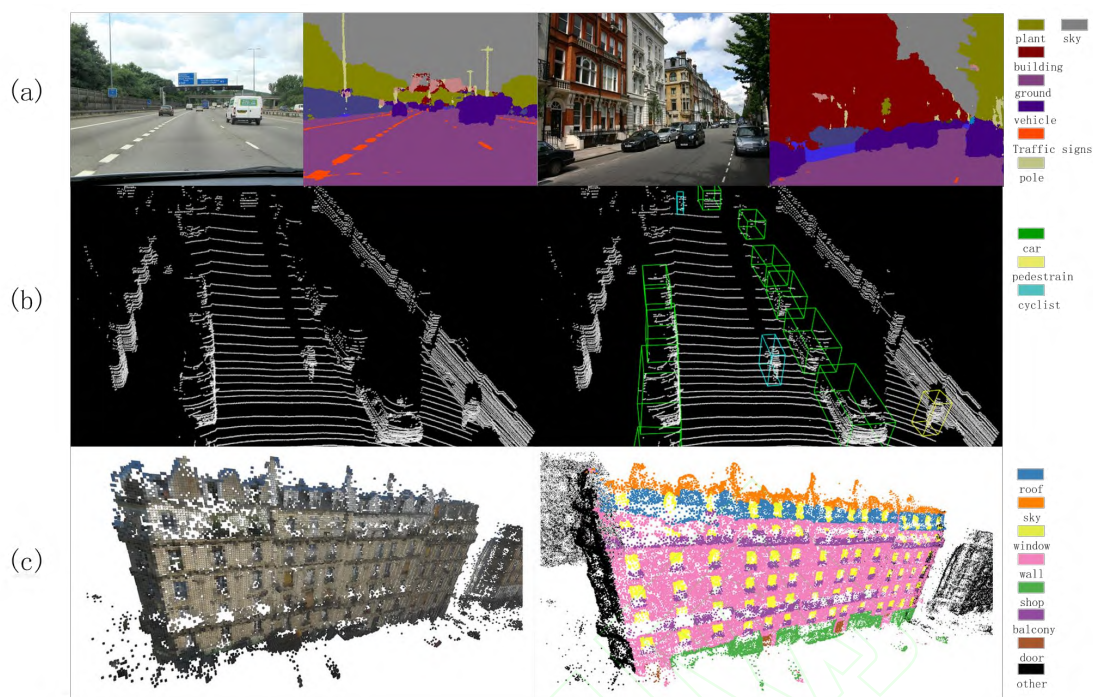
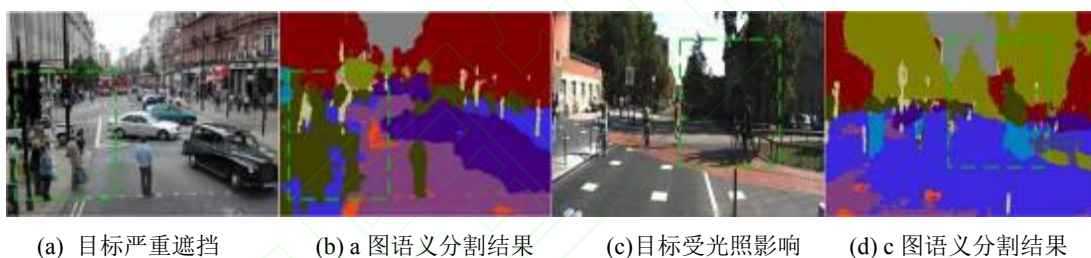


图 2 场景理解示例

Fig. 2 Examples of scene understanding.



(a) 目标严重遮挡

(b) a 图语义分割结果

(c)目标受光照影响

(d) c 图语义分割结果

图 3 二维图像受环境因素影响示例





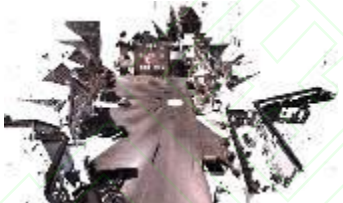
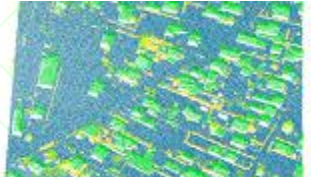

Fig.3 2D images affected by environmental factors.(a) Target severe occlusion, (b) Semantic segmentation of (a) , (c)Target affected by light, (d) Semantic segmentation of (c).

为了更加明显的对比各种场景理解的数据，对现有的数据进行了分类比较，表 1 所示为面向场景理解的不同数据类型，通过对常见的二维图像数据与三维建模数据的进行对比可知：二维 RGB 图像具有丰富的场景信息，在细节理解上优势明显；Kinect 探测距离较近，适用于室内场景的三维建模；基于激光雷达的点云数据具有高精度的三维空间信息和激光强度信息，此外，激光雷达获取空间信息数据具有速度快、探测距离远以及精度高等优点，可以深入到复杂的室外环境或空间进行三维建模。然而，基于激光雷达的点云数据缺乏颜色信息，这给场景的精细理解增加了困难。因此，具有颜色信息的激光点云对场景理解全面理解的具有明显优势。

表 1 场景理解不同数据的优缺点对比

Table1 Comparison of the advantages and disadvantages of different data in the scenario

Name	Instance	Advantage	Disadvantage
------	----------	-----------	--------------

2D RGB image		2D information, color textures, etc.	Image is susceptible to illumination, lacks depth information, cannot directly acquire geometric information.
Point cloud data based on image 3D reconstruction		3D scenes, color textures and other sparse point clouds, with depth information, distance	Point cloud is susceptible to light and environment, and reconstruction information is lost
Kinect-based RGB-D point cloud data		3D indoor small scene, dense point cloud, only for close range, with depth.	Point cloud is susceptible to light, only for indoor scenes, small field of view.
Point cloud data based on vehicle lidar		high-precision, dense point clouds, and three-dimensional spatial information and intensity information. Less affected by environmental factors.	Limited by the platform, only point cloud scenes with linear road trajectories can be generated. Unable to extract scene color
Point cloud data based on static lidar		3D outdoor scene, high-precision, dense point cloud, 3D spatial information and intensity	The collection device cannot be moved, and the point cloud scene is incomplete.
Point cloud data based on aerial lidar		3D outdoor scene, high-precision information, sparse point cloud, suitable for large-scale rough modeling	Unable to reproduce ground details, laser radar cannot extract scene color information
Collision-based panoramic image and laser point cloud fusion		3D outdoor scenes, high-precision information, dense point clouds, mobile modeling, full-frame 3D space, color and intensity.	The scene color information cannot be extracted by the laser point cloud alone.

本文定义具有颜色信息的激光点云为彩色激光点云，该数据通过 RGB 图像和激光点云联合三维建模生成，不仅拥有 RGB 图像中的颜色信息，而且具备激光点云的空间结构特征，有利于目标物体多样、纹理复杂以及噪点多的室内外环境场景理解。图 4 为基于图 3（b）的数据采集系统获取的 Semantic3D^[8]彩色点云室外场景，该场景是静态扫描仪获取的围绕一个中心点扫描，丢失后部信息，不够完整的场景。Loic Landrieu 等人^[9]通过语义分割能够准

确提取出道路、草丛以及硬体等几何特征存在相似的目标物体。

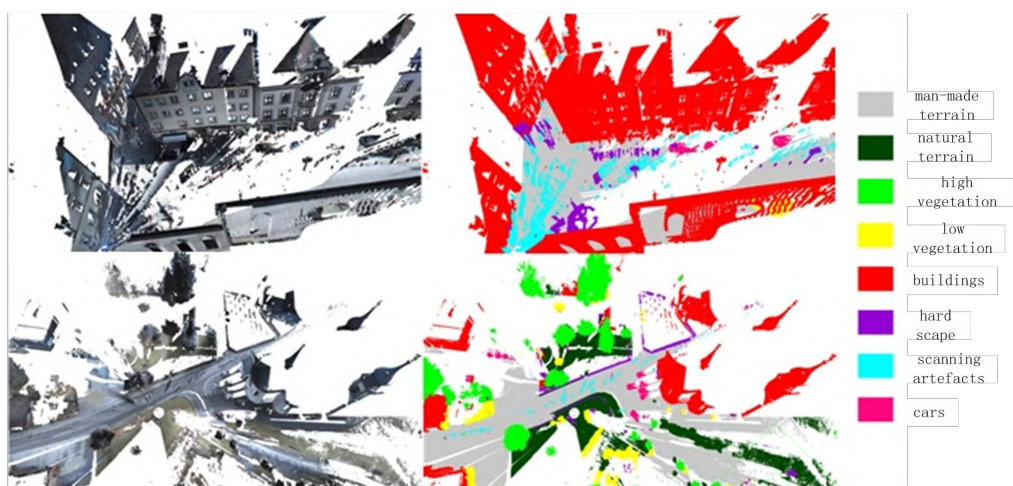


图 4 Semantic3D 彩色点云室外场景语义分割

Fig.4 Semantic3D color point cloud outdoor scene semantic segmentation.

此外，相比于车载激光雷达、静态激光雷达以及航拍激光雷达，背负式激光雷达在复杂环境下的数据采集具有更多的优势。如图 5、6 所示为基于背负式全景影像和激光点云数据采集系统所采集的室外场景点云数据。该系统实现了全景影像和激光点云的高精度配准，彩色激光点云数据具有空间信息(x,y,z)、颜色信息(R,G,B)以及强度信息(I)。并且，针对复杂的场景通过性好，不受铺装道路的行走限制，获取的点云数据更为完整，可以复现出场景中丰富的几何信息，更适合室外环境的场景理解，如图 6 所示。



图 5 小区彩色点云图像

Fig.5 Community color point cloud image

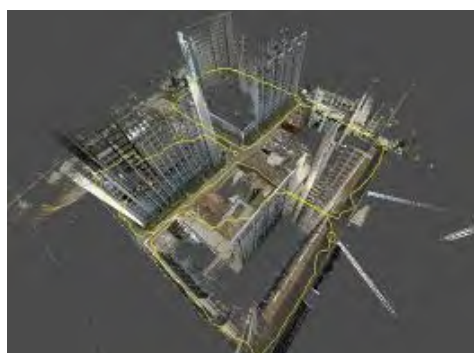


图 6 背负式全景影像移动轨迹(黄色线条)

Fig.6 Knapsack panorama image trajectory (yellow lines)

2.2、三维场景数据库

目前，三维点云公共数据库主要包含基于 Kinect 采集的数据库和基于激光点云的数据库：

- 基于 Kinect 采集的数据库，如 NYUv2 dataset^[10]和 SUN RGB-D dataset^[11]等。此类数据库面向室内场景居多，且数据精度及点云模型的完整性并不高；
- 基于激光点云的数据库，如 Laser Registration Datasets^[12]；基于单站式的静态激光扫描的公共数据集 Semantic 3D^[8]，航拍影像与激光扫描结合的点云数据库，如 Aerial Inspection Datasets^[7]；其他车载激光点云数据，如德国 Toposcan 公司通过 Lynx 平台获取的 Enschede 数据、芬兰赫尔辛基大学通过 ROAMER 平台获取的 Espoonlahti 数据、Sydney Urban Object^[13]、ModelNet^[14]等。此类数据库多数缺乏颜色信息或数据库表征场景单一或精度较低、场景完整度较差，在场景理解应用中效果较差。

内容丰富的彩色激光点云数据库，尤其是彩色激光点云能够准确复现三维场景的丰富信息，场景理解算法需要通过完整、有效的数据集进行验证，如图 7 中所示的场景示例。可见，针对室外场景理解的不同研究内容，高效、准确地标记数据库样本以及构建高精度、多场景的彩色点云数据库是场景理解未来发展趋势中的重要基础工作。



图 7 完整的彩色点云室外场景

Fig. 7 Completed Colorful Point Cloud Outdoor Scene

3 场景理解关键技术

三维场景理解是在空间认知的基础上，利用视觉分析、测绘测量等技术，挖掘场景数据中的特征与模式，形成对场景中物体的分割和识别，并对整个场景进行描述等的过程。基于三维点云的场景理解关键技术主要包括：点云的滤波、点云的特征提取、点云的分割、点云的分类、场景的语义分割与理解等内容。

3.1 点云的滤波

点云数据较复杂，目前的点云数据会存在数据缺失的现象，如扫描的模型被遮挡，造成部分点云数据的缺失；点云中通常含有噪声点及冗余信息；点云中存在旋转问题，如同一个物体，会存在不同的点云表达形式。

在点云的场景理解中，地面区域会对场景的分割等造成干扰，需要滤除地面区域，以便对地面以上感兴趣的信息进行有效的理解。点云滤波是点云处理的基本步骤，主要分为对点云中的噪声点进行滤波和对地面点的滤波。对噪声点的滤波主要有直通滤波、半径滤波和统计滤波。如表 2 所示，给出了常用的几类滤波方法的比较。

表 2 噪声点的滤波方法比较
Table 2 Noise filtering method comparison

Name	Principle	Characteristic
Through filter	According to the point cloud, the threshold should be set in the range of X, Y, and Z axes, and then the threshold is filtered to remove points that are out of range.	Fast, but not accurate enough, mostly used for rough processing in the first step
Statistical filtering	The noise point is removed according to the density of the point. The denser the point of a certain area, the larger the amount of information, the noise information is useless, and the amount of information is small. By calculating the average distance of each point to its nearest k points, The Gaussian distribution of the distances of all points in the point cloud to eliminate noise	The effect is better than the pass-through filter, which can accurately filter out the noise points inside the bounding box.
Radius filtering	Given a radius threshold, calculate the number of points in its radius for each point in the point cloud. When the number is greater than a given number of thresholds, the point is retained, otherwise the point is taken as a noise point and rejected.	Filter out the noise points inside the bounding box, faster

目前，地面滤波的方式主要有基于高程的滤波方法、基于模型的滤波方法、基于区域生长的滤波方法、基于窗口移动的方法和基于三角网的滤波方法。如表 3 所示，给出了现有地面滤波方法的现状及比较。

表 3 地面滤波方法比较
Table 3 Ground filter methods comparison

Name	Principle	Characteristic
Elevation based filtering method	According to the point distribution filtering in the point cloud, manually set or adaptively find the z-direction threshold, and filter the point where the z-value is less than the threshold in the point cloud as ground ground.	Fast, low robustness
Model based filtering method	Select a model to fit the ground (such as RANSAC-based planar model, CSF cloth simulation) and use the fitted inner point as the ground point	The algorithm is suitable for specific environments, and the robustness is poor, but the filtering effect is relatively good.
Filtering	Taking the normal vector direction of the point as the criterion	When the ground is not

method based on region growing	for regional growth, first adaptively find the point that is most likely to be the ground. Based on this, based on the angular difference between the normal vector direction of its neighborhood point and its normal vector direction, it is judged whether it grows or not. Continuous iteration to find all ground points	undulating, the algorithm can separate the ground well, but the time and space cost are relatively large
Method based on window movement	Points distributed on the ground should be of a continuous nature. Set a suitable window size to find the lowest point in the current window. Then set the threshold by the lowest point calculation model and filter out all points where the elevation difference exceeds the threshold.	Faster, but the window size is too dependent on manual settings, and only local features are considered
Triangulation based filtering method	The discrete points are connected according to a certain rule into a plurality of triangles covering the entire area without overlapping each other to form an irregular triangular network. The sparse TIN is generated by the seed point, and the slope of the model is analyzed to perform initial segmentation, and the slope is eliminated. Large triangular regions, and then through connectivity analysis to obtain features such as elevation differences for each segment	It avoids data redundancy when the terrain is flat and flat, but the data structure is complex and the space complexity is high.

3.2 点云的特征提取与分割

点云的鲁棒特征提取与分割是室外场景理解的重要研究内容。鲁棒的特征是描述三维场景的基础表达。图 8 为点云特征提取以及点云分割研究现状^[15-31]。如表 4 所示，给出了常用的点云特征描述方法及比较。点云鲁棒特征的提取能够更好地表征场景内的目标对象，而且，有效点云块的初始化分割减少了后续场景理解工作的计算量。然而，对于彩色激光点云的鲁棒特征的提取及分割，目前的研究相对较少，尚未面向室外场景。

表 4 常用的点云特征描述方法及比较

Table 4 Common method and comparison of point cloud feature description

Name	Type	Principle	Characteristic
spin image	Local feature	Count the projection coordinates of all vertices to the base plane to obtain the descriptor	Resistance to rigid transformation and background interference; sensitive to density changes
3DSC	Local feature	Counting the number of points in different grids of spherical neighborhoods to obtain descriptors	Strong discrimination, anti-noise
PFH	Local feature	Parameterize the spatial difference between the point and the neighborhood and form a multi-bit histogram	Highly robust to point cloud density changes; high computational complexity
FPFH	Local feature	Recalculate the K neighborhood by calculating the tuple of the query point and its neighbors compared to the PFH	Retains most of the recognition capabilities of PFH, and reduces computational complexity compared to PFH
SHOT	Local	Spherical neighborhood rasterization, construct	Descriptive, anti-noise; large amount of

	feature	a histogram according to the angle of the normal vector, and then concatenate the histogram	calculation, sensitive to density changes
ESF	Global feature	Describe the angle, distance, and triangle area of three random points in a point cloud	No need for pre-processing, strong feature description
VFH	Global feature	Add pilot direction to relative normal calculation after extending FPFH	Strongly discernible

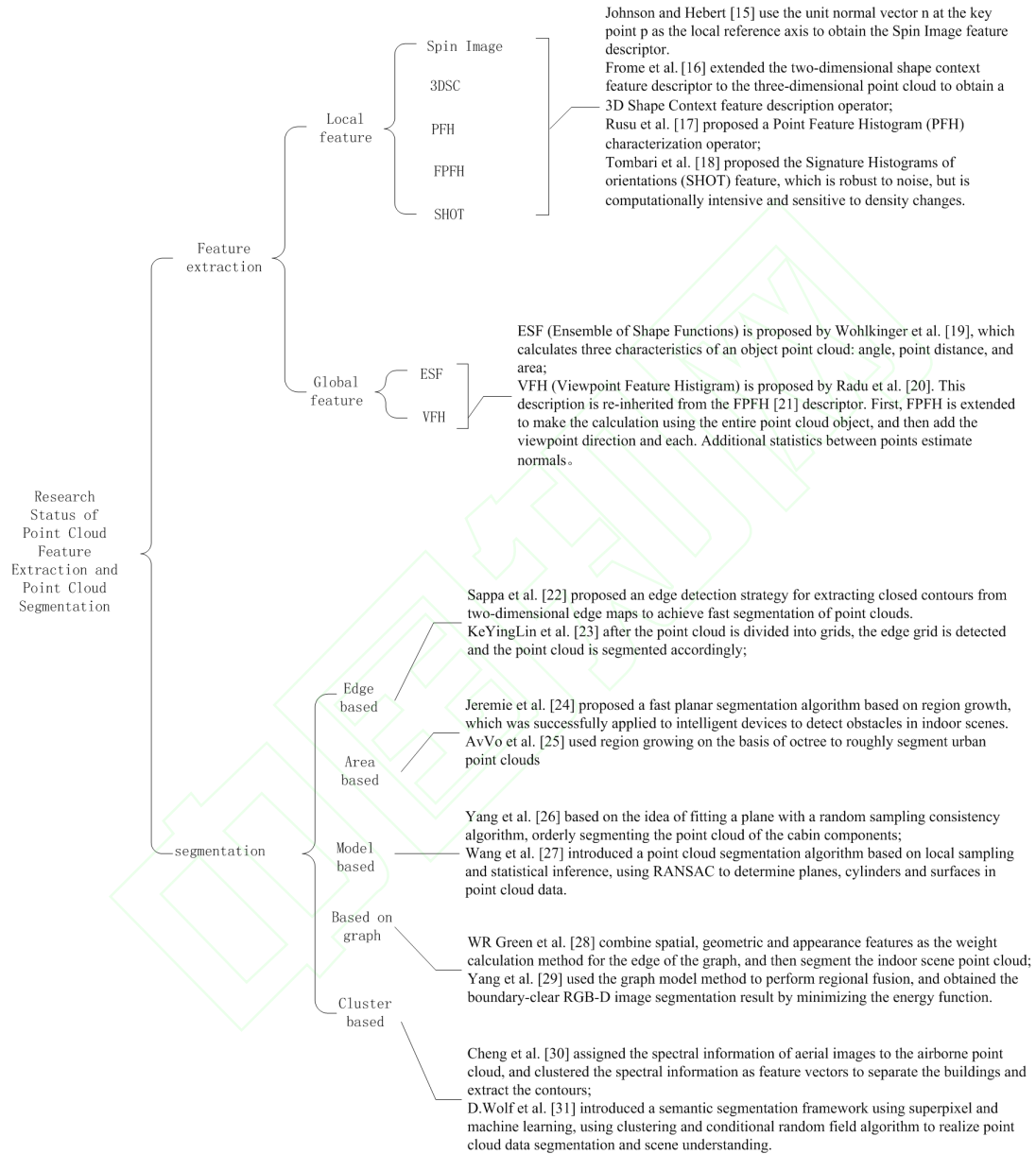


图 8 点云特征提取及点云分割研究现状

Fig.8 The research status of point cloud feature extraction and point cloud segmentation

3.3 三维点云的语义理解

语义分割是场景理解的基础和前提，即在语义分割基础上，结合其他信息，可以继续推理得到其他高层语义，如场景类别（街道、海边、城市抑或室内），真实世界中的三维结构，

场景中正在发生的事件等,进而实现场景的理解^[15-16]。点云数据中的几何信息通常以原始无序的、密度分布不均的三维空间点集的形式存在,数量从几千到上千万个不等,数据量较大^[21]。面对着越来越庞大需要处理的三维点云数据,传统的人工手段解决海量三维数据的语义解析的工作量较大,且效率低,而且计算机硬件的不断增强和软件技术的发展也使得处理三维数据的能力越来越强。尤其是机器学习和深度学习理论及其在多个领域的成功应用,为基于三维点云的语义分割与场景理解提供了重要的技术手段。

随着机器学习理论不断完善使得基于三维点云的分类、识别等场景理解研究,取得了较多的研究成果。国内外的研究者针对点云分类提出了大量算法,主要分为两大类,即基于点的分类和基于分割的分类。

- 基于点的分类是直接对每一个三维点进行分类,如 Timo Hackel 等^[32]就是根据密度构建尺度金字塔,计算每个点共 144 维特征向量,并使用分类器训练来对室外场景点云直接分类, Martin Weinmann 等^[33] 首先选取尺度大小不同的最近邻,之后提取多种特征并使用分类器分类大场景点云。但是,由于点的特征提取受到尺度限制,因此在环境多样化的室外场景,基于点的分类效果并不好。
- 基于分割的分类首先对点云数据进行分割,之后再判别每个分割块的类别。如 Xiang B 等^[34]根据法线信息对每个点构建邻接关系,并组成大的分割块再利用 SVM 判别来实现城市道路场景的点云分类。AK Aijazi 等^[35]将三维点体素化后合并成超体素,再根据设定的阈值对超体素进行划分进而完成点云分类。基于分割的分类相对计算速度较快,能提取更加丰富的特征信息。然而,该类算法对前期的初始化分割比较敏感。

这类基于机器学习的方法主要是通过大量样本训练有效的分类器,然而,有效的特征描述是这个分类模型的关键内容。用于点云描述的特征如本文 3.2 节所述,虽然目前提出了大量的特征描述方法,但是其表征能力有限,并不适用于所有的点云。目前点云分类常用的机器学习分类器主要有 support vector machine (SVM)^[36], cascaded AdaBoost^[37], and random forest^[38,39]等,但这些分类器的训练还是依赖于有效特征的选取。

此外,室外场景彩色激光点云的数据量大、目标信息多样以及受环境干扰明显等特性增加了点云分类难度。因此,面向室外场景彩色激光点云的分类仍需进行深入地研究。

计算能力的提高及张量数据理论的发展推动了深度学习在场景理解方面的广泛应用,对于点云理解的深度网络被不断地提出,目前,基于深度学习的点云理解,主要面临三个方面的挑战:(1)点云数据是分布在空间中的 XYZ 点,属于非结构化的数据,由于没有结构化的网格而无法直接使用 CNN 滤波器。(2)点云本质上是三维的一长串点,在几何上,点的顺序不影响它在底层矩阵结构中的表示方式,相同的点云可以由两个完全不同的矩阵表

示。(3) 点云的数量不同于图像中像素, 像素的数量是一个给定的常数, 取决于相机。然而, 点云的数量是不确定的, 点云的多少取决于传感器及扫描的场景等因素。

目前, 针对点云的深度学习网络主要分为三大类, 即: 基于 2D 投影的深度学习网络、基于 3D 体素化的卷积神经网络和基于点云中单个点的网络模型。

(1) 基于 2D 投影的深度学习网络

受深度学习在 2D 图像上取得较好的效果的启发, 文献^[40]和文献^[41]将 3D 点云投影到 2D 图像后作为卷积神经网络的输入。常用的 2D 投影图像有: 基于虚拟相机的 RGB 图像, 基于虚拟相机的深度图, 基于传感器获取的距离图像, 全景图像等映射图像。这些投影的方法通过利用 2D 图像中已经经过大量图像训练好的目标检测^[41]与语义分割的网络模型作为预训练的模型进行微调, 可以在 2D 图像上获取较好的检测与分类效果。常用的预训练模型主要有: VGG-16、Alexnet、Googlenet、FCN、U-NET 等。

MVCNN^[40], Snapnet^[42], DeePr3SS^[43]等方法采用多视角投影的方法进行点云的分类。目前, Snapnet 和 DeePr3SS 的分类效果在 Semantic 3D 数据集的 reduced-8 上的 Benchmark 排名分别为第 7 和第 9^[44]。由此, 可知, 这类方法还存在很多改进的空间。此外, 这类方法会容易造成三维结构信息的丢失, 而且投影角度的选取, 同一角度的投影对物体的表征能力也不同, 对网络的泛化能力有一定的影响。

(2) 基于 3D 体素化的卷积神经网络

为了保留更多的三维结构信息, 研究人员在对 2D CNN 模型的基础上进行改进, 通过对点云进行体素化等预处理操作, 提出了基于 3D CNN 模型。最早的是由 Daniel 等^[45]基于监督三维卷积神经网络和点云体素化提出了 VoxNet 网络模型; Gernot 团队^[46]提出 OctNet 网络模型, 通过非平衡八叉树和三维卷积神经网络, 对点云进行了高分辨率体素表征, 并且将表征后的结果在目标分类、点云标记等工作中进行了验证。这类方法保留了大量的三维结构信息, 但如文献^[47]提出的基于 3D CNN 的大场景标注方法虽取得了不错的效果, 但只考虑了单一的体素尺度, 后来, 为了提高体素的表征能力, 相继的提出了多种多尺度体素的卷积神经网络方法, 如 Semantic 3D Net^[8], MS3_DVS^[48], MVSNet^[49]。

体素化的是这类方法的基础, 目前体素化的方法通常有: 基于 0-1 表示是否有点的体素方法, 基于体素网络密度的方法, 基于网格点的方法^[45]。现有的体素化的尺寸主要有: $11 \times 11 \times 11$ ^[49], $16 \times 16 \times 16$ ^[8], $20 \times 20 \times 20$ ^[47], $32 \times 32 \times 32$ ^[48]此外, 由于点云的数据量大, 有研究人员采用降采样体素化的方法, 降低数据量^[47], 也研究者通过体素化的方式将点云结构化, 然后利用 3D CNN、3D FCN 等方法进行语义分割^[50]。

基于 3D 体素化的卷积神经网络方法在 Semantic 3D 数据集上的 Benchmark 排名相对出色。此外, 这类方法通过网格化提供了点云结构, 网格的转换解决了排列问题, 同时也得到了数量不变的体素。然而 3D CNN 卷积时计算量非常大, 通常会降低体素的分辨率, 增加了量化的误差。此外, 网络中仅利用了点云的结构信息, 没有考虑到点云的颜色、强度等信

息。

(3) 基于点云中单个点的网络模型

为了充分利用点云的多模态信息,减少预处理过程中的计算复杂度,基于点云中单个点的网络模型逐渐被提出。如斯坦福大学 Charles 团队^[51]针对室内的点云场景提出了 PointNet,对室内点云数据进行了分类、部分分割、语义分割三部分工作,而后,针对 PointNet 中局部特征信息进行了改进,推向多尺度,综合局部特征,提出了 PointNet++网络模型^[52]; Yangyan Li^[53]等人在 PointNet 基础上进行改进,提出 PointCNN,该方法通过对点云特性的分析,提出了一种基于点云中点学习到的 X 变换方法,然后将其用于同时加权与点关联的输入特征和将它们重新排列成潜在隐含的规范顺序,之后再在元素上应用求积和求和运算。虽然所提算法能够对点云卷积处理的性能有所提高,但 X-transformations 还有较大的改进空间,尤其是在排序方面。

此外,最新的 SO-net^[54]的提出以及对 pointnet 改进提出的 PointFlowNet^[55]等网络模型均取得了较好的效果,但是对于大规模点云还具有一定的局限性,但这些网络为后续的大规模点云的语义分割提供了重要基础。

鉴于基于点的网络模型在点云分类上取得的较好效果,越来越多的人对这类方法进行改进。由于原始的 Pointnet 对于大规模点云具有局限性,有研究者通过对大规模点云进行预先的分割或者分块处理,然后利用 pointnet 网络进行分类,如 SPGraph^[9],该方法在目前的 Semantic 3D 数据集的 Semantic-8 和 reduced-8 上的 Benchmark 均排名第一。

以上三类方法各有各的优缺点,此外,目前关于点云场景语义理解的方法大多是针对单体目标的识别、室内或者室外小场景的语义分割,针对于室外大场景稠密激光点云的场景理解网络模型还有很多问题亟待解决。因此,对于室外复杂大场景的点云理解还有待进一步的研究。

4 研究展望

面向真实场景的多通道高效表征和场景理解具有重要价值和现实意义。三维场景理解是真实三维地图、L5 级自动驾驶以及机器智能等研究的关键。基于三维点云的语义分割与场景理解等智能感知方面因其数据的丰富具有明显的优势,但是目前基于三维点云的场景理解效果还有众多内容有待去深入研究,尤其是针对彩色激光点云的场景理解。本文通过对现有三维点云数据获取及场景理解的关键技术进行分析总结,归纳出了以下几点仍需重点研究。

(1) 针对基于学习方法需要大量有标注的样本问题,海量三维数据,尤其是彩色激光点云数据人工标注效率低,需要研究快速的高精度的自动标注方法;此外,对于内容丰富的有标签的室外点云场景理解数据集相对较少,构建一套作为 benchmark 的数据库也十分重要。

(2) 目前多数三维点云场景理解网络模型需要对点云进行大量预处理,如体素化等,增加了算法的复杂度。并且很多方法将三维信息投影到二维,再利用深度学习网络模型进行场景理解,丢失了部分三维信息。为了提高场景理解的效率,充分利用点云的三维信息,基于深度学习的端到端场景理解模型,即构建出由点到点标签的鲁棒性的网络模型是重要的研究方向。

(3) 由于点云数据样本中各类目标的点数并不均衡,且室外大场景的点云较为复杂、差异大,例如街景场景中车辆、行人、建筑物等目标的个数不同,且各类目标中的点云的点数相差较大,从而导致在深度学习网络模型训练时存在样本不均衡问题,容易导致网络模型的识别结果偏向于样本较大的类别。因此,解决训练样本不均衡问题的方法是重要的研究内容。

(4) 目前获取的三维点云场景数据的精度和点云与颜色配准的效果还有待进一步提高。近年来,随着传感器的发展,三维点云数据获取装置的研究更注重其高精度、信息丰富性、高通过性、获取数据规模的广阔性以及装置的便携性等。

(5) 目前基于三维点云数据的场景理解大多是对点云数据进行目标检测与语义分割,获取场景中的目标位置、种类和语义信息等内容,感知场景中感兴趣的内容,对整个场景的理解还不够充分。除了端到端更加完善的目标检测、语义分割和场景分类的深度网络模型是关键的技术重点外,对场景中各个目标的拓扑关系、场景中目标的行为与分析等内容的理解也是场景理解需要突破的关键技术与发展方向。

参考文献

- [1] Kim B S, Kohli P, Savarese S. 3D scene understanding by voxel-CRF[C]//Proceedings of IEEE International Conference on Computer Vision. IEEE Computer Society, 2013:1425-1432.
- [2] Zhuang Y, He G, Hu H, *et al.* A novel outdoor scene-understanding framework for unmanned ground vehicles with 3D laser scanners[J]. Transactions of the Institute of Measurement & Control, 2015,37(4):435-445.
- [3] 佟国峰, 杜宪策, 李勇, 等. 基于切片采样和质心距直方图特征的室外大场景三维点云分类[J].中国激光,2018,45(10):1004001.
Tong G F, Du X C, Li Y, *et al.* Three-dimensional point cloud classification of large outdoor scenes based on vertical slice sampling and centroid distance histograms[J]. Chinese Journal of Lasers, 2018,45(10):1004001.
- [4] 闫飞, 庄严, 王伟. 移动机器人基于多传感器信息融合的室外场景理解[J]. 控制理论与应用, 2011, 28(8):1093-1098.
Yan F, Zhuang Y, Wang W. Outdoor scene comprehension of mobile robot based on multi-sensor information fusion[J]. Control Theory & Applications, 2011, 28(8):1093-1098.

- [5] Izadi S, Kim D, Hilliges O, *et al.* Kinect Fusion: real-time 3D reconstruction and interaction using moving depth camera[C]// Proceedings of the 24th annual ACM symposium on User interface software and technology, ACM, 2011 :559—568.
- [6] 卢秀山,俞家勇,田茂义,等. 车载激光点云与序列化全景影像融合方法[J]. 中国激光, 2018, 45(5): 0510004
 Lu X S,Yu J Y,Tian M Y, *et. al.* Fusion method of vehicle laser point cloud and serialized panoramic image[J]. Chinese Journal of Lasers, 2018, 45(5): 0510004
- [7] Bircher A., Alexis K, Burri M., *et. al.* Structural inspection path planning via iterative viewpoint resampling with application to aerial robotics[C]// Proceedings of IEEE International Conference on Robotics & Automation, May 26-30, 2015 , Seattle, Washington, USA.
- [8] Hackel T, Savinov N, Ladicky L, *et al.* Semantic3D.net: A new large-scale point cloud classification benchmark[C]// Proceedings of ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences, 2017, IV-1-W1, 91–98.
- [9] Landrieu L, Simonovsky M. large-scale point cloud semantic segmentation with superpoint graphs[C] //Proceedings of IEEE CVPR 2018.
- [10] Silberman N, Hoiem D, Kohli P, *et al.* Indoor segmentation and support inference from RGBD images[C]//Proceedings of European Conference on Computer Vision. Springer, Berlin, Heidelberg, 2012:746-760.
- [11] Song S, Lichtenberg S P, Xiao J. SUN RGB-D: A RGB-D scene understanding benchmark suite[C]//Proceedings of IEEE Computer Vision and Pattern Recognition, 2015:567-576.
- [12] Liu M, Colas F, Siegwart R. Challenging data sets for point cloud registration algorithms[J]. International Journal of Robotics Research, 2012, 31(14):1705-1711.
- [13] <http://www.acfr.usyd.edu.au/papers/SydneyUrbanObjectsDataset.shtml>
- [14] Wu Z, Song S, Khosla A, *et al.* 3D ShapeNets: A deep representation for volumetric shapes[C]//Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, 2014:1912-1920.
- [15] Johnson A E. Spin-Images: A representation for 3-D surface matching[D]. Carnegie Mellon University, 1997.
- [16] Frome A, Huber D, Kolluri R, *et al.* Recognizing objects in range data using regional point descriptors[J]. Lecture Notes in Computer Science, 2004, 3023:224--237.
- [17] Rusu R B, Blodow N, Marton Z C, *et al.* Aligning point cloud views using persistent feature histograms[C]//Proceedings of Ieee/rsj International Conference on Intelligent Robots and Systems, 2008:3384-3391.
- [18] Salti S, Tombari F, Stefano L D. SHOT: Unique signatures of histograms for surface and texture description [J]. Computer Vision & Image Understanding, 2014, 125(8):251-264.

- [19] Wohlking W, Vincze M. Ensemble of shape functions for 3D object classification [C]//Proceedings of IEEE International Conference on Robotics and Biomimetics, 2012: 2987-2992.
- [20] Rusu R B, Bradski G, Thibaux R, et al. Fast 3D recognition and pose using the Viewpoint Feature Histogram[C]//Proceedings of Ieee/rsj International Conference on Intelligent Robots and Systems, 2014:2155-2162.
- [21] Rusu R B, Blodow N, Beetz M. Fast point feature histograms (FPFH) for 3D registration[C]//Proceedings of IEEE International Conference on Robotics and Automation, 2009:3212-3217.
- [22] A.Sappa, M.Devy, Fast range image segmentation by an edge detection strategy[M].In 3D Digital Imaging and Modeling, Los Alamitos, CA, USA, 2001:292-299.
- [23] 柯映林, 单东日. 基于边特征的点云数据区域分割[J]. 浙江大学学报(工学版), 2005, 39(3):377-380.
Ke Y L, Shan D R. Edge-based segmentation of point cloud data[J]. Journal of Zhejiang University, 2005, 39(3):377-380.
- [24] Papon.J, Abramov.A, Schoeler.M, Worgotter.F. Voxel cloud connectivity segmentation-supervoxels for point clouds[C]//Proceedings of Computer Vision and Pattern Recognition, 10.1109/CVPR.2013:264.
- [25] Vo A V, Truong-Hong L, Laefer D F, *et al.* Octree-based region growing for point cloud segmentation[J]. Isprs Journal of Photogrammetry & Remote Sensing, 2015, 104:88-100.
- [26] 杨泽鑫,程效军,李 泉,等.平面舱壁类型的船舱点云分割方法[J]. 中国激光, 2017, 44(10): 1010006.
Yang Z X, Cheng X J, Li Q, *et al.* Segmentation of point cloud in tank of plane bulkhead type[J]. Chinese Journal of Lasers, 2017, 44(10): 1010006.
- [27] Wang Y, Shi H. A segmentation method for point cloud based on local sample and statistic inference[J]. Communications in Computer and Information Science, 2015 (482) :274-282.
- [28] Green W R, Grobler H. Normal distribution transform graph-based point cloud segmentation[C]// Proceedings of Pattern Recognition Association of South Africa and Robotics and Mechatronics International Conference, 2015:54-59.
- [29] Yang J, Gan Z, Li K, *et al.* Graph-based segmentation for RGB-D data using 3-D geometry enhanced superpixels[J]. IEEE Transactions on Cybernetics, 2015, 45 (5) :913-926.
- [30] 程效军,程小龙,胡敏捷,等. 融合航空影像和 LIDAR 点云的建筑物探测及轮廓提取[J]. 中国激光, 2016, 43(5): 0514002
Cheng X J, Cheng X L, Hu M J, *et al.* Buildings detection and contour extraction by the fusion of aerial images and LIDAR point cloud[J]. Chinese Journal of Lasers, 2016, 43(5): 0514002.
- [31] Wolf, J Prankl, M Vincze. Fast semantic segmentation of 3D point clouds using a dense CRF with learned parameters[C]//Proceedings of 2015 IEEE International Conference on Robotics and Automation, 2015.
- [32] Hackel T, Wegner J D, Schindler K. Fast semantic segmentation of 3d point clouds with strongly varying density[J]. Isprs Annals of Photogrammetry Remote Sensing & Spatial Informa, 2016, III-3:177-184.

- [33] Weinmann M, Jutzi B, Hinz S, *et al.* Semantic point cloud interpretation based on optimal neighborhoods, relevant features and efficient classifiers[J]. *Isprs Journal of Photogrammetry & Remote Sensing*, 2015, 105:286-304.
- [34] Xiang B, Yao J, Lu X, *et al.* Segmentation-based classification for 3D urban point clouds[C]//*Proceedings of IEEE International Conference on Information and Automation*, 2017:172-177.
- [35] Aijazi A K, Checchin P, Trassoudaine L. Segmentation based classification of 3D urban point clouds: A super-voxel based approach with evaluation[J]. *Remote Sensing*, 2013, 5(4):1624-1650.
- [36] Ghamisi, P., Höfle, B. LiDAR data classification using extinction profiles and a composite kernel support vector machine[J]. *IEEE Geosci. Remote Sens. Lett.* 2017, 14, 659–663.
- [37] Lodha, S.K, Fitzpatrick, D.M, Helmbold, D.P. Aerial lidar data classification using AdaBoost[C] //*Proceedings of the 6th International Conference on 3-D Digital Imaging and Modeling*, Montreal, QC, Canada, 21–23 August 2007; pp. 435–442.
- [38] Ni, H.; Lin, X.; Zhang, J. Classification of ALS point cloud with improved point cloud segmentation and random forests[J]. *Remote Sens.* 2017, 9, 288.
- [39] Niemeyer, J.; Rottensteiner, F.; Soergel, U. Classification of urban LiDAR data using conditional random field and random forests[C]// *Proceedings of the Urban Remote Sensing Event*, Sao Paulo, Brazil, 21–23 April 2013; Volume 856, pp. 139–142.
- [40] Su H, Maji S, Kalogerakis E, *et. al.* L. Multi-view convolutional neural networks for 3d shape recognition[C]//*Proceedings of ICCV*, 2015.
- [41] Qi C. R, Su H, Nießner M., *et al.* Volumetric and multi-view cnns for object classification on 3d data[C] //*Proceedings of Computer Vision and Pattern Recognition(CVPR)*, IEEE, 2016.
- [42] Boulch A, Guerry J, Saux B L, *et al.* SnapNet: 3D point cloud semantic labeling with 2D deep segmentation networks[J]. *Computers & Graphics*, 2017, 71.
- [43] Felix J L, Martin D, Patrik T, *et al.* Deep projective 3D semantic segmentation[J], *Computer Analysis of Images and Patterns*, *Lecture Notes in Computer Science*, 2017, 10424, 95-107.
- [44] http://www.semantic3d.net/view_results.php?chl=2.
- [45] Daniel M and Sebastian S. VoxNet: A 3D convolutional neural network for real-time object recognition[C] //*Proceedings of 2015 IEEE/RSJ International Conference on Intelligent Robots and Systems*, Congress Center Hamburg, Sept 28 - Oct 2, 2015. Hamburg, Germany, 2015, 922-928.
- [46] Riegler G, Ulusoy A O, and Geiger A. Octnet: Learning deep 3d representations at high resolutions[J]. *arXiv preprint arXiv:1611.05009*, 2016.
- [47] Huang J, You S Y. Point cloud labeling using 3D Convolutional Neural Network[C]//*Proceedings of ICPR*, IEEE, Dec. 4-8, 2016, Cancun, Mexico.
- [48] Xavier R, Deschaud J E, François G. Classification of point cloud scenes with multiscale voxel deep

- network[J], arXiv:1804.03583, 2018.
- [49] Wang L, Huang Y, Shan J, et al. MSNet: multi-scale convolutional network for point cloud classification[J]. Remote Sensing, 2018, 10(4):612.
- [50] Tchapmi L. P , Choy C. B., Armeni I., *et al.* SEGCloud: Semantic segmentation of 3D point clouds[C] //Proceedings of International Conference on 3D Vision (3DV), 2017.
- [51] C. R. Qi, H. Su, K. Mo, and L. J. Guibas. Pointnet: Deep learning on point sets for 3d classification and segmentation[J]. arXiv preprint arXiv:1612.00593, 2016.
- [52] Qi C R, Yi L, Su H, *et al.* PointNet++: Deep Hierarchical Feature Learning on Point Sets in a Metric Space[J]. arXiv preprint arXiv:1706.02413, 2017.
- [53] Li Y Y, Bu R, Sun M C, *et al.* PointCNN[J]. arXiv:1801.07791,2018.
- [54] J. Li, B. Chen, and G. Lee.SO-Net: Self-Organizing Network for Point Cloud Analysis[C]//Proceedings of CVPR, arxiv:1803.04249, 2018.
- [55] Behl, Aseem; Paschalidou, Despoina; Donné, Simon; Geiger, Andreas.PointFlowNet: Learning Representations for 3D Scene Flow Estimation from Point Clouds[J].2018, arXiv180602170B.

网络首发:

标题: 三维点云场景数据获取及其场景理解关键技术综述

作者: 李勇, 佟国峰, 杨景超, 张立强, 彭浩, 高华帅

收稿日期: 2018-08-06

录用日期: 2018-09-06

DOI: 10.3788/lop56.040002

引用格式:

李勇, 佟国峰, 杨景超, 张立强, 彭浩, 高华帅. 三维点云场景数据获取及其场景理解关键技术综述[J]. 激光与光电子学进展, 2019, 56(04): 040002.

网络首发文章内容与正式出版的有细微差别, 请以正式出版文件为准!

您感兴趣的其他相关论文:

基于卷积神经网络特征和改进超像素匹配的图像语义分割

郭呈呈 于凤芹 陈莹

江南大学物联网工程学院, 江苏 无锡 214122

激光与光电子学进展, 2018, 55(8): 081005

基于SHOT的多尺度关键点检测技术研究

贾勇杰 熊风光 韩燮 况立群

中北大学计算机科学与技术学院, 山西 太原 030051

激光与光电子学进展, 2018, 55(7): 071013

深度学习在视觉定位与三维结构恢复中的研究进展

鲍振强 李艾华 崔智高 袁梦

火箭军工程大学, 陕西 西安 710025

激光与光电子学进展, 2018, 55(5): 050007

光波导激光陶瓷的研究进展与展望

李江 姜楠 葛琳 赵玉

中国科学院上海硅酸盐研究所透明光功能无机材料重点实验室, 上海 200050

激光与光电子学进展, 2018, 55(3): 030001

激光熔覆制备纳米结构涂层的研究进展

姚建华 张伟

浙江工业大学激光加工技术工程研究中心, 杭州 310014

激光与光电子学进展, 2006, 43(4):