Artificial Intelligence Review (2024) 57:83
https://doi.org/10.1007/s10462-024-10713-6

3D point cloud-based place recognition: a survey

Kan Luo[1,2] · Hongshan Yu[2] · Xieyuanli Chen[3] · Zhengeng Yang[2,4] · Jingwen Wang[2] · Panfei Cheng[2] · Ajmal Mian[5]

Accepted: 23 January 2024 / Published online: 7 March 2024

Abstract

Place recognition is a fundamental topic in computer vision and robotics. It plays a crucial role in simultaneous localization and mapping (SLAM) systems to retrieve scenes from maps and identify previously visited places to correct cumulative errors. Place recognition has long been performed with images, and multiple survey papers exist that analyze image-based methods. Recently, 3D point cloud-based place recognition (3D-PCPR) has become popular due to the widespread use of LiDAR scanners in autonomous driving research. However, there is a lack of survey paper that discusses 3D-PCPR methods. To bridge the gap, we present a comprehensive survey of recent progress in 3D-PCPR. Our survey covers over 180 related works, discussing their strengths and weaknesses, and identifying open problems within this domain. We categorize mainstream approaches into feature-based, projection-based, segment-based, and multimodal-based methods and present an overview of typical datasets, evaluation metrics, performance comparisons, and applications in this field. Finally, we highlight some promising research directions for future exploration in this domain.

Keywords 3D point cloud · Place recognition · LiDAR · Localization · Mapping

1 Introduction

Where am I? Determining the place in a given reference database or map is still an ongoing challenge in computer vision, robotics, and autonomous driving (Masone and Caputo 2021). Place recognition is a perception-based method that can use images, 3D point clouds, and other information acquired by robots to identify previously visited places by comparing the similarity between query frame information and map database information. Place recognition can help robots improve the accuracy of loop-closure detection by providing reference places in the environment, that is, using the initial reference places provided by place recognition to lock the loop-closure area. This eliminates cumulative errors and helps to achieve high-precision and reliable simultaneous localization and mapping (SLAM). This critical task of place recognition has obtained significant research attention over the last few decades. Since GPS-based methods may not always be accurate and sometimes even completely fail in cities with high-rise buildings and bridges, numerous

Extended author information available on the last page of the article research efforts are dedicated to developing solutions for image-based and 3D point cloud-based place recognition.

Image-based place recognition, also known as visual place recognition (VPR), involves providing an image of a place and recognizing whether the image corresponds to a previously visited place (Lowry et al. 2015). Since the camera is the most commonly used sensor for this purpose, conventional image feature extraction algorithms such as SIFT (Lowe 2004), SURF (Bay et al. 2006), BRIEF (Calonder et al. 2010) and ORB (Rublee et al. 2011) have been conveniently applied to VPR. Consequently, VPR has received extensive attention from researchers, and numerous advancements have been made in the past two decades (Lowry et al. 2015; Zhang et al. 2021; Masone and Caputo 2021; Barros et al. 2021; Yin et al. 2022). Lowry (Lowry et al. 2015) defined the "place" concept in their survey and discussed how VPR solutions can implicitly or explicitly account for changes in the environment's appearance.

With the advent of deep learning-based image classification methods, recent surveys (Zhang et al. 2021; Masone and Caputo 2021) have focused on their application in VPR. VPR methods can be classified into two categories, depending on the camera sensor modalities used: partial-observable camera and fully-observable camera (Yin et al. 2022). The partial-observable camera includes pin-hole, fish-eye, and stereo cameras. Most VPR methods and datasets are developed based on this modality (Zaffar et al. 2021). However, observing the same area under different perspectives is still a significant challenge in partial-observable camera-based VPR, which may result in significantly different observations of the same place. This problem is overcome by fully-observable camera systems which provide a 360-degree field of view (Scaramuzza 2014), and have the inherent advantage of viewpoint-invariant localization.

Although VPR has achieved great success, its performance is inevitably influenced by various environmental factors (Lai et al. 2022), such as lighting conditions, viewpoint variations, seasonal changes, weather conditions, etc. In contrast to image-based methods, 3D point cloud-based place recognition (3D-PCPR) methods utilize range sensors, such as LiDAR or RGB-D sensors, to acquire 3D geometric information about the surrounding environment. The obtained 3D information is then used to identify whether the place in the environment has been visited before. The use of 3D point clouds makes 3D-PCPR more robust to changes in lighting, viewpoint, seasons, and weather conditions (Uy and Lee 2018), enabling SLAM technology to adapt to these challenging scenarios. Driven by the recent advancements in point cloud sensor technology, there has been a surge of interest among researchers in exploring and advancing 3D-PCPR techniques. This has resulted in remarkable advancements in 3D-PCPR (Yin et al. 2018; Uy and Lee 2018; Liu et al. 2019; Du et al. 2020; Zhou et al. 2021; Komorowski 2021; Sun et al. 2020; Fan et al. 2020; Xiang et al. 2021; Hou et al. 2022; He et al. 2016; Kim and Kim 2018; Kim et al. 2021; Yin et al. 2020, 2021; Jiang et al. 2020; Schaupp et al. 2019; Chen et al. 2021; Xu et al. 2021; Wang et al. 2020; Dubé et al. 2017; Dube et al. 2020; Vidanapathirana et al. 2021; Li et al. 2021; Zhu et al. 2020; Lu et al. 2020; Pan et al. 2021; Komorowski et al. 2021; Cramariuc et al. 2021; Yin et al. 2021; Chen et al. 2020a; Ma et al. 2023).

In the face of such rapid advancements in 3D-PCPR techniques, there is a pressing need for a comprehensive and up-to-date survey that encompasses the broader scope of 3D data sources beyond just LiDAR sensors. While existing surveys (Wang et al. 2019; Elhousni and Huang 2020; Yin et al. 2022, 2023) have made valuable contributions, they either focus on specific aspects of 3D-PCPR or provide limited coverage of the topic. For example, Wang et al. (2019) only provide a summary of loop-closure detection methods with 3D data sources, but their discussion is confined to a restricted number of methods.

Similarly, Elhousni (Elhousni and Huang 2020) provide a brief survey on 3D LiDAR-based localization methods, primarily centered around LiDAR-aided pose tracking for autonomous vehicles. Yin et al. (2022) conduct a general place recognition survey with a focus on real-world autonomy, offering limited coverage of 3D-PCPR methods. Even the recent survey (Yin et al. 2023) does not cover the topic comprehensively and is restricted to LiDAR-based global localization topics.

Considering the existing literature, it becomes evident that a comprehensive survey specifically dedicated to 3D-PCPR methods, encompassing a broader range of 3D data sources, while also giving insights into the limitations of existing methods and highlighting promising future directions to explore in this domain is lacking. This survey covers the gap and serves as an invaluable resource for researchers, enabling them to grasp the current state-of-the-art, identify research gaps, and drive further advancements in the rapidly evolving field of 3D-PCPR.

Our survey covers more than 180 important works related to place recognition. We mainly considered papers published in well-known journals or conferences in the fields of robotics, computer vision and artificial intelligence, such as IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI), International Journal of Computer Vision (IJCV), IEEE Transactions on Robotics (T-RO), IEEE International Conference on Intelligent Robots and Systems (IROS), IEEE International Conference on Robotics and Automation (ICRA), and IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Besides that, we also considered some latest preprint papers on arXiv that already gained much attention. Based on an extensive and comprehensive literature survey, we propose a novel categorization scheme that classifies 3D-PCPR methods into four distinct categories. By categorizing existing methods and examining each category in detail, our survey provides a deeper understanding of the current state of the field and identifies promising avenues for future research.

The rest of this article is organized as follows: Sect. 2 briefly introduces the background of 3D point clouds, including the acquisition, development, and applications of 3D point clouds, etc. Sections 3 to 6 respectively introduce the four categories of 3D-PCPR methods, namely feature-based, projection-based, segment-based, and multimodal-based methods, as shown in Fig. 1. Section 7 describes the main datasets, evaluation metrics and performance comparisons commonly used in 3D-PCPR. Section 8 introduces related applications of 3D-PCPR. Particularly, we discuss the future promising research directions of 3D-PCPR technology. Finally, we conclude our survey in Sect. 9.

2 Acquisition of 3D point cloud

A point cloud is a set of geometric points situated on the surfaces of 3D objects in Euclidean space. These points are typically captured using 3D sensors like LiDAR, laser scanners, structured light scanners, or Time-of-Flight (ToF) RGB-D cameras. A point $\mathbf{p}_i$ in a point cloud $\mathcal{P}$ can be represented by its x, y, and z Cartesian coordinates, denoted as $\mathcal{P} = \{\mathbf{p}_i = (x_i, y_i, z_i)\}_{i=1}^{N}$, where N is the number of points in $\mathcal{P}$. In this section, we provide a brief overview of the acquisition and applications of 3D point clouds.

The acquisition of a 3D point cloud involves measuring the depth or range of obstacles (from the sensor) and then calculating the 3D coordinates and attributes of points in Euclidean space (Xu and Stilla 2021). Various sensors are available to acquire 3D point clouds. These sensors are typically grouped as being either active or passive (Lillesand et al. 2015) as shown in Fig. 2.

Active sensors include structured light technology (e.g. Kinect-1 Zennaro 2014; Lun and Zhao 2015) whereby an infrared light pattern is projected onto the scene and then sensed by a camera to measure the 3D distance using the principle of triangulation. Another type of sensor is laser scanners (Wang et al. 2020; Vosselman and Maas 2010) that project a single light stripe and scan it over the scene to generate dense point clouds, using the triangulation principle. ToF principle is used by LiDARs (Wandinger 2005) that transmits and scans multiple laser beams to generate sparse point clouds of a large scene. Radar (Knott and Skolnik 2008) sensors also use the ToF principle but they transmit electromagnetic radiation and then measure the reflections returned from the target object. Another class of ToF sensors transmit a modulated single light (usually IR) and measure the phase difference of the reflected light from various points in the scene to measure the time and hence their distance from the sensor. Kinect-2 (Fankhauser et al. 2015; Wasenmüller and Stricker 2016) and PrimeSense (Breuer et al. 2014; Kuan et al. 2019) sensors use this technology. Only lasers based and radar sensors work in outdoor environments due to the strong sunlight. Structured light and ToF sensors are designed for indoor use only. These sensors measure the distances to obstacles to compute their 3D coordinates, resulting in a point cloud of xyz coordinates.

Passive sensors, such as photogrammetry (Lillesand et al. 2015) and stereo (Beltran and Basañez 2014) cameras, capture 3D data of the environments without actively emitting any energy. These sensors typically measure the geometry structure of environments using multiple observations, estimate the depth of objects within the scene through photogrammetric approaches such as multi-view geometry, and finally generate point clouds from the 3D reconstructions. In addition to 3D coordinates, a point cloud can also contain other information such as intensity and color, and normal vectors can be calculated using the local neighborhood geometry (Chen et al. 2020a, b). We show more details of different sensors in Table 1.

Early 2D laser scanners (Thrun 2002), also known as single-line LiDAR, have a single-line beam emitted by the laser source to generate low-resolution 2D planar scans. Due to its simple structure and high reliability, it has been widely studied and used in real-world robots (Hess et al. 2016; Kuang et al. 2023). However, 2D laser scanners can only perform plane scanning and generate low-resolution point cloud information, limiting its use for place recognition (Zhang and Ghosh 2000; Olson 2009; Zimmerman et al. 2023). The development of sensor technology has promoted point cloud sensing from 2D to 3D. Compared to 2D scans, 3D point clouds present more information for robots to better understand their surroundings. Therefore, 3D sensors have developed rapidly in the past three decades.

3D point cloud data finds applications in many fields (Guo et al. 2020), including computer vision, autonomous driving, robotics, remote sensing, medical treatment, cultural relic reconstruction, etc. The rest of this article will mainly discuss research and applications of 3D point clouds in place recognition.

3 Feature based methods

We show a chronological overview of the four main categories of 3D-PCPR methods in Fig. 3. As depicted, feature-based approaches are fundamental methods for 3D-PCPR. The main idea of these methods is to extract features from the 3D point clouds and then match such features to perform subsequent place recognition. We divide the feature-based

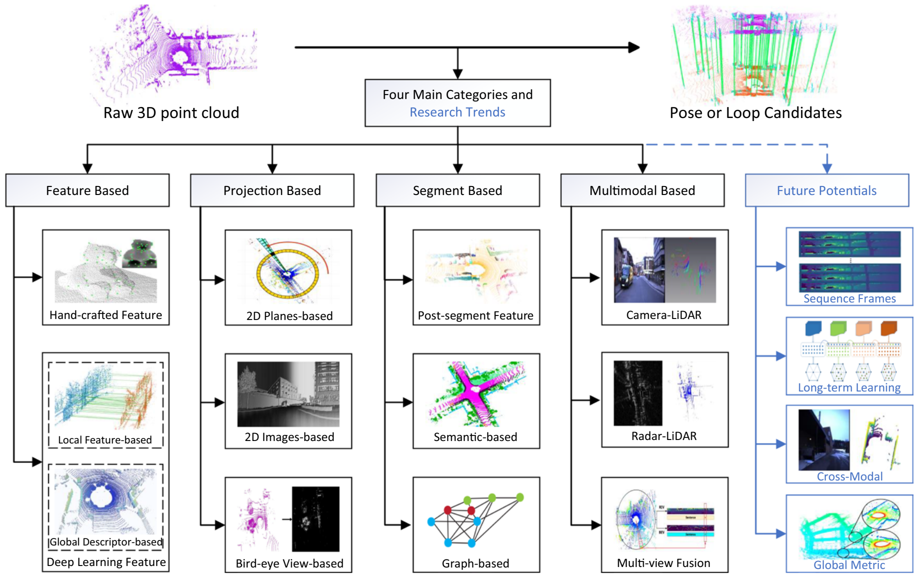

Fig. 1 Four categories of 3D-PCPR methods namely feature-based, projection-based, segment-based, and multimodal-based, as well as future research trends

methods into two categories: hand-crafted feature-based methods and deep learning feature-based methods.

3.1 Hand-crafted feature-based methods

Hand-crafted feature-based methods have been extensively researched for several decades, resulting in significant advancements in 3D-PCPR. These advancements have played a crucial role in driving the continuous development of this field. Magnusson et al. (2009a, 2009b) conducted early research inspired by NDT (Biber and Straßer 2003) and proposed a loop detection approach based on surface shape and orientation histograms using only 3D point cloud data. The main idea behind their method is to calculate the similarity of two scans from the histogram of the NDT descriptors and achieve good recall rates at low false negative (For more detailed description about recall and false negative, please refer to Sect. 7.2) in environments with different characteristics.

Steder et al. (2010) presented a robust approach for 3D place recognition using range data. Their method uses interest feature points and scores candidate transformations. Although this method produces accurate relative pose transformations between two scans and has high recognition rates, it cannot achieve real-time performance and orientation invariance. To overcome these shortcomings, they later proposed another method (Steder et al. 2011) using a combination of a bag-of-words and a point-feature-based estimation of relative poses, which is more efficient and rotational invariant compared to the former approach.

A loop closure detection method using small-sized signatures from 3D LiDAR data was presented by Muhammad and Lacroix (2011). This method extracts histogram-based signatures from 3D LiDAR data and uses them for loop closure detection. These features are

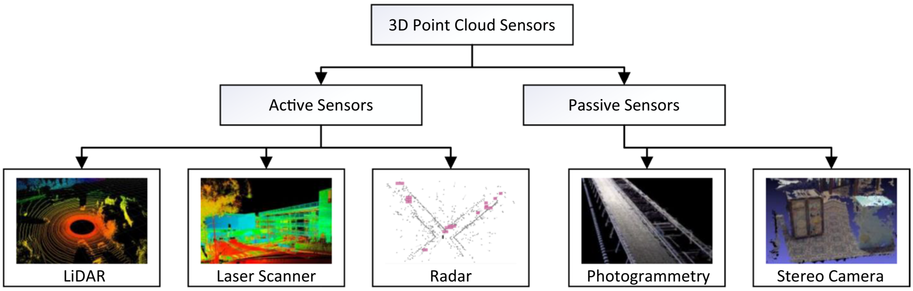

Fig. 2 Common 3D point cloud acquisition sensor types along with sample acquired point clouds

based on histograms of local surface normals for 3D point clouds. To design a 3D-PCPR method recognizing complex indoor scenes and tackling moving objects' disturbance effectively, Zhuang et al. (2012) proposed an approach that can extract and match the Speed Up Robust Features (SURF) (Bay et al. 2006) from bearing-angle images generated by a self-built rotating 3D laser scanner. Using both the local SURF features and the global spatial features, their place recognition framework has shown validity and robustness in dynamic indoor environments.

Collier et al. (2012) presented a LiDAR-based place recognition system that can extract highly descriptive features called the variable dimensional local shape descriptors from 3D point cloud data to encode environmental features. Their system can run on a military research vehicle equipped with a highly accurate, 360-degree field of view LiDAR and detect loops regardless of the sensor's orientation.

Bosse and Zlot (2013) presented a noteworthy method for place recognition in large 3D point cloud datasets, utilizing keypoint voting. This approach involves extracting keypoints directly from the 3D point cloud and describing them using handcrafted 3D gestalt descriptors. The keypoints then vote to their nearest points, and based on the resulting scores, loops are detected.

Röhling et al. (2015) proposed a fast histogram-based similarity measure for detecting loop closures in 3D point cloud data. Their method can avoid computationally expensive features and compute histograms from simple global statistics of the LiDAR scans. Hence, high precision and recall rates are achieved in a computationally efficient manner.

Another fast, complete, 3D point cloud-based loop closure for LiDAR odometry and mapping method was proposed by Lin and Zhang (2019). They compute 2D histograms of local map patches and then use the normalized cross-correlation of the 2D histograms as the similarity metric. This method selects some keyframes from the LiDAR input and the offline map, and can quickly evaluate the similarity between keyframes to form a relatively simple and practical system for place recognition. However, this method is mainly based on a small field of view and does not propose a very effective calculation method for relative pose estimation between the keyframes.

3.2 Deep learning feature-based methods

The rapid advancement of technological innovations along with the proliferation of big data and the exponential enhancement of computational capabilities, has significantly propelled the widespread adoption of deep learning techniques (LeCun et al. 2015) in a

Table 1 A brief comparison of commonly used 3D point cloud sensors

Sensor	Popular products	Type	Method	Range	Accuracy	Density	Texture	Speed	Cost	Environment
LiDAR (Wandinger 2005)	Velodyne, Pandar	Active	Laser	Long	High	Sparse	No	Medium	High	Indoor/Outdoor
Laser Scanner (Vosselman and Maas 2010)	FARO, Leica	Active	Laser	Long	High	Dense	No	Very Slow	Very High	Indoor/Outdoor
MMW Radar (Knott and Skolnik 2008)	Navtech, RoboSense	Active	Millimeter wave	Long	Low	Sparse	No	Medium	Medium	Indoor/Outdoor
RGB-D Sensor (Endres et al. 2013)	Kinect, RealSense	Mixed	Infrared	Short	Low	Medium	Yes	Fast	Low	Indoor
Photogrammetry (Lillesand et al. 2015)	Metashape, Pix4Dmapper	Passive	Image-based	Varies	Medium	Dense	Yes	Slow	high	Indoor/Outdoor
Stereo Camera (Beltran and Basañez 2014)	ZED, PointGrey	Passive	Image-based	Medium	Medium	Medium	Yes	Fast	Medium	Indoor/Outdoor

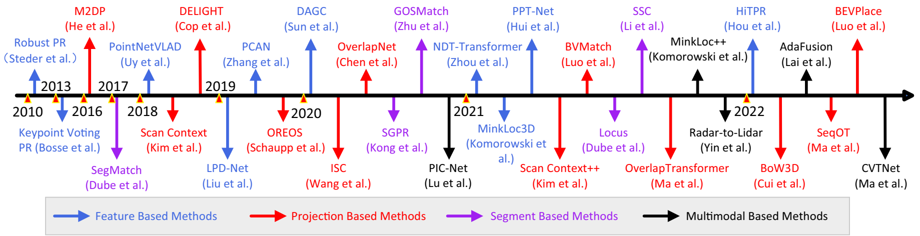

Fig. 3 Chronological overview of the four main categories of 3D-PCPR methods. Each category of representative methods is marked with different colors

myriad of domains. In this section, we discuss some deep learning feature-based methods in 3D-PCPR. These methods mainly learn from the raw 3D point clouds to extract features that are useful for performing subsequent place recognition.

In the absence of any prior knowledge, the task of 3D point cloud-based global localization poses a formidable challenge. To tackle this issue, Yin et al. (2018) proposed a semi-handcrafted approach that leverages siamese LocNets for representation learning from LiDAR point clouds. By employing LocNet representations in the Euclidean space, a crucial global prior map can be constructed, which helps in enhancing robustness. Nonetheless, achieving global localization in dynamic environments remains a daunting task even with their method. The inherent disorderliness of point clouds complicates the extraction of local features, making the encoding of these features into global descriptors for addressing the 3D-PCPR problem more challenging.

Uy and Lee (2018) proposed a deep learning network called PointNetVLAD (see Fig. 4) to extract local features using PointNet (Qi et al. 2017) which are then passed to NetVLAD (Arandjelovic et al. 2016) to generate the final discriminative global descriptor. This work presents the first end-to-end trainable network for extracting global descriptors directly from raw 3D point clouds. However, PointNetVLAD overlooks the spatial distribution of similar local features, which is of significant importance in capturing the static structural information in expansive dynamic environments. To address this limitation, Liu et al. (2019) presented LPD-Net, which can extract more discriminative and generalizable global descriptors by employing an adaptive local feature extraction module and a graph-based neighborhood aggregation module for extracting local structures and revealing the spatial distribution of local features within large-scale point clouds.

For implementing practical vehicle platforms that possess limited computing and storage resources, the author proposed SeqLPD (Liu et al. 2019), a lightweight variant derived from LPD-Net. SeqLPD aims to tackle the place recognition problem by integrating deep learning-based point cloud description and a coarse-to-fine sequence matching strategy, resulting in notable improvements in loop closure detection. Despite the success of LPD-Net, it is still resource-intensive. In an effort to enhance performance while mitigating resource demands, Hui et al. (2022) proposed EPC-Net, an efficient point cloud learning network specifically designed for 3D-PCPR tasks. EPC-Net achieves commendable performance while effectively reducing memory requirement and inference time.

Yin et al. (2018) proposed an end-to-end framework that utilizes low-dimensional feature matching instead of geometry matching for LiDAR-based long-term place recognition. This approach combines dynamic octree mapping and place feature inference modules and the feature learning is performed in a fully unsupervised manner. In the aggregation of local features into a global descriptor, it is important to reweigh the contributions of each local point feature, thereby allocating greater attention to regions that are more relevant to the task. Drawing inspiration from this concept, Zhang and Xiao (2019) proposed the Point Contextual Attention Network (PCAN), which leverages point context to predict the significance of individual local point features, offering an efficient means to encode the local features into a discriminative global descriptor.

Indoor place recognition represents an important yet relatively less explored area. The SpoxelNet (Chang et al. 2020) neural network architecture was proposed as a 3D-PCPR method tailored for crowded indoor spaces. SpoxelNet effectively encodes input voxels into global descriptor vectors. The method involves voxelizing point clouds in spherical coordinates and defining voxel occupancy through ternary values. SpoxelNet has been evaluated on diverse indoor datasets, yielding promising results for the task of place recognition.

Du et al. (2020) proposed DH3D, the first approach that unifies global place recognition and local 6DoF pose refinement. DH3D incorporates a deep hierarchical network and utilizes NetVLAD to generate more discriminative global descriptors. However, the obtained descriptors lack rotational invariance and often exhibit shortcomings in reverse revisits. Zhou et al. (2021) introduced NDT-Transformer, a real-time and large-scale 3D-PCPR method. Taking inspiration from the success of the NDT (Biber and Straßer 2003) and Transformer (Vaswani et al. 2017) models, NDT-Transformer condenses raw point clouds through 3D NDT representation and subsequently learns global descriptors through a novel NDT-Transformer network. Notably, this approach obviates the need for handcrafted features and can serve as a crucial module in real-time SLAM systems.

Acquiring high-quality point cloud data along with ground truth registration in real-world scenarios for training place recognition models is time-consuming and resource-intensive. Qiao et al. (2021) address this problem by proposing a novel registration-aided 3D domain adaptation network named VLPD-Net (Virtual Large-Scale Point Cloud Descriptor Network) for 3D-PCPR. Recognizing the importance of the neighborhood context of each point, the method takes into account the tactical contributions of different local features, which may vary unevenly. Xia et al. (2021) adopted a point orientation encoding module to capture neighborhood information from various orientations. Additionally, a self-attention unit is employed to encode the spatial relationships of local features for weighted aggregation. This end-to-end architecture enables one-stage training, generating a discriminative and compact global descriptor directly from a given 3D point cloud by exploring the relationships between raw 3D point clouds and the varying importance of local features to perform large-scale 3D-PCPR tasks.

Existing PointNet-like methods primarily process unordered point clouds and may not adequately capture local geometric structures. Consequently, a large-scale 3D-PCPR method named MinkLoc3D was introduced by Komorowski (2021). MinkLoc3D leverages a sparse voxelized point cloud representation and sparse 3D convolutions to compute a discriminative 3D point cloud descriptor, as depicted in Fig. 5. The efficacy of this method can be attributed to two key factors. Firstly, the sparse convolutional architecture effectively generates informative local features. Secondly, enhancements in the training process facilitate efficient and effective training by accommodating larger batch sizes. However, MinkLoc3D solely utilizes the geometry of 3D point clouds for place recognition. To address this limitation, the author proposed MinkLoc3D-SI (Żywanowski et al. 2021), which integrates both spherical representation and measurement intensities. MinkLoc3D-SI improves performance when a single 3D LiDAR scan is used. Experimental results demonstrate the superior performance of MinkLoc3D-SI on single scans from 3D LiDAR and its excellent generalization ability.

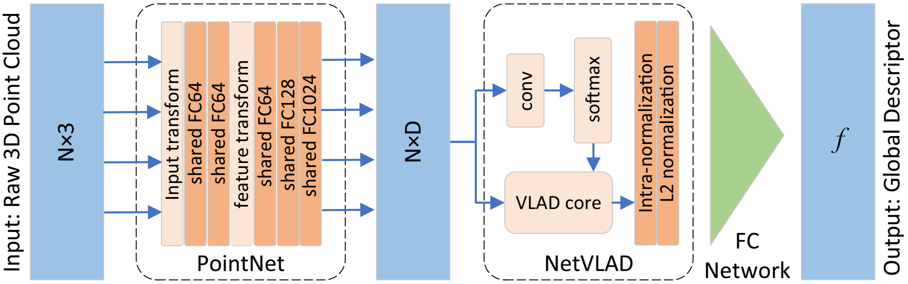

Fig. 4 A lightweight architecture of PointNetVLAD (from Uy and Lee (2018)). N denotes the number of input 3D points, D denotes the dimension of the learned point features, and f denotes the final global descriptor vector

Komorowski (2022) has recently proposed an improved 3D-PCPR method that incorporates a ranking-based loss and large batch training technique. This method employs a simple and efficient 3D convolutional feature extraction process to enhance channel attention blocks. The network architecture is an advancement over the MinkLoc3D Komorowski (2021) point cloud descriptor and surpasses the performance of most recent methods with more complex architectures.

However, many existing algorithms tend to overlook long-range contextual properties and exhibit large model sizes, thereby limiting their widespread applicability. To overcome these challenges, Fan et al. (2022) introduced SVT-Net, which is a lightweight sparse voxel transformer designed for large-scale 3D-PCPR tasks. To mitigate issues related to moving objects, size disparities among objects, and long-range contextual information, Xu et al. (2021) proposed TransLoc3D, another large-scale 3D-PCPR method that employs adaptive receptive fields. TransLoc3D achieves impressive results across multiple datasets, including Oxford RobotCar (Maddern et al. 2017), USRABD dataset(including a university sector (U.S.), a residential area (R.A.), and a business district (B.D.)) (Uy and Lee 2018).

Sun et al. (2020) proposed DAGC that leverages dual attention and graph convolution techniques to perform 3D-PCPR. The dual attention and residual graph convolution network modules contribute to the extraction of discriminative and generalizable features for describing a point cloud. By simultaneously considering the importance of points and features, DAGC utilizes the point relationships to extract local features, which are subsequently passed through a feature fusion block to generate global descriptors by a NetVLAD (Arandjelovic et al. 2016) module. Whereas DAGC effectively captures the relationship between points and the discriminative power of different features in generating global descriptors, it does not account for the spatial relationships between local features nor the long-range dependence of different features.

To take full advantage of the contextual semantic features of the scene and mitigate the influence of dynamic noise, such as moving cars and pedestrians, Fan et al. (2020) proposed SRNet, a 3D scene recognition network using static graphs and dense semantic fusion. SRNet comprises Static Graph Convolution, a Spatial Attention Module, and Dense Semantic Fusion. These modules help the network learn a deep understanding of the contextual scene semantics. After obtaining naive embedded features, the final global descriptors used for recognition are aggregated by an additional NetVLAD module. Benefiting from strong local feature learning, contextual semantics understanding,

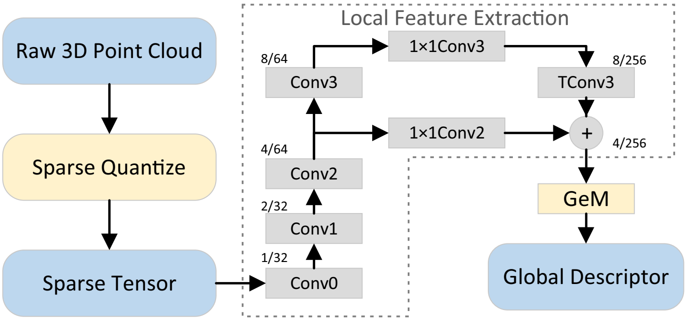

Fig. 5 An overview of MinkLoc3D architecture (from Komorowski 2021). Raw 3D point cloud is quantized into a sparse, single-channel 3D tensor. Based on the extracted local feature, a global point cloud descriptor is generated using Generalized-mean (GeM) pooling

and dynamic noise avoidance capabilities, combined with network flexibility, SRNet can be easily integrated into other point cloud architectures for tasks beyond place recognition.

Most existing algorithms struggle when dealing with reverse loops. Cattaneo et al. (2022) proposed LCDNet for simultaneous deep loop closure detection and point cloud registration in LiDAR-based SLAM. LCDNet leverages a shared encoder and two heads for generating global descriptors and estimating the relative pose, enabling simultaneous identification of previously visited places and estimation of the 6-DoF relative transformation between the current scan and the map. Considering the sparsity of point clouds, Hui et al. (2021) proposed PPT-Net, a pyramid point cloud transformer network designed for large-scale place recognition. PPT-Net extracts discriminative local features to form a global descriptor. It incorporates a pyramid point transformer module, which adaptively learns the spatial relationships among different KNN neighboring points, and a pyramid VLAD module, which aggregates the multi-scale feature maps of point clouds into comprehensive global descriptors.

Habich et al. (2021) proposed an extension of graph-based SLAM to exploit the potential of 3D laser scans for loop detection. Their method extracts global features from the point cloud and then uses a trained detector to determine the presence of a loop. Their algorithm is considered an extension of the widely used state-of-the-art RTAB-Map (Labbé and Michaud 2019) library. In the domain of indoor LiDAR mobile mapping, Xiang et al. (2021) proposed FastLCD, a compact and efficient method for precise loop closure detection using comprehensive descriptors such as statistics, geometry, planes, range histogram, and intensity histogram. These features are invariant to rotation and are encoded to uniquely describe each point cloud scan, making FastLCD a feasible and reliable loop closure detection algorithm.

Hou et al. (2022) introduced a novel Hierarchical Transformer for Place Recognition (HiTPR), specifically designed to address the challenges of LiDAR-based large-scale place recognition such as robustness and real-time performance. HiTPR avoids the use of the memory-intensive and inefficient approach of global information aggregation through NetVLAD (Arandjelovic et al. 2016). HiTPR comprises four main components namely, point cell generation, short-range transformer, long-range transformer, and global descriptor aggregation, enabling HiTPR to achieve superior performance in terms of average recall rate.

Many existing methods fail to produce consistent descriptors for the same scene under different viewpoints, making rotation invariance crucial. Therefore, Li et al. (2022) designed an efficient 3D LiDAR-based place recognition using a rotation invariant neural network that exploits the fact that autonomous robots generally rotate only in the yaw direction. Their method combines semantic and geometric features to improve descriptiveness and employs a rotation-invariant siamese neural network to predict the similarity of descriptor pairs. To acquire more repeatable global descriptors and improve performance in 3D place recognition, Vidanapathirana et al. (2022) presented an end-to-end trainable locally guided global descriptor learning network (LoGG3D-Net) for 3D-PCPR. To tackle both tasks of loop closing and relocalization, Shi et al. (2023) proposed a novel multi-head network namely LCR-Net. Based on the input 3D point clouds, the method utilizes a novel feature extraction and pose-aware attention mechanism to accurately estimate the similarities and 6-DoF poses between pairs of LiDAR scans.

Compression techniques have become popular to store large-scale point cloud maps (Golla and Klein 2015; Huang and Liu 2019; Wiesmann et al. 2021). To address the problem of place recognition in a compressed point cloud map, Wiesmann et al. (2022) presented Retriever, a novel deep neural network architecture that directly operates on compressed feature representation, then uses a NetVLAD (Arandjelovic et al. 2016) layer to aggregate local features with an attention mechanism between local features and a latent code.

3.3 Summary

In summary, feature-based methods are the most common methods in 3D-PCPR. Their main idea is to directly use hand-crafted or deep learning-based methods on 3D point clouds to extract local or global features, then similarity matching is performed on the extracted point cloud query features and 3D reference map features, and finally achieve the task of subsequent place recognition. The hand-crafted place recognition methods have good interpretability and high computational efficiency. Representative algorithms include AL3D (Magnusson et al. 2009a), Robust PR (Steder et al. 2010), Keypoint Voting PR (Bosse and Zlot 2013), etc. However, they are unable to extract all relevant features from 3D point clouds. On the other hand, deep learning features-based place recognition methods have gained more popularity as they can automatically learn and characterize relevant features from the raw 3D point clouds. Deep learning feature-based place recognition methods can be divided into local feature-based methods and global descriptor-based methods. Local feature-based methods, such as LPD-Net (Liu et al. 2019), PCAN (Zhang and Xiao 2019), PPT-Net (Hui et al. 2021), MinkLoc3D (Komorowski 2021), etc., first extract local features based on pointnet-like (Qi et al. 2017) methods, and then aggregate them into global descriptors through VLAD-like (Arandjelovic et al. 2016) methods for subsequent place recognition. Global descriptor-based methods generally use transformer-like (Vaswani et al. 2017) methods to directly represent global features from the raw 3D point clouds, such as NDT-Transformer (Zhou et al. 2021), LCDNet (Cattaneo et al. 2022), RINet (Li et al. 2022), etc. Overall, deep learning feature-based place recognition methods have achieved advanced performance, such as HiTPR (Hou et al. 2022), MinkLoc3Dv2 (Żywanowski et al. 2021), etc. However, the point cloud features they extract are not easy to interpret (Minh et al. 2022) and require a large amount of training data and powerful computing hardware (Han et al. 2023). Nonetheless, feature-based place recognition methods, especially end-to-end deep learning based 3D-PCPR methods, will remain the preferred research direction in the foreseeable future due to their state-of-the-art performance.

4 Projection based methods

Projection-based methods are another category of methods in 3D-PCPR where the main idea is to project the raw 3D point clouds to 2D planes, 2D images, or bird-eye view (BEV) information. These projections are then subsequently processed to achieve place recognition. In this section, we discuss the projection-based methods by dividing them into three categories: 2D planes based methods, 2D image-based methods, and bird-eye view-based methods.

4.1 2D Planes based

The 2D planes based 3D-PCPR methods involve an initial step of projecting the raw 3D point cloud onto a 2D representation which is then utilized for subsequent place recognition tasks. A pioneering approach in this domain is M2DP (He et al. 2016), which projects a 3D point cloud onto a sequence of 2D planes that capture various viewpoints of the cloud. By characterizing the point projections, M2DP extracts multiple density distributions or signatures from a single point cloud.

Scan Context (Kim and Kim 2018) is an egocentric spatial descriptor, which summarizes a place as a plane matrix for 3D-PCPR and offers robustness to structural changes such as dynamic objects and seasonal changes, as shown in Fig. 6. It projects the maximum height of the point cloud in different bins to generate a 2D global descriptor. However, using only the maximum height information does not offer much invariance in the lateral direction. It also uses brute-force search which is highly inefficient. Wang et al. (2020) proposed another method called Intensity Scan Context (ISC) which codes intensity information and geometry relations for loop closure detection. It explores the intensity information properties for place recognition and to reduce the computational cost, it performs a two-stage hierarchical re-identification process including a binary-operation-based fast geometric relation retrieval and an intensity structure re-identification.

Cai and Yin (2021) introduced a robust global descriptor, known as Weighted Scan Context (WSC), for 3D-PCPR by leveraging the enhanced information provided by intensity data in comparison to sparse height features. WSC employs the intensity information of the points to sparsify geometric features in the height direction. Furthermore, it utilizes a hybrid distance metric that combines cosine distance and Euclidean distance to quantify the similarity between two scenes. This integration of distance metrics reduces the sensitivity typically associated with cosine distance and enhances the overall performance of the method. Due to the absence of a unified reference frame and the usage of a simplified vector instead of a complete matrix in Scan Context, Shi et al. (2021) introduced a robust global place recognition method by enhancing the Scan Context approach. Their improved Scan Context employs a three-stage matching algorithm, which effectively enhances the performance of place recognition. To further advance the concept of a rotation invariance introduced by Scan Context, Kim et al. (2021) proposed Scan Context++ (SC++) which is capable of generating a versatile descriptor that is resilient to rotation and translation. SC++ extends the capabilities of its predecessor by incorporating two sub-descriptors, enabling topological place retrieval, and facilitating 1-DOF semi-metric localization thereby bridging the gap between topological place retrieval and metric localization. An additional benefit is that SC-like methods can easily integrate into existing LiDAR-based SLAM systems (Kim et al. 2022).

To recognize places by analyzing the projection of observed points along the gravity direction, Sánchez-Belenguer et al. (2020) proposed a robust global matching technique specifically designed for 3D mapping applications. By leveraging a global projection direction, the method introduces a matching algorithm that effectively compares places in a two-dimensional (2D) space and retrieves the corresponding relative three-dimensional (3D) transformations between maps.

Yin et al. (2020) proposed SeqSphereVLAD for orientation-invariant 3D-PCPR. This method can recognize places from previous trajectories regardless of variations in viewpoint and temporal observation differences. SeqSphereVLAD achieves this by projecting a 3D point cloud onto a spherical view to extract place descriptors which are then utilized in a coarse-to-fine sequence matching module designed to enhance the accuracy of 3D-PCPR. To achieve viewpoint-invariant 3D-PCPR while simultaneously balancing matching accuracy and search efficiency, Yin et al. (2021) introduced a fast sequence-matching enhanced viewpoint-invariant 3D-PCPR framework. This framework comprises two key modules: a spherical harmonic place descriptor extraction (SphereVLAD) and fast sequence matching. By leveraging this framework, the authors aimed to emulate human-like place recognition abilities by employing a novel 3D feature learning method. The SphereVLAD module is responsible for extracting unique place descriptors using spherical harmonics, while the fast sequence matching module focuses on efficient and accurate sequence matching.

Jiang et al. (2020) introduced LiPMatch for 3D LiDAR point cloud-based loop-closure detection and loop-closure correction. LiPMatch formulates each keyframe as a fully connected graph, where nodes represent planes. The method constructs a plane graph for each keyframe and leverages the geometric properties of the planes and their relative positions to detect loop closures. By identifying matched planes between keyframes, LiPMatch enhances the accuracy and robustness of SLAM algorithms, thereby improving the overall performance of the system.

To overcome the limitations of existing methods in terms of real-time loop recognition and full 6-DoF loop pose correction, Cui et al. (2023) introduced BoW3D, a novel bag-of-words approach for 3D LiDAR SLAM. BoW3D addresses these challenges by leveraging the LinK3D (Cui et al. 2024), which is an efficient, pose-invariant, and accurate point-to-point matching method specifically designed for 3D LiDAR data. By building a bag-of-words representation based on LinK3D, BoW3D efficiently recognizes revisited loop locations while also enabling real-time correction of the full 6-DoF loop pose.

4.2 2D Images based

The 2D images based methods in 3D-PCPR first project the raw 3D point clouds to 2D images, and then use the 2D images for place recognition. A notable method in this category was proposed by Cao et al. (2018), which accomplishes loop closure detection for SLAM. This method adopts an image model named Bearing Angle (BA) to convert 3D laser point clouds to 2D images. It then utilizes the ORB features (Rublee et al. 2011) extracted from BA images to perform scene matching and uses a visual Bag of Words (BoW) approach (Angeli et al. 2008; Gálvez-López and Tardos 2012) to improve the search efficiency. However, the performance of this method in large-scale unstructured environments has not been fully verified.

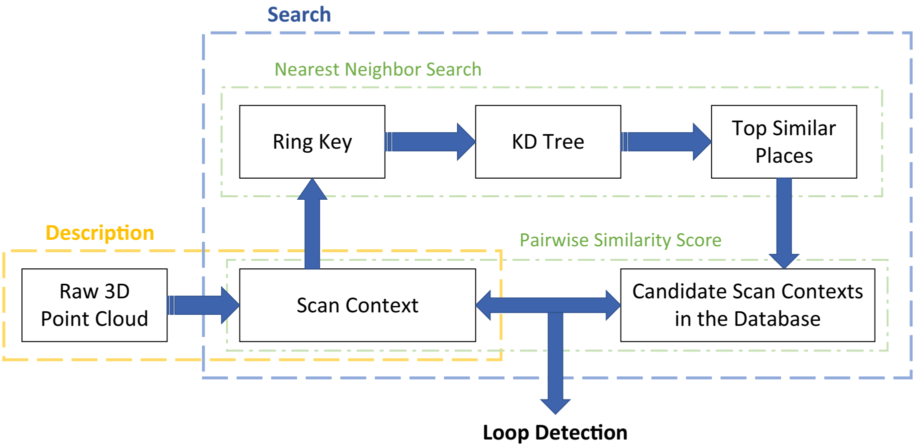

Fig. 6 A brief illustration of Scan Context algorithm (from Kim and Kim (2018)). A raw 3D point cloud is encoded into scan context and a 1-dimensional vector is used for retrieving the nearest candidates. For loop detection, the retrieved candidates are compared to the query Scan Context

Cop et al. (2018) introduced DELIGHT, a highly efficient global localization descriptor that solely relies on LiDAR data without requiring robot motion information. This pioneering work leverages the LiDAR intensity image data and encodes it into a unique descriptor comprising a collection of histograms. DELIGHT stands out as the first solution that offers a near real-time approach to global localization by utilizing global intensity descriptors. Guo et al. (2019) introduced ISHOT, a local descriptor designed to enhance robust place recognition by integrating the advantages of both geometry and appearance using LiDAR intensity image data. ISHOT combines geometric and texture information obtained from calibrated LiDAR intensity images to form a comprehensive local descriptor. The method then employs a probabilistic keypoint voting strategy for place recognition.

Kim et al. (2019) proposed a long-term LiDAR localization technique that leverages the structural information of an environment by projecting a raw point cloud to an image. They present a novel Scan Context Image descriptor for point clouds and an end-to-end CNN that effectively summarizes the unstructured point cloud into a structured form for robust long-term place recognition. Unlike existing methods such as M2DP (He et al. 2016) and PointNetVLAD like (Uy and Lee 2018) approaches, which rely on pairwise comparisons between a query and scans in a database, this method offers faster processing. Moreover, experimental results demonstrate consistent and robust year-round localization performance, even when trained in just a single day.

Schaupp et al. (2019) proposed OREOS, oriented place recognition in outdoor scenarios using LiDAR scans. Their approach involves several stages: firstly, the current raw 3D LiDAR point cloud is projected onto a 2D range image. Next, a CNN is employed to extract compact descriptors, followed by yaw estimation and local point cloud registration. To enhance performance, retrieve nearby place candidates, and estimate yaw discrepancy, the method utilizes a triplet loss function during training and incorporates a hard negative mining strategy. Cao et al. (2020) proposed a season-invariant and viewpoint-tolerant 3D-PCPR method to achieve long-term robust localization. To achieve robust place recognition across seasons, the method designs a compact cylindrical image model to project 3D point clouds to 2D images representing the prominent geometric relationships of scenes.

The structure of the algorithm mainly consists of two parts: a novel cylindrical image representation of a 3D point cloud and an efficient descriptor based on contexts and layouts of the scenes. Additionally, a sequence-based temporal consistency check is introduced to handle similar scenes and local structural changes.

OverlapNet (Chen et al. 2021) is a loop closing method for 3D LiDAR-based SLAM. It exploits the different cues generated from the point cloud such as range, normal, and intensity images, and semantic data to provide overlap and relative yaw angle estimates between pairs of 3D scans. The 3D point cloud is first converted to a 2D image, and the rotation information is represented as translation of the image. This translation is estimated by a differentiable phase correlation. OverlapNet estimates an image overlap generalized to range images and provides a relative yaw angle estimate between pairs of scans. Ma et al.proposed OverlapTransformer (Ma et al. 2022), an efficient yaw-angle-invariant transformer network for LiDAR-based place recognition. OverlapTransformer has three modules: Range Image Encoder, Attentional Feature Transformer, and VLAD. It is a lightweight transformer network that leverages range images projected from raw 3D point clouds to achieve faster online inference. In follow-up works, Ma et al.process sequential LiDAR scans with a transformer network, named SeqOT (Ma et al. 2023), and multiple different views (depth view and BEV) with another transformer network, named CVTNet (Ma et al. 2023), for more robust and reliable long-term place recognition.

Wang et al. (2020) proposed a global 3D LiDAR point cloud descriptor to improve the speed and accuracy of loop-closure detection. Their method projects a point cloud to a binary signature image after a couple of Gabor-filtering and thresholding operations on the LiDAR-Iris image representation. Point cloud pairs are matched by calculating the Hamming distance of their corresponding binary signature images. This work is somewhat similar to Scan Context (Kim and Kim 2018) but differs in three main ways: Firstly, it encodes the height information as the pixel intensity of the LiDAR-Iris image. Secondly, it extracts a binary feature map from the LiDAR-Iris image for loop-closure detection. Thirdly, the loop-closure detection is rotation-invariant with respect to the LiDAR's pose.

Leveraging high-resolution 3D LiDAR point clouds, Shan et al. (2021) proposed a method for robust, real-time place recognition. Their method extracts ORB features from the intensity images of point clouds and encodes them into bag-of-words vectors. Candidate frames are found by matching visual feature descriptors, and outliers are rejected by applying PnP and RANSAC. This method is specifically designed for LiDAR imaging and is the first to use projected LiDAR intensity images for place recognition. Di Giammarino et al. (2021) used different datasets to investigate the practicality of applying techniques from VPR to LiDAR intensity data. Their results suggest that visual representations (such as intensity images) of places are useful for place recognition and are an effective means for determining loop closures.

In order to solve the problem that the network models of place recognition methods with higher detection accuracy are usually very large, while the application speed of the methods with smaller network models are not fast enough in actual scenarios, Ye et al. (2022) proposed an efficient 3D-PCPR approach based on feature point extraction and transformer(FPET-Net). The method first projects the raw 3D point cloud to range image to get the horizontal index and scan index of each point and then calculates the curvature value to filter the feature points. Then, a point transformer module is developed to extract global descriptors. Finally, a feature similarity network module is used to calculate global descriptor similarity.

Ma et al. (2023) presented SeqOT, a transformer-based network designed for place recognition using sequential 3D LiDAR scans obtained from an onboard sensor. The method aims to effectively utilize the temporal and spatial information present in the sequential range images derived from the LiDAR data. SeqOT is an end-to-end network for long-term place recognition and uses a multiscale transformer to generate global descriptors for each LiDAR range image sequence. It finds similar places by matching the descriptor of the current query sequence with the descriptors stored in the map.

4.3 Bird-eye view based

Bird-eye view (BEV) based methods first project the raw 3D point clouds to BEV, then use the BEV information for subsequent place recognition. A prominent method in this category is DiSCO (Xu et al. 2021) (Differentiable Scan Context with Orientation) which can simultaneously find the scan at a similar place and estimate the relative orientation. The main idea of DiSCO is to convert the rotation-invariant signature to the translation-invariant frequency spectrum. It efficiently estimates the global optimal relative orientation by projecting a 3D point cloud to a polar BEV image and reorganizes the same height voxel values into image channels to construct a multi-layer BEV. Low overlap between input point clouds may lead to registration failures, especially in scenes where non-overlapping regions contain similar structures. To solve this problem and inspired by DiSCO (Xu et al. 2021), Li et al. (2023) presented a unified BEV model for jointly learning of 3D local features and overlap estimation for simultaneous pairwise registration and loop closure.

BVMatch (Luo et al. 2021) is a LiDAR-based frame-to-frame place recognition method that is able to estimate 2D relative poses. Since the ground area can be approximated as a plane, BVMatch employs a BEV image which is projected from the raw 3D point cloud as the intermediate representation and introduces the BVFT descriptor to perform matching. Leveraging the BVFT descriptors, the method unifies the 3D-PCPR task and pose estimation. However, BVMatch cannot generalize well to unseen environments. In a follow-up work, the authors proposed a rotation-invariant network called BEVPlace (Luo et al. 2023), as shown in Fig. 7. It uses group convolution (Cohen and Welling 2016) to extract rotation-equivariant local features from BEV images, and NetVLAD (Arandjelovic et al. 2016) for global feature aggregation. Furthermore, BEVPlace observes that the distance between BEV features correlates with the geometric distance of point clouds. Based on the above structure, BEVPlace is able to estimate the position of the query point cloud for place recognition.

4.4 Summary

To summarize, the main idea of the projection-based 3D-PCPR methods is to first project the raw 3D point clouds to 2D planes, images, or BEV information, and then use the projected information for subsequent processing to achieve place recognition. Based on the planar projection, it lays the foundation for hand-crafted descriptors, such as Scan Contex (Kim and Kim 2018), BOW3D (Cui et al. 2023), and other methods. Projection is followed by image feature extraction or a deep learning network to construct place recognition algorithms, such as OREOS (Schaupp et al. 2019), OverlapNet (Chen et al. 2020a), OverlapTransformer (Ma et al. 2022), SeqOT (Ma et al. 2023), etc. Furthermore, the projected image or BEV information sequence (or their combination) can also be used to construct new place recognition algorithms, such as BEVPlace (Luo et al. 2023), etc. Projection-based methods have achieved great success recently, however, in the process of projecting 3D point clouds to planes, images, or BEV, there is inevitable information loss, which can undermine the place recognition accuracy. Multi-projection-based methods (Ma et al. 2023) can mitigate the information loss, however, such methods increase the processing time, resulting in a trade-off between accuracy and time cost.

5 Segment based methods

Segment-based methods are another popular category in 3D-PCPR. The main idea is to segment the raw point cloud and then use the post-segment features, semantic information, or graph structure for subsequent processing to realize place recognition. We divide segment-based methods into three categories: post-segment features-based, semantic-based, and graph-based methods.

5.1 Post-segment features based

Post-segment features-based methods utilize the post-segment features of the raw 3D point clouds for subsequent place recognition. A pioneering method in this category is SegMatch (Dubé et al. 2017) which is the first to present real-time loop-closure detection and localization in 3D point clouds (see Fig. 8). SegMatch first segments the raw 3D point cloud into sets of point clusters and then uses post-segment features encoded by a CNN on the clusters to find place matches. Finally, a geometric verification step is applied to turn the candidate matches into place recognition candidates. Since segmentation provides a good compromise between local and global descriptions by combining their advantages while mitigating their disadvantages, this method not only reduces the matching time but also decreases the likelihood of false matches. In a follow-up work, Dubé et al. (2018) proposed an incremental segment-based localization for 3D-PCPR, which utilizes an incremental segmentation algorithm to track the evolution of single segments. It is the first work to propose combining incremental solutions to normal estimation, segmentation, and recognition for finding global associations in 3D point clouds. It is interesting to investigate incremental updates of learning-based descriptors that can potentially gain discriminative power and reliability over time. To precisely estimate a robot's pose in unstructured, dynamic environments, Dube et al. (2020) also put forward SegMap, a 3D segment mapping method using Data-Driven Descriptors. SegMap decomposes the robot's surroundings into a set of segments, each represented by a distinctive, low-dimensional learning-based descriptor. It is the first work on robot localization proposing to reuse the extracted features for reconstructing 3D environments and extracting semantic information.

Tinchev et al. (2018) proposed Natural Segmentation and Matching (NSM), an extension of SegMatch (Dubé et al. 2017), for place recognition in both urban and natural environments. Their method first uses a feature extraction module to extract stable and reliable object-sized segments from point clouds. Next, repeatable oriented key poses are extracted and matched with a reliable shape descriptor using Random Forests to estimate the current sensor's position within the target map. The key poses extraction module segments and defines consistent orientated coordinate frames for object-sized segments, and the descriptor is employed to recognize different instances of the same segment. To adapt to online applications, Tinchev et al. (2019) then explored laser-based localization in both urban and natural environments and proposed a deep learning approach capable of learning meaningful descriptors directly from 3D point clouds as well as a feature space representation for the set of segmented point clouds.

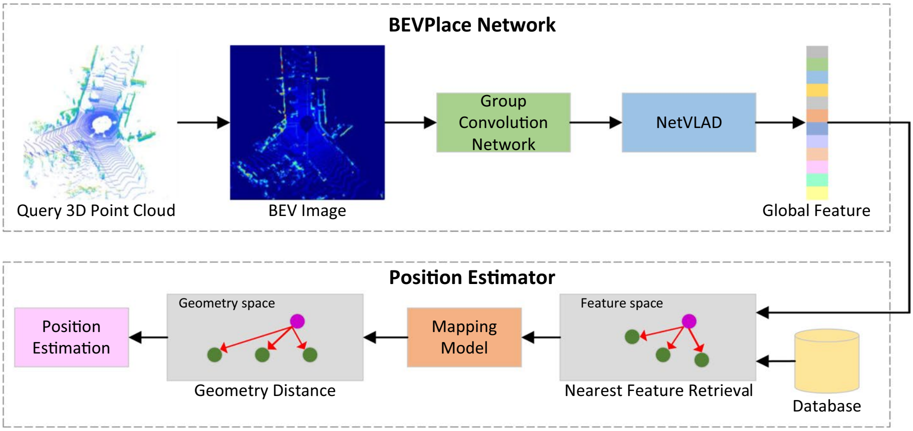

Fig. 7 Illustration of BEVPlace modules (from Luo et al. (2023)). The BEVPlace network projects point clouds to BEV images and extracts rotation-invariant global features. The position estimator module recovers geometry distances from feature space and estimates the positions of query point clouds

Their main contribution is a novel description method for segment-based 3D-PCPR, using a lightweight model that can be deployed using only a CPU.

SEED (Fan et al. 2020) is a segmentation-based egocentric 3D point cloud descriptor for loop closure detection. For robustness to noise and low/varying resolution, the method first obtains different segmented objects and then encodes their topological information into descriptors. The method is rotation invariant and insensitive to translation variations. However, its performance drops significantly when there are fewer objects in the scene.

Tomono (2020) proposed a method that uses geometric segments, such as planes, lines, and balls, to reduce the number of matching elements in the point cloud registration process for loop detection. Their method uses geometric constraints between the segments to achieve robustness and reduce matching element combinations, for real-time loop detection. However, when the environment lacks a sufficient number of salient objects and physical features, it struggles to find good loop hypotheses due to the lack of geometric segments. Locus (Vidanapathirana et al. 2021) is another 3D-PCPR method for large-scale environments. It encodes topological and temporal information related to components (obtained through segmentation) of the scene. To generate a fixed-length global descriptor, a second-order pooling along with a nonlinear transform is used to aggregate the extracted multi-level features.

Wietrzykowski and Skrzypczyński (2021) proposed an extension to the segment-based global localization method for LiDAR SLAM using descriptors learned from the visual context of the segments. This method represents one of the pioneering approaches that utilize intensity images to enhance the learned descriptors of 3D segments and investigate the learning of segment descriptions that are visible in images. The solution falls between learning to describe segments that occupy part of the image and finding the context in the description. This method is inherited from SegMap (Dube et al. 2020) but achieves better performance.

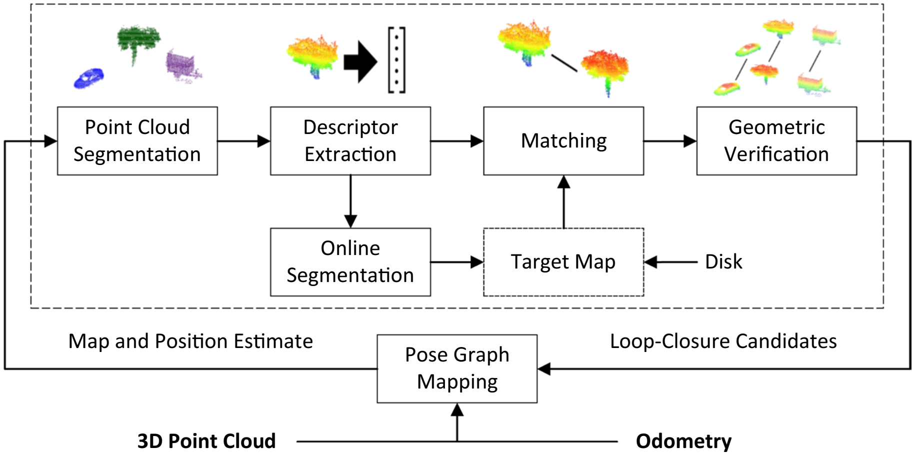

Fig. 8 Illustration of SegMatch block (from Dubé et al. (2017)). SegMatch is a modular place recognition algorithm composed of 4 main modules: point cloud segmentation, feature extraction, segment matching, and geometric verification

5.2 Semantic based

Semantic-based methods for 3D-PCPR utilize the semantic information of the segmented 3D point cloud for subsequent place recognition. For example, Zaganidis et al. (2019) presented a SLAM pipeline based on semantic-assisted NDT and PointNet++ (Qi et al. 2017) for place recognition and loop closure detection. Their method first segments the raw 3D point cloud and then utilizes geometric and semantic information of the environment and a single deep semantic segmentation network for registration and loop closure detection.

Semantic scan context (SSC) (Li et al. 2021) is a large-scale place recognition method that leverages high-level semantics features and corrects the translation between point clouds for improved accuracy. The algorithm framework mainly consists of two parts: a two-stage global semantic iterative closest point (ICP) (Besl et al. 1992) algorithm and a semantic scan context (SSC). Semantic segmentation is first performed on the raw 3D point cloud, and then the semantic information is used to preserve representative objects and project them into the x-y plane. The two-stage global semantic ICP is performed on the projected point cloud to obtain the 3D pose which is used to align the original point cloud and generate global descriptors. Finally, the similarity score is obtained by matching the global descriptors. Similar to most place recognition methods, the SSC method does not consider pitch and roll angles, leading to a possible failure in some extreme cases. Li et al. (2021) presented a global semantic descriptor for 3D-PCPR. To resolve ambiguous geometric features in scenes containing similar objects, their algorithm mainly relies on static semantic information such as trunks, poles, traffic signs, buildings, roads, and sidewalks. The descriptors not only record the geometrical structure of a 3D LiDAR scan but also encode the semantic distribution information.

Yin et al. (2021) proposed PSE-Match, a viewpoint-free place recognition method with parallel semantic embedding. PSE-Match incorporates a divergence place learning network to capture different semantic attributes in parallel through the spherical harmonics. This way, the observed variance of semantic attributes is smaller than the original point cloud.

5.3 Graph based

Graph-based 3D-PCPR methods utilize the graph structure information of the segmented point cloud for subsequent place recognition. Semantic graph-based place recognition (SGPR) (Kong et al. 2020) is a pioneer semantic graph representation and graph matching method for 3D-PCPR. Getting its inspiration from how humans perceive the environment by distinguishing scenes through semantic objects and their topological relations, SGPR utilizes semantic segmentation on raw 3D point clouds to obtain instances and further collects semantic and topological information together to acquire nodes forming the semantic graph. It leverages the semantic level to achieve superior robustness to environmental changes. The method is rotation invariant since the network can capture topological and semantic information from the point cloud. However, given its reliance on semantic segmentation, SGPR still suffers from bottlenecks, such as the test dataset's pre-defined semantic classes.

Zhu et al. (2020) proposed GOSMatch, a graph-of-semantics matching method for loop detection and 3D-PCPR. GOSMatch leverages the spatial relationship between semantics to generate descriptors and employs a coarse-to-fine strategy to efficiently search for loop closures. Once the loop closure is confirmed, GOSMatch can give an accurate 6-DOF initial pose estimate. This is the first method that leverages object-level semantics graphs to detect loop closures in 3D laser data. Instead of manually constructing the graph, Shi et al. (2021) employed an extension of graph data analysis methods Graph Neural Network (GNN) (Waikhom and Patgiri 2022) to facilitate the keypoint matches between two point clouds, which were subsequently utilized for point cloud registration and place recognition (Shi et al. 2023). Utilizing a GNN-based approach allows for the extraction of improved point matches, leading to enhanced accuracy and robustness in pose estimation and place recognition outcomes. SA-LOAM (Li et al. 2021) is a semantic-aided LiDAR SLAM method with loop closure detection. It leverages a semantic-assisted ICP, including semantic matching, downsampling, and planar constraint, and integrates a semantic graph-based place recognition method in the loop closure detection module. SA-LOAM exploits semantic information to improve the accuracy of point cloud registration and designs a semantic-graph-based loop closure detection module to eliminate the accumulated error.

To leverage the spatial relations of internal structures for place recognition, Gong et al. (2021) presented a two-level framework based on a spatial relation graph. The framework first segments the 3D point cloud into multiple clusters, then extracts features from each cluster and the spatial relation descriptors between clusters to represent the 3D point cloud scene. Finally, a two-level matching model is proposed for accurately and efficiently matching the spatial relation graph. Dai et al. (2022) proposed a new place recognition method named SC-LPR, which uses spatiotemporal contextual information from LiDAR scans to increase the capacity of feature representation. A semantic graph is constructed to represent the topological geometric map, and an end-to-end network is designed to predict similarity.

5.4 Summary

To sum up, the main idea of the segment-based 3D-PCPR methods is to segment the raw 3D point cloud and then use the segmented point cloud features, semantic information, or graph structure information for place recognition. Post-segment features-based place recognition methods can reduce the number of calculations and extract more effective point cloud features, such as SegMatch (Dubé et al. 2017), Locus (Vidanapathirana et al. 2021), etc. Semantic-based place recognition methods introduce high-level semantic information after segmentation to improve the accuracy of place recognition, such as SSC (Li et al. 2021), PSE-Match (Yin et al. 2021), etc. Graph-based place recognition methods use the instance information formed after segmentation to construct the graph structure and recognize the scene by identifying object-object relationships, such as SGPR (Kong et al. 2020), GOSMatch (Zhu et al. 2020), etc. Segment-based 3D-PCPR methods have further promoted the development of place recognition algorithms. However, this category of methods relies heavily on the accuracy of point cloud segmentation and semantic recognition. Further research is required to overcome this bottleneck.

6 Multimodal based methods

In adverse conditions, place recognition with a single-sensor or single-method could become challenging. Therefore, multimodal-based methods are also popular where the main idea is to combine 3D point clouds with other data (or sensor) modalities, such as RGB images, range images, and/or BEV information, etc. The combined multimodal data is then used as input for the subsequent place recognition processing. We divide multimodal-based methods into three categories: camera-LiDAR based, radar-LiDAR based, and multi-view fusion based methods.

6.1 Camera-LiDAR based

Camera-LiDAR based methods for 3D-PCPR mainly combine 3D point clouds with camera image information as input data for subsequent place recognition. LiDAR suffers from limitations such as motion distortion, degenerate environment, and limited range (since the laser may not reflect back with sufficient strength from far off objects). On the other hand, cameras do not have these limitations but encounter problems associated with varying illumination, occlusions, and season changes. Therefore, increasing attention has been paid to developing methods for fusing the information from cameras and LiDAR sensors. For example, Żywanowski et al. (2020) made a comparison of camera-based, 3D LiDAR-based, and joint camera-LiDAR-based place recognition across different weather conditions and concluded the need for more research on loop closures performed with multisensory fusion.

Xie et al. (2020) proposed a fusion algorithm that robustly captures the image and point cloud descriptors to solve the place recognition problem. In their method, point cloud descriptors are obtained with PointNetVLAD (Uy and Lee 2018) and image-based descriptors are extracted using ResNet50. A fully-connected layer is then employed to produce a compact global multimodal descriptor for each place. Their network finally learns an optimal metric to describe the similarity of the fused global descriptors for end-to-end place recognition. Lu et al. (2020) proposed PIC-Net, a point cloud and image collaboration network for large-scale place recognition. PIC-Net uses spatial attention VLAD to fuse the discriminative points and pixels, and mines the complementary information between the image and point cloud. Comparative results show that PIC-Net outperforms the image-based and point cloud-based methods. Pan et al. (2021) presented CORAL, a bi-modal descriptor place recognition method that can extract a compound global descriptor from camera and LiDAR data. It first builds an elevation image generated from the 3D point cloud as a structural representation. The elevation image is then enhanced with projected RGB image features and processed using a deep neural network. A NetVLAD layer is employed to aggregate the extracted local features.

MinkLoc++ (Komorowski et al. 2021) was proposed by Komorowski et al.as a LiDAR and monocular image fusion method for place recognition. As shown in Fig. 9, MinkLoc++ puts forward a discriminative multimodal descriptor based on a point cloud from a LiDAR and an image from an RGB camera. The method uses a fusion approach, where each modality is processed separately and fused in the final part of the processing pipeline. MinkLoc++ is an effective solution for the problem of dominating modality which adversely affects the discriminability of a multimodal descriptor.

By leveraging the benefits of semantic understanding, Cramariuc et al. (2021) introduced SemSegMap, an extension of SegMap (Dube et al. 2020), which seamlessly combines color and semantic information from an RGB camera with LiDAR data in real-time. SemSegMap introduces novel processes for segmentation and descriptor extraction. The integration of cameras into a LiDAR-equipped platform is typically straightforward in real-world robotic applications. This method has demonstrated commendable performance and holds promising prospects for practical applications.

Many existing camera-LiDAR fusion methods simply combine the two sensors without considering their performance characteristics in different environments. To address this limitation, Lai et al. (2022) introduced AdaFusion, an adaptive weighting visual-LiDAR fusion method. AdaFusion goes beyond conventional approaches by dynamically learning the weights for both image and 3D point cloud features. By utilizing an attention branch network, AdaFusion adaptively assigns weights to the camera and LiDAR sensors based on the current environmental conditions which enhances the system's recognition accuracy and robustness across various environments. AdaFusion represents a significant improvement in fusion techniques, enabling more effective utilization of camera and LiDAR data.

6.2 Radar-LiDAR based

Radar-LiDAR based 3D-PCPR methods combine 3D point clouds obtained from a radar and a LiDAR to perform subsequent place recognition. A notable multimodal range dataset for this line or research is MulRan (Kim et al. 2020) that contains radar and LiDAR data of urban environments. MulRan focuses on range sensor-based place recognition and provides 6D baseline trajectories of a vehicle for ground truth place recognition. This dataset is expected to promote the development of range-LiDAR based place recognition technology.

Yin et al. (2021) introduced Radar-to-LiDAR, a heterogeneous measurement-based framework for long-term place recognition. This method retrieves query radar scans from an existing LiDAR map. Initially, the radar and LiDAR points are encoded using Scan Context (Kim and Kim 2018) and then a shared U-Net transforms the handcrafted features to learned representations. Applying this method on a current radar scan, a robot can recognize the revisited LiDAR submaps.

Traditional 3D-PCPR methods assume that reliable prior maps are available. Tang et al. (2021) proposed a different approach which assumes that an overhead view of the workspace is available instead. The overhead view is used as a map for radar and LiDAR based localization. Their method consists of three steps: rotation inference, image generation, and pose estimation. To compare overhead imagery with ground-range sensor data, they propose a learned metric localization method that handles modality differences. This metric is cost-effective to train and can learn in a self-supervised manner without the need for metric-accurate ground truth. Based on this idea, off-the-shelf, publicly available overhead imagery (such as Google satellite imagery) can become a ubiquitous, low-cost, and powerful localization tool when prior maps are not available or convenient.

6.3 Multi-view fusion based

Multi-view fusion based 3D-PCPR methods mainly combine 3D point clouds and multi-view fusion information for place recognition. A notable method in this category is FusionVLAD (Yin et al. 2021), which is a parallel fusion network structure that learns the point cloud representations from multi-view projections and embeds them into viewpoint-free low-dimensional place descriptors for efficient global recognition. This method consists of a spherical-view branch for orientation invariant feature extraction and a top-down view branch for translation insensitive feature extraction. Moreover, a parallel fusion module is designed to enhance the combination of region-wise feature connection between the two branches.

Many existing 3D-PCPR methods adopt a shared representation of the input point cloud, disregarding different views and potentially underutilizing the LiDAR sensor's information. Ma et al. (2023) proposed CVTNet, a novel approach based on cross-view transformers, aimed at fusing range image views and Bird's Eye View (BEV) representations derived from LiDAR data. CVTNet leverages intra-transformers to capture correlations within each view and inter-transformers to capture correlations between the two distinct views. By utilizing CVTNet, a yaw-invariant global descriptor is generated for each LiDAR point cloud in an end-to-end fashion. This descriptor enables the retrieval of previously visited places by matching descriptors between the current query scan and a pre-built database.

The task of localizing images on a large-scale point cloud map is still relatively unexplored. To address this challenge, Li et al. (2023) introduced I2P-Rec, a method designed for image recognition on large-scale point cloud maps using BEV Projections. The BEV image serves as an intermediate representation, which is then fed into a CNN to extract global descriptors for matching purposes.

6.4 Summary

In summary, multimodal-based 3D-PCPR methods aim to overcome the limitations of single-sensor or single-modality approaches by fusing point cloud information with other modalities. This fusion leverages the complementary nature of multimodal information, such as Camera-LiDAR, Radar-LiDAR, and Multi-view Fusion, among others. These methods, exemplified by MinkLoc++ (Komorowski et al. 2021), AdaFusion (Lai et al. 2022), Radar-to-LiDAR (Yin et al. 2021), CVTNet (Ma et al. 2023), and others, strive to achieve robust and adaptable place recognition in complex and dynamic environments. Whereas multimodal 3D-PCPR methods enhance the robustness of place recognition, they also require careful synchronization and calibration of sensors, which can be a challenging task. Continued development in the field of multimodal-based 3D-PCPR holds promise for further advancements and improvements in place recognition capabilities.

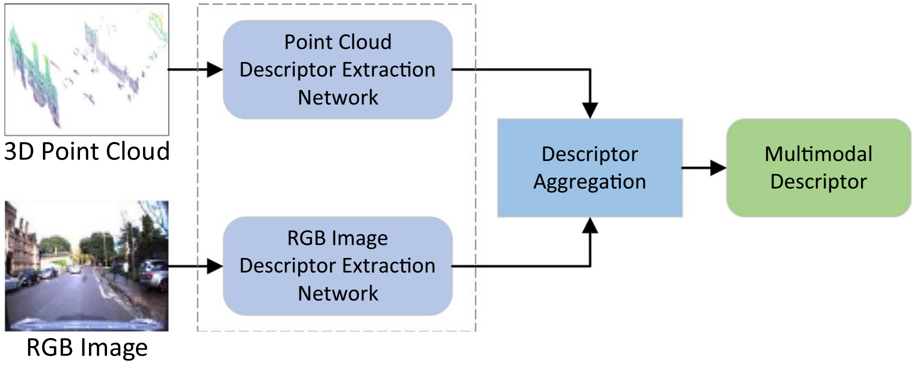
Point Cloud Descriptor Extraction Network
3D Point Cloud
RGB Image
RGB Image Descriptor Extraction Network
Descriptor Aggregation
Multimodal Descriptor

Fig. 9 Illustration of MinkLoc++ architecture. 3D point cloud and RGB image are processed by separate networks to extract their respective descriptors which are then aggregated to produce a fused multimodal descriptor

7 Datasets and performance

Given the emergence of numerous advanced algorithms for 3D-PCPR, conducting a comprehensive and unbiased performance evaluation and comparison of existing methods becomes crucial. In this section, we present a selection of prominent 3D-PCPR datasets and evaluation metrics commonly utilized for assessing the performance of these methods. Additionally, we present a performance comparison of mainstream algorithms in the field of 3D-PCPR to aid readers in gaining a clearer understanding of the strengths and limitations associated with different existing approaches.

7.1 Datasets

Public datasets play a pivotal role in advancing 3D-PCPR research. Numerous 3D point cloud datasets have been utilized to evaluate the performance of place recognition algorithms, serving as benchmark baselines and providing valuable ground truth information. These datasets enable researchers worldwide to conduct their investigations without being constrained by system or data limitations. In the following, we introduce a selection of popular and representative public datasets within the field, which are listed in Table 2. These datasets serve as valuable resources for evaluating and comparing different approaches, fostering progress and innovation in the domain of 3D-PCPR.

Ford Campus (Pandey et al. 2011): Ford Campus Vision and LiDAR dataset was collected by an autonomous ground vehicle testbed. The dataset consists of time-registered data from sensors mounted on the vehicle, collected while driving around the Ford Research campus and downtown Dearborn, Michigan during November-December 2009. The vehicle paths in the Ford campus dataset contain several large and small-scale loop closures, to assist in developing and testing place recognition algorithms. The dataset contains the vehicle's ground truth pose in the local coordinate system, including the vehicle's 3D rotation angle (roll, pitch, and yaw), 3D acceleration, 3D velocity, and timestamp.

KITTI (Geiger et al. 2013): KITTI (Karlsruhe Institute of Technology and Toyota Technological Institute) is one of the most popular datasets in mobile robotics, autonomous driving, and computer vision research. It consists of hours of traffic scenarios recorded with

Table 2 Representative public 3D point cloud datasets for place recognition research. GT in the header stands for Ground Truth

Dataset name	Year	Point cloud sensor	GT Acquisition	Type	Brief introduction
Ford Campus (Pandey et al. 2011)	2009	64-beam LiDAR, Stereo camera	POS-LV/GPS	Urban	Including time-registered data and several large and small-scale loop closures
KITTI (Geiger et al. 2013)	2012	64-beam LiDAR, Stereo camera	GPS/OXTS	Outdoor	6 h of traffic scenarios at 10–100 Hz using a variety of sensor modalities
NCLT (Carlevaris-Bianco et al. 2016)	2015	32-beam LiDAR, Stereo camera	RTK GPS	Outdoor	Including 34.9 h of logs covering 147.4 km of robot trajectory
Oxford RobotCar (Maddern et al. 2017)	2017	Stereo camera, LiDAR	GPS/INS	Outdoor	1,000 km, contains over 100 traversals of a consistent route
University Sector (U.S.) (Uy and Lee 2018)	2018	64-beam LiDAR	GPS/INS	Urban	10 km repeatedly at different time periods in a university sector
Residential Area (R.A.) (Uy and Lee 2018)	2018	64-beam LiDAR	GPS/INS	Suburban	8 km repeatedly at different time periods in a residential area
Business District (B.D.) (Uy and Lee 2018)	2018	64-beam LiDAR	GPS/INS	Urban	5 km repeatedly at different time periods in a business district
Oxford Radar (Barnes et al. 2020)	2020	32-beam LiDAR, Radar	GPS/INS	Outdoor	280 km, encompassed a variety of weather, traffic, and lighting conditions
MulRan (Kim et al. 2020)	2020	64-beam LiDAR, Radar	FOG/VRS-GPS	Urban	Multimodal dataset for radar and LiDAR, with 6D ground-truth poses
Haomo (Ma et al. 2022)	2022	32-beam LiDAR	RTK GNSS	Outdoor	5 long-term and reverse repeated trajectories to test the rotation invariance
HPointLoc (Yudin et al. 2023)	2022	RGB-D	-	Indoor	Consists 76,000 frames and 49 scenes from the Matterport3D (Chang et al. 2017) dataset
Perth-WA (Ibrahim et al. 2023)	2023	64-beam LiDAR	-	Urban	Including a 4 km^b map of the Perth CBD area at 3 different times and weather
Wild-Places (Knights et al. 2023)	2023	16-beam LiDAR	INS	Natural	Contains 8 lidar sequences over the course of 14 months, and 63K undistorted LiDAR submaps

a variety of sensor modalities, including high-resolution RGB, grayscale stereo cameras, and a 3D laser scanner. 3D-PCPR methods mainly use the 3D LiDAR data. The dataset provides 11 sequences with ground truth trajectories for training (00–10) and 11 sequences without real trajectories for evaluation (11–21).

NCLT (Carlevaris-Bianco et al. 2016): NCLT is a large-scale, long-term autonomy dataset, including 34.9 h of logs covering 147.4 km of robot trajectory collected on the University of Michigan's North Campus. It consists of omnidirectional imagery, 3D LiDAR, planar LiDAR, GPS, and ground-truth pose information. It has 27 sessions, each containing both indoor and outdoor environments, spaced approximately biweekly over the course of 15 months. Although the same area is repeatedly explored, the path for each session is varied, as is the time of the day for each session-from early morning to just after dusk. NCLT can facilitate long-term place recognition research in challenging environments such as moving obstacles, changing lighting, varying viewpoints, seasonal and weather changes, and long-term structural changes caused by construction projects. The dataset uses LiDAR scan matching and high-precision RTK GPS to provide ground truth robot pose.

Oxford RobotCar (Maddern et al. 2017): This dataset was collected by repeatedly traversing an approximately 10 km route in central Oxford, UK for over one year. It contains 100+ traversals of a consistent route, capturing the large variation in appearance and structure of a dynamic city environment over long periods of time. The dataset contains images, LiDAR, GPS, and INS ground truth data, captured in many different combinations of weather, traffic, and pedestrians, along with longer-term changes such as construction and roadworks.

USRABD (Uy and Lee 2018): USRABD dataset is a collection of three datasets proposed by the authors of PointNetVLAD (Uy and Lee 2018) for 3D-PCPR. The three datasets include a university sector (U.S.), a residential area (R.A.), and a business district (B.D.) dataset. USRABD dataset was collected using a LiDAR sensor mounted on a car. The data collection vehicle traveled through three areas of U.S., R.A., and B.D. covering a distance of 10, 8, and 5 km repeatedly at different time periods. This dataset has been used as a mainstream benchmark often together with the Oxford RobotCar dataset (Maddern et al. 2017). Ground truth GPS coordinates for the three datasets can be found in the corresponding csv files.

Oxford Radar RobotCar (Barnes et al. 2020): This dataset is a radar extension to The Oxford RobotCar dataset. It mainly utilizes a Navtech CTS350-X Millimetre-Wave FMCW radar and Dual Velodyne HDL-32E LiDARs for 280 km of driving around Oxford, UK. The dataset was gathered in January 2019 over 32 traversals of a central Oxford route and includes a variety of weather, traffic, and lighting conditions. In addition to the raw sensor recordings from all sensors, this dataset provides an updated set of calibrations, ground truth trajectories for the radar sensor as well as MATLAB and Python development tools for leveraging the data.

MulRan (Kim et al. 2020): MulRan is a multimodal range dataset for radar and LiDAR specifically targeting the urban environment. It focuses on the 3D-PCPR problem and provides 6D baseline trajectories of a vehicle for place recognition ground truth. MulRan captures both temporal and structural diversities for 3D place recognition research.

Haomo (Ma et al. 2022): This dataset was collected in urban environments of Beijing by a mobile robot built by HAOMO.AI Technology company equipped with a HESAI PandarXT 32-beam LiDAR sensor, a SENSING-SG2 wide-angle camera, and an ASENSING-INS570D RTK GNSS. There are currently five sequences: seq 1–1 and 1–2 were collected from the same route on 8th December 2021 with opposite driving directions. An additional seq 1–3 from the same route is utilized as the online query with respect to both 1–1 and 1–2 respectively to evaluate place recognition performance of forward and reverse driving. Seq 2–1 and 2–2 are collected along a much longer route from the same direction, but on different dates i.e. 28th December 2021 and 13th January 2022, respectively. The former is used as a database while the latter one is used as query. The two sequences are for evaluating the performance for large-scale long-term place recognition.

HPointLoc (Yudin et al. 2023): HPointLoc is a point cloud-based indoor place recognition dataset with synthetic RGB-D images. It is based on the popular Habitat (Savva et al. 2019) simulator from 49 photorealistic indoor scenes from the Matterport3D (Chang et al. 2017) dataset and contains 76,000 frames. The HPointLoc dataset is split into two parts: the validation HPointLoc-Val, which contains only one scene, and the complete HPointLoc-All dataset, containing all 49 scenes, including HPointLoc-Val. Although the dataset does not have ground truth poses, it provides an estimate of an average registration error between corresponding surface points of 1 cm or less.

Perth-WA (Ibrahim et al. xxx): Perth-WA dataset was first presented in (Ibrahim et al. 2023), and contains 6DoF annotations for localization in a 3D point cloud map of the Perth city in Western Australia. The 3D map is constructed using a 64-channel LiDAR and covers 4 km^2 region of the Perth Central Business District. The dataset scenes contain commercial structures, residential areas, food streets, complex routes, and hospital buildings etc. Perth-WA was collected in 3 different 2-hour sessions under day/night conditions with sunny and cloudy weather. Particularly, its labels come directly from the LiDAR frames themselves. This dataset leverages the map creation process itself to extract ground truth poses and contains loop data with LiDAR frames and their ground truth pose labels in text files.

Wild-Places (Knights et al. 2023): Wild-Places is a challenging large-scale dataset for 3D-PCPR in unstructured, natural environments. It contains 8 LiDAR sequences collected with a handheld sensor payload over the course of 14 months, containing a total of 63K undistorted LiDAR submaps along with accurate 6DoF ground truth. Wild-Places contains multiple revisits and uses Wildcat (Ramezani et al. 2022) system to generate accurate intra-sequence ground truth.

7.2 Evaluation metrics

Numerous evaluation metrics have been proposed to test the effectiveness of place recognition methods (Li et al. 2021; Cui et al. 2023; Ferrarini et al. 2020). Here, we introduce some of the commonly used evaluation metrics.

Precision (P): Precision denotes the ratio between the correct matches and the total of the predicted positive matches. Precision is defined as:

$$P = \frac{TP}{TP + FP}, \tag{1}$$

where TP are the number of True Positives (i.e. correct matches), FP are False Positives (i.e. incorrect matches), FN are False Negatives (i.e. matches erroneously excluded from the query results).

Recall (R): Recall is the proportion of real positive cases that are correctly identified as positive matches. Formally:

$$R = \frac{TP}{TP + FN} \tag{2}$$

Recall@N is also commonly used which measures the proportion of relevant items retrieved in the top N results. Particularly, Average Recall@1 (AR@1) measures the proportion of relevant items retrieved as the first item in the list of results and Average Recall@1% (AR@1%) is calculated taking into account top-k matches, where k is 1% of the database size. Higher values of Recall@N, Recall@1, and Recall@1% indicate better performance.

PR-Curve: *PR*-Curve is a graph with recall values on the x-axis and precision values on the y-axis. It shows the relationship between precision and recall values.

F_1 score: F_1 score combines the precision and recall values into a single metric by taking the harmonic mean of P and R. It treats P and R as equally important and measures the overall performance of the test systems. The F_1 score is defined as:

$$F_1 = 2 \times \frac{P \times R}{P + R} \tag{3}$$

where P and R represent the Precision and Recall values, respectively.

Extended Precision (*EP*): Extended Precision (*EP*) provides more comprehensive insights into place recognition performance. It is designed specifically for evaluating place recognition algorithms. The Extended Precision is defined as:

$$EP = \frac{1}{2}\left(P_{R0} + R_{P100}\right), EP \in [0, 1] \tag{4}$$

where P_{R0} is the precision at minimum Recall value, and R_{P100} is the max Recall at 100% Precision, i.e. it is the highest value of the recall that can be reached without any False Positives (*FP*).

7.3 Performance comparison

We present the comparative performance of some representative algorithms in 3D-PCPR on typical public datasets, including Oxford RobotCar dataset (Maddern et al. 2017), USRABD dataset (Uy and Lee 2018): University Sector (U.S.), Residential Area (R.A.), Business District (B.D.), and KITTI dataset (Geiger et al. 2013). The results are collected from the original papers (Uy and Lee 2018; Komorowski 2021, 2022; Kong et al. 2020; Li et al. 2021; Cui et al. 2023; Liu et al. 2019; Luo et al. 2023; Hou et al. 2022; Xu et al. 2021; Lai et al. 2022).

In Table 3, we present a performance comparison of some state-of-the-art 3D-PCPR methods (including PointNetVLAD[1] (Uy and Lee 2018), PCAN[2] (Zhang and Xiao 2019), LPD-Net[3] (Liu et al. 2019), EPC-Net[4] (Hui et al. 2022), SOE-Net[5] (Xia et al. 2021), HiTPR (Hou et al. 2022), MinkLoc3D[6] (Komorowski 2021), NDT-Transformer[7] (Zhou

[1] https://github.com/mikacuy/pointnetvlad.git.

[2] https://github.com/XLechter/PCAN.git.

[3] https://github.com/Suoivy/LPD-net.git.

[4] https://github.com/fpthink/EPC-Net.git.

[5] https://github.com/Yan-Xia/SOE-Net.git.

[6] https://github.com/jac99/MinkLoc3D.git.

[7] https://github.com/dachengxiaocheng/NDT-Transformer.git.

et al. 2021), PPT-Net[8] (Hui et al. 2021), SVT-Net[9] (Fan et al. 2022), TransLoc3D[10] (Xu et al. 2021), MinkLoc3Dv2[11] (Komorowski 2022), OREOS (Schaupp et al. 2019), Scan Context[12] (Kim and Kim 2018), DiSCO[13] (Xu et al. 2021), BEVPlace[14] (Luo et al. 2023), PIC-Net (Lu et al. 2020), MinkLoc++[15] (Komorowski et al. 2021), CORAL (Pan et al. 2021), AdaFusion[16] (Lai et al. 2022)) according to the categories of feature-based, projection-based, and multimodal-based methods.The evaluation is based on the AR@1 and AR@1% metrics. The performance of each method shown in this table is mainly evaluated on the Oxford RobotCar dataset (Maddern et al. 2017) and USRABD dataset (Uy and Lee 2018). These results provide valuable insights into the performance of the examined methods under specific conditions and facilitate comparison and analysis within the field of 3D-PCPR.

Table 3 provide a comprehensive overview of the advancements in the field of 3D-PCPR, highlighting the emergence of numerous state-of-the-art algorithms in recent years. Starting from the pioneering algorithm PointNetVLAD (Uy and Lee 2018), significant improvements have been made, as seen in the LPD-Net algorithm (Liu et al. 2019), which exhibits enhanced performance. More recently, the MinkLoc3D series algorithms (Komorowski 2021, 2022), the BEVPlace algorithms (Luo et al. 2023), and the AdaFusion (Lai et al. 2022) have further advanced the state-of-the-art. These algorithms demonstrate impressive performance on standard benchmark datasets and continue to progress and evolve in the field of 3D-PCPR.

According to the categories of feature-based, projection-based, and segment-based methods, Table 4 gives a performance comparison of some state-of-the-art methods (PointNetVLAD[17] (Uy and Lee 2018), M2DP[18] (He et al. 2016), ISC[19] (Wang et al. 2020), LiDAR Iris[20] (Wang et al. 2020), Scan Context[21] (Kim and Kim 2018), OverlapNet[22] (Chen et al. 2021), BoW3D[23] (Cui et al. 2023), SGPR[24] (Kong et al. 2020), SSC-RN[25] (Li et al. 2021)) in terms of the F_1 max scores and Extended Precision (F_1/EP), along with their capability to accurately correct the full 6-DoF loop pose on the KITTI dataset (Geiger et al. 2013). The evaluation focuses specifically on the sequences with loop closures (00, 02, 05, 06, 07, and 08) from the KITTI dataset. These sequences are selected to facilitate a convenient and standardized evaluation process. By examining the results in Table 4,

[8] https://github.com/fpthink/PPT-Net.git.

[9] https://github.com/ZhenboSong/SVTNet.git.

[10] https://github.com/slothfulxtx/TransLoc3D.git.

[11] https://github.com/jac99/MinkLoc3Dv2.git.

[12] https://github.com/irapkaist/scancontext.git.

[13] https://github.com/MaverickPeter/DiSCO-pytorch.git.

[14] https://github.com/zjuluolun/BEVPlace.git.

[15] https://github.com/jac99/MinkLocMultimodal.git.

[16] https://github.com/MetaSLAM/AdaFusion.git.

[17] https://github.com/mikacuy/pointnetvlad.git.

[18] https://github.com/LiHeUA/M2DP.git.

[19] https://github.com/wh200720041/iscloam.git.

[20] https://github.com/BigMoWangying/LiDAR-Iris.git.

[21] https://github.com/irapkaist/scancontext.git.

[22] https://github.com/PRBonn/OverlapNet.git.

[23] https://github.com/YungeCui/BoW3D.git.

[24] https://github.com/kxhit/SG_PR.git.

[25] https://github.com/lilin-hitcrt/SSC.git.

Table 3 Evaluation results of common 3D-PCPR methods according to the categories of feature-based, projection-based, and multimodal-based method (AR@1 and AR@1%)

		Oxford		U.S.		R.A.		B.D.		Mean	
Categories	Methods	AR@1	AR@1%	AR@1	AR@1%	AR@1	AR@1%	AR@1	AR@1%	AR@1	AR@1%
Feature based [a]	PointNetVLAD (Uy and Lee 2018)	62.8	80.3	63.2	72.6	56.1	60.3	57.2	65.3	59.8	69.6
	PCAN (Zhang and Xiao 2019)	69.1	83.8	62.4	79.1	56.9	71.2	58.1	66.8	61.6	75.2
	LPD-Net (Liu et al. 2019)	86.3	94.9	87.0	96.0	83.1	90.5	82.5	89.1	84.7	92.6
	EPC-Net (Hui et al. 2022)	86.2	94.7	–	96.5	–	88.6	–	84.9	–	91.2
	SOE-Net (Xia et al. 2021)	–	96.4	–	93.2	–	91.5	–	88.5	–	92.4
	*HiTPR** (Hou et al. 2022)	86.6	93.7	80.9	90.2	78.2	87.2	74.3	79.8	80.0	87.7
	MinkLoc3D (Komorowski 2021)	93.0	97.9	86.7	95.0	80.4	91.2	81.5	88.5	85.4	93.2
	NDT-Transformer (Zhou et al. 2021)	93.8	97.7	–	–	–	–	–	–	–	–
	PPT-Net (Hui et al. 2021)	93.5	98.1	90.1	97.5	84.1	93.3	84.6	90.0	88.1	94.7
	SVT-Net (Fan et al. 2022)	93.7	97.8	90.1	96.5	84.3	92.7	85.5	90.7	88.4	94.4
	TransLoc3D (Xu et al. 2021)	95.0	98.5	–	94.9	–	91.5	–	88.4	–	93.3
	MinkLoc3Dv2 (Komorowski 2022)	96.3	98.9	90.9	96.7	86.5	93.8	86.3	91.2	90.0	95.1
Projection based[b]	OREOS (Schaupp et al. 2019)	46.46	68.47	–	–	–	–	–	–	–	–
	Scan Context (Kim and Kim 2018)	64.59	81.88	–	–	–	–	–	–	–	–
	DiSCO (Xu et al. 2021)	75.01	88.44	–	–	–	–	–	–	–	–
	BEVPlace⋆ (Luo et al. 2023)	96.5	99.0	96.9	99.7	92.3	98.7	95.3	99.5	95.3	99.2
Multimodal based[c]	PIC-Net (Lu et al. 2020)	–	98.22	–	–	–	–	–	–	–	–
	MinkLoc++ (Komorowski et al. 2021)	96.70	99.10	–	–	–	–	–	–	–	–
	CORAL (Pan et al. 2021)	88.93	96.13	–	–	–	–	–	–	–	–
	AdaFusion (Lai et al. 2022)	98.18	99.21	–	–	–	–	–	–	–	–

[a] The data marked with ∗ come from HiTPR (Hou et al. 2022) and the rest of the data in this category come from MinkLoc3Dv2 (Komorowski 2022)

[b] The data marked with ⋆ come from BEVPlace (Luo et al. 2023) and the rest of the data in this category come from DiSCO (Xu et al. 2021)

[c] All data in this category come from AdaFusion (Lai et al. 2022)

Table 4 Comparative results using F_1 max scores and Extended Precision (F_1 / EP) on KITTI dataset. LC in the header stands for Loop correction

Categories	Methods	00	02	05	06	07	08	Mean	LC
Feature based[a]	PointNetVLAD (Uy and Lee 2018)	0.779/0.641	0.727/0.691	0.541/0.536	0.852/0.767	0.631/0.591	0.037/0.500	0.595/0.621	–
Projection based[b]	M2DP (He et al. 2016)	0.708/0.616	0.717/0.603	0.602/0.611	0.787/0.681	0.560/0.586	0.073/0.500	0.575/0.600	–
	ISC (Wang et al. 2020)	0.657/0.627	0.705/0.613	0.771/0.727	0.842/0.816	0.636/0.638	0.408/0.543	0.670/0.661	–
	LiDAR Iris (Wang et al. 2020)	0.668/0.626	0.762/0.666	0.768/0.747	0.913/0.791	0.629/0.651	0.478/0.562	0.703/0.674	–
	Scan Context (Kim and Kim 2018)	0.750/0.609	0.782/0.632	0.895/0.797	0.968/0.924	0.662/0.554	0.607/0.569	0.777/0.681	1-DoF
	OverlapNet (Chen et al. 2021)	0.869/0.555	0.827/0.639	0.924/0.796	0.930/0.744	0.818/0.586	0.374/0.500	0.790/0.637	1-DoF
	BoW3D (Cui et al. 2023)	0.977/0.981	0.578/0.704	0.965/0.969	0.985/0.985	0.906/0.929	0.900/0.866	0.885/0.906	6-DoF
Segment based[c]	SGPR (Kong et al. 2020)	0.820/0.500	0.751/0.500	0.751/0.531	0.655/0.500	0.868/0.721	0.750/0.520	0.766/0.545	–
	SSC-RN (Li et al. 2021)	0.939/0.826	0.890/0.745	0.941/0.900	0.986/0.973	0.870/0.773	0.881/0.732	0.918/0.825	3-DoF

[a,b,c] The three categories of data in this table come from SSC (Li et al. 2021) and BoW3D (Cui et al. 2023)

valuable insights can be gained regarding the performance and effectiveness of the analyzed state-of-the-art methods in the context of loop pose correction on the KITTI dataset.

As can be seen from Table 4, based on the KITTI dataset and F_1 max scores and Extended Precision (*EP*) evaluation metrics, many advanced 3D-PCPR algorithms have emerged (He et al. 2016; Uy and Lee 2018; Wang et al. 2020, 2020; Kong et al. 2020; Kim and Kim 2018; Chen et al. 2021; Li et al. 2021; Cui et al. 2023). The BOW3D (Cui et al. 2023) algorithm recently proposed by Cui et al.has not only achieved excellent performance (mean F_1/*EP*: 0.885/0.906) but also can be used to correct the full 6-DoF loop pose.

8 Applications and future trends

This section delves deeper into the downstream applications of 3D-PCPR technology and highlights the anticipated trends in future development. By exploring these applications and trends, researchers can gain a more holistic and expedited understanding of the potential uses and advancements within the realm of 3D-PCPR methods.

8.1 Applications

3D-PCPR is a key task in the navigation and localization of mobile robots, especially in large-scale, long-term, and complex scenes with closed loops. It has a wide range of applications, ranging from land to air and even interstellar exploration (Yin et al. 2022).

Firstly, major on land applications of 3D-PCPR on a large scale include robotics and autonomous driving. Autonomous driving is crucial for achieving intelligent transportation in the future. Currently, most research and development vehicles for autonomous driving are equipped with high-precision LiDAR sensors, enabling real-time acquisition of 3D point cloud data of the surrounding environment. We can expect that 3D-PCPR will play a key role in realizing Simultaneous Localization and Mapping (SLAM) (Chen et al. 2020a; Kim et al. 2022) in autonomous vehicles. Additionally, in the domain of robotics, there are numerous smart application scenarios that can benefit from 3D-PCPR, such as smart logistics distribution (Wang et al. 2019) and smart indoor navigation (Xiang et al. 2018), among others. These applications demonstrate the versatility and potential impact of 3D-PCPR in advancing various robotics-related endeavors.

In aerial settings, the widespread adoption and utilization of unmanned aerial vehicles (UAVs) equipped with high-precision LiDAR sensors have paved the way for the extensive application of place recognition technology. This technology holds great potential in various fields such as smart agriculture, aerial photography localization, rapid delivery, and even military reconnaissance (Maffra et al. 2018; Patel et al. 2020; Hongming et al. 2022; Aslan et al. 2022). With the aid of UAVs, place recognition technology can significantly contribute to enhancing aerial navigation and mapping capabilities, enabling efficient and accurate operations in these domains.

Finally, in the realm of interstellar exploration, where traditional positioning signals like GPS or Beidou are unavailable in outer space or on alien planets, the significance and criticality of autonomous localization and navigation based on 3D point clouds become paramount. This technology finds practical application in well-known interstellar missions such as NASA's robotic rover (Perseverance) operating on Mars and CNSA's teleoperated rover (Yutu-2) on the Moon (Witze 2020; Ding et al. 2022). The utilization of 3D-PCPR technology in these missions demonstrates its crucial role in enabling precise positioning, navigation, and mapping in extraterrestrial environments. As humankind ventures further into space exploration, the reliance on 3D-PCPR for autonomous localization and navigation will continue to grow, making it an essential component of interstellar missions and the exploration of alien worlds.

8.2 Research trends

3D-PCPR technology has witnessed widespread application and rapid development. Single frame-based 3D-PCPR methods, exemplified by high-accuracy algorithms like MinkLoc3Dv2 (Komorowski 2022) and others, have achieved remarkable progress. However, despite these advancements, numerous challenges and open issues remain to be addressed in this field. Based on our comprehensive analysis of over 180 research works, we now delve into the future research trends of 3D-PCPR. By providing a concise overview, we aim to inspire and guide future researchers in their exploration of this domain. By tackling these challenges, we can further enhance the accuracy, robustness, and efficiency of 3D-PCPR methods. The identified research trends are poised to shape the future of 3D-PCPR and drive its continued evolution as an essential technology for robotics, autonomous vehicles, aerial mapping, interstellar exploration, and beyond.

8.2.1 Based on sequence frames

Sequence-based 3D-PCPR methods leverage serialized multi-frame point clouds as input, enabling spatio-temporal feature fusion and descriptor generation. These methods surpass single-frame approaches by incorporating a broader range of information, mitigating the risk of overemphasizing intra-frame features. By employing inter-frame continuous consistency detection, sequence-based methods can capture more comprehensive and discriminative features, resulting in superior recognition performance over extended time periods. Prominent examples of sequence-based approaches, such as SeqLPD (Liu et al. 2019), SeqsphereVLAD (Yin et al. 2020), FSEPR (Yin et al. 2021), SeqOT (Ma et al. 2023), among others, have showcased inspiring and representative work in this direction. Figure 10 illustrates the structural diagram of SeqOT (Ma et al. 2023), a sequence-based method proposed by our team members. We anticipate that future research will yield further advancements in sequence-based 3D-PCPR, facilitating even more robust and accurate place recognition capabilities.

8.2.2 Based on long-term learning

To address the rapid decline in robustness exhibited by many classic place recognition algorithms when there is a significant time gap between the given map data and the input query data, researchers have recently proposed long-term learning-based strategies for 3D-PCPR. Long-term 3D-PCPR aims to dynamically update the model as new data streams in, enabling continuous learning of the evolving environment. This approach also tackles the challenge of catastrophic forgetting, which involves preserving the memory of the original environment while incorporating new information.

Minimizing catastrophic forgetting is a key challenge in long-term learning place recognition. Several noteworthy approaches have emerged in this area, including 1-Day Learning 1-Year Localization (Kim et al. 2019), Radar-to-LiDAR (Yin et al. 2021), SVLPR (Cao et al. 2020), InCloud (Knights et al. 2022), CCL (Cui and Chen 2023), among others.

These methods have undertaken meaningful and valuable explorations, yielding promising results in the context of long-term learning for 3D-PCPR.

8.2.3 Cross-modal localization

As the application scenarios for place recognition continue to expand, there are situations where the sensors used to collect offline map data and query data differ or multiple sensors are employed for data collection. In such cases, conventional 3D-PCPR methods that rely on single-modal information suffer from severe performance degradation. Hence, cross-modal place recognition has emerged as a promising research direction. Cross-modal place recognition aims to address the challenges of integrating data from different modalities for improved performance. Several notable approaches have paved the way for future cross-modal 3D-PCPR, including PIC-Net (Lu et al. 2020), MinkLoc3D++ (Komorowski et al. 2021), Get to the point (Tang et al. 2021), AdaFusion (Lai et al. 2022), Text2Pos (Kolmet et al. 2022), $(LC)^2$ (Lee et al. 2023), I2P-Rec (Li et al. 2023), UnLoc (Ibrahim et al. 2023), among others. However, several challenges remain in achieving more efficient cross-modal data synchronization, calibration, fusion, and the integration of high-dimensional semantic information with 3D point cloud data. Cross-modal place recognition still has a long way to go and researchers need to overcome these challenges to unlock its full potential in the future.

8.2.4 Global metric localization

Conventionally, place recognition focused only on identifying the current localization within a given map. However, to enable more advanced navigation and localization tasks, there is a growing demand for place recognition methods that can estimate precise poses or 6-DoF (degrees of freedom) while recognizing the place, ultimately providing global pose localization. Fortunately, some researchers have recognized this need and made significant contributions in this direction. Methods such as DH3D (Du et al. 2020), LCDNet (Cattaneo et al. 2022), BoW3D (Cui et al. 2023), Slice Transformer (Ibrahim et al. 2023), and others have emerged, offering the capability to estimate 6-DoF poses alongside place recognition. The development of these methods has accelerated the progress of 3D-PCPR based on global pose localization. It is foreseeable that this will become a prominent research trend in the future as the demand for precise pose estimation and global localization continues to grow in robotics and related fields.

9 Conclusion

This article presents a comprehensive survey of 3D-PCPR (3D Point Cloud-based Place Recognition) methods, aiming to provide readers with a thorough understanding of the field. The survey categorizes 3D-PCPR methods into four main categories based on the source of extracted features: feature-based, projection-based, segment-based, and multi-modal-based methods. Each category is discussed in detail, providing relevant introductions and explanations. To enhance readers' understanding, the survey also introduces common public datasets and evaluation methods used in the field of 3D-PCPR. Additionally, it compares the performance of mainstream methods in 3D-PCPR to highlight the algorithm performance of various methods. The article further explores the technical applications and

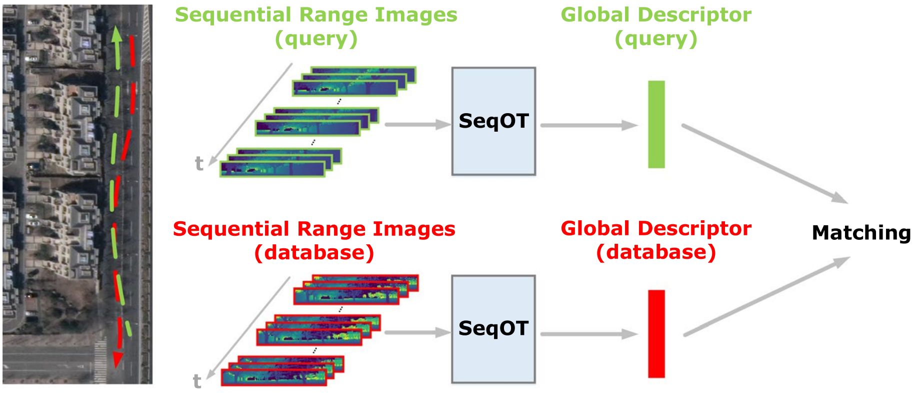
Sequential Range Images (query)
Global Descriptor (query)
SeqOT
t
Sequential Range Images (database)
Global Descriptor (database)
Matching
SeqOT
t

Fig. 10 Illustration of our SeqOT (Ma et al. 2023) which utilizes sequential range images projected from LiDAR sensors as input, simultaneously extracts spatial and temporal features, and generates a final sequence-augmented global descriptor

future development directions in the field of 3D-PCPR. Importantly, this survey represents the first comprehensive overview of 3D-PCPR methods that utilize 3D point clouds from different resources. It is intended to provide future researchers with a comprehensive view of the field, enabling them to contribute to the further advancement of 3D-PCPR.

Acknowledgements This work was supported in part by the National Natural Science Foundation of China under Grant U2013203, 62373140, U21A20487, U1913202, 62103137, and 61973106, the National Key Research and Development Program of China(2023YFB4704503), the Project of Science Fund for Distinguished Young Scholars of Hunan Province (2021JJ10024), the Leading Talents in Science and Technology Innovation of Hunan Province (2023RC1040), the Natural Science Fund of Hunan Province (2022JJ40100, 2022JJ30024), the Project of Talent Innovation and Sharing Alliance of Quanzhou City(2021C062L), the Key Research and Development Project of Science and Technology Plan of Hunan Province (2022GK2014), the Changsha Normal University Cultivation Project(K84022011). Professor Ajmal Mian is the recipient of an Australian Research Council Future Fellowship Award (project number FT210100268) funded by the Australian Government.

Author contributions K.L. wrote the main manuscript text, H.Y. conceived and provided overall revisions for the article, X.C. adjusted and refined the manuscript's structure, Z.Y. conducted the analysis and made revisions. J.W. contributed to the manuscript's framework, P.C. performed data analysis, and A.M. provided final polishing and revisions. All authors reviewed the manuscript.

Competing interests The authors declare no competing interests.

References

Angeli A, Filliat D, Doncieux S, Meyer J-A (2008) Fast and incremental method for loop-closure detection using bags of visual words. IEEE Trans Robot 24:1027–1037

Arandjelovic R, Gronat P, Torii A, Pajdla T, Sivic J (2016) NetVLAD: CNN architecture for weakly supervised place recognition. In: Proceedings of the IEEE conference on computer vision and pattern recognition (CVPR)

Aslan MF, Durdu A, Sabanci K, Ropelewska E, Gültekin SS (2022) A comprehensive survey of the recent studies with UAV for precision agriculture in open fields and greenhouses. Appl Sci 12:1047

Barnes D, Gadd M, Murcutt P, Newman P, Posner I (2020) The oxford radar Robotcar dataset: a radar extension to the oxford Robotcar dataset. In: IEEE International conference on robotics and automation (ICRA)

Barros T, Pereira R, Garrote L, Premebida C, Nunes UJ (2021) Place recognition survey: an update on deep learning approaches. arXiv:2106.10458

Bay H, Tuytelaars T, Van Gool L (2006) Surf: speeded up robust features. Lect Notes Comput Sci 3951:404–417

Beltran D, Basañez L (2014) A comparison between active and passive 3D vision sensors: Bumblebeexb3 and Microsoft Kinect. In: First Iberian robotics conference: advances in robotics

Besl PJ, McKay ND (1992) Method for registration of 3-d shapes. In: Sensor fusion IV: control paradigms and data structures, vol 1611, pp 586–606

Biber, P, Straßer W (2003) The normal distributions transform: a new approach to laser scan matching. In: IEEE/RSJ international conference on intelligent robots and systems (IROS) (Cat. No. 03CH37453)

Bosse M, Zlot R (2013) Place recognition using keypoint voting in large 3d lidar datasets. In: IEEE international conference on robotics and automation (ICRA)

Breuer T, Bodensteiner C, Arens M (2014) Low-cost commodity depth sensor comparison and accuracy analysis. In: Electro-optical remote sensing, photonic technologies, and applications VIII; and military applications in hyperspectral imaging and high spatial resolution sensing II, pp 77–86

Cai X, Yin W (2021) Weighted scan context: global descriptor with sparse height feature for loop closure detection. In: International conference on computer, control and robotics (ICCCR)

Calonder M, Lepetit V, Strecha C, Fua P (2010) Brief: binary robust independent elementary features. In: European conference on computer vision (ECCV)

Cao F, Zhuang Y, Zhang H, Wang W (2018) Robust place recognition and loop closing in laser-based SLAM for UGVs in urban environments. IEEE Sens J 18:4242–4252

Cao F, Yan F, Wang S, Zhuang Y, Wang W (2020) Season-invariant and viewpoint-tolerant lidar place recognition in GPS-denied environments. IEEE Trans Ind Electron 68:563–574

Carlevaris-Bianco N, Ushani AK, Eustice RM (2016) University of Michigan north campus long-term vision and lidar dataset. Int J Robot Res 35:1023–1035

Cattaneo D, Vaghi M, Valada A (2022) Lcdnet: deep loop closure detection and point cloud registration for lidar slam. IEEE Trans Robot 38:2074–2093

Chang MY, Yeon S, Ryu S, Lee D (2020) Spoxelnet: spherical voxel-based deep place recognition for 3d point clouds of crowded indoor spaces. In: IEEE/RSJ international conference on intelligent robots and systems (IROS)

Chang A, Dai A, Funkhouser T, Halber M, Niebner M, Savva M, Song S, Zeng A, Zhang Y (2017) Matterport3d: Learning from RGB-D data in indoor environments. In: International conference on 3D vision (3DV)

Chen X, Läbe T, Milioto A, Röhling T, Behley J, Stachniss C (2021) OverlapNet: a siamese network for computing LiDAR scan similarity with applications to loop closing and localization. Auton Robots 46:61–81

Chen X, Läbe T, Milioto A, Röhling T, Vysotska O, Haag A, Behley J, Stachniss C (2020) Overlapnet: loop closing for lidar-based slam. In: Proceedings of robotics: science and systems (RSS), pp 1–10

Chen X, Läbe T, Nardi L, Behley J, Stachniss C (2020) Learning an overlap-based observation model for 3D LiDAR localization. In: Proceedings of the IEEE/RSJ international conference on intelligent robots and systems (IROS)

Cohen T, Welling M (2016) Group equivariant convolutional networks. In: International conference on machine learning (ICML)

Collier J, Se S, Kotamraju V, Jasiobedzki P (2012) Real-time lidar-based place recognition using distinctive shape descriptors. In: Unmanned systems technology XIV, vol 8387, pp 271–281

Cop KP, Borges PV, Dubé R (2018) Delight: an efficient descriptor for global localisation using lidar intensities. In: IEEE international conference on robotics and automation (ICRA)

Cramariuc A, Tschopp F, Alatur N, Benz S, Falck T, Brühlmeier M, Hahn B, Nieto J, Siegwart R (2021) Semsegmap–3D segment-based semantic localization. In: IEEE/RSJ international conference on intelligent robots and systems (IROS)

Cui Y, Chen X, Zhang Y, Dong J, Wu Q, Zhu F (2023) Bow3d: bag of words for real-time loop closing in 3d lidar slam. IEEE Robot Autom Lett 8:2828–2835

Cui Y, Zhang Y, Dong J, Sun H, Chen X, Zhu F (2024) Link3d: linear keypoints representation for 3d lidar point cloud. IEEE Robot Autom Lett. https://doi.org/10.1109/LRA.2024.3354550

Cui J, Chen X (2023) Ccl: continual contrastive learning for lidar place recognition. arXiv:2303.13952

Dai D, Wang J, Chen Z, Bao P (2022) SC-LPR: spatiotemporal context based lidar place recognition. Pattern Recognit Lett 156:160–166

Di Giammarino L, Aloise I, Stachniss C, Grisetti G (2021) Visual place recognition using lidar intensity information. In: IEEE/RSJ international conference on intelligent robots and systems (IROS)

Ding L, Zhou R, Yuan Y, Yang H, Li J, Yu T, Liu C, Wang J, Li S, Gao H et al (2022) A 2-year locomotive exploration and scientific investigation of the lunar farside by the Yutu-2 rover. Sci Robot 7:6660

Dubé R, Gollub MG, Sommer H, Gilitschenski I, Siegwart R, Cadena C, Nieto J (2018) Incremental-segment-based localization in 3-d point clouds. IEEE Robot Autom Lett 3:1832–1839

Dube R, Cramariuc A, Dugas D, Sommer H, Dymczyk M, Nieto J, Siegwart R, Cadena C (2020) SegMap: segment-based mapping and localization using data-driven descriptors. Int J Robot Res 39:339–355

Dubé R, Dugas D, Stumm E, Nieto J, Siegwart R, Cadena C (2017) Segmatch: segment based place recognition in 3D point clouds. In: IEEE international conference on robotics and automation (ICRA)

Du J, Wang R, Cremers D (2020) Dh3d: deep hierarchical 3d descriptors for robust large-scale 6dof relocalization. In: European conference on computer vision (ECCV)

Elhousni M, Huang X (2020) A survey on 3D lidar localization for autonomous vehicles. In: IEEE intelligent vehicles symposium (IV), pp 1879–1884

Endres F, Hess J, Sturm J, Cremers D, Burgard W (2013) 3-D mapping with an RGB-D camera. IEEE Trans Robot 30:177–187

Fan Y, He Y, Tan U-X (2020) Seed: a segmentation-based egocentric 3d point cloud descriptor for loop closure detection. In: IEEE/RSJ international conference on intelligent robots and systems (IROS)

Fankhauser P, Bloesch M, Rodriguez D, Kaestner R, Hutter M, Siegwart R (2015) Kinect v2 for mobile robot navigation: evaluation and modeling. In: International conference on advanced robotics (ICAR), pp 388–394

Fan Z, Liu H, He J, Sun Q, Du X (2020) Srnet: a 3d scene recognition network using static graph and dense semantic fusion. In: Computer graphics forum, vol 39, pp 301–311

Fan Z, Song Z, Liu H, Lu Z, He J, Du X (2022) Svt-net: super light-weight sparse voxel transformer for large scale place recognition. In: Proceedings of the AAAI conference on artificial intelligence

Ferrarini B, Waheed M, Waheed S, Ehsan S, Milford MJ, McDonald-Maier KD (2020) Exploring performance bounds of visual place recognition using extended precision. IEEE Robot Autom Lett 5:1688–1695

Gálvez-López D, Tardos JD (2012) Bags of binary words for fast place recognition in image sequences. IEEE Trans Robot 28:1188–1197

Geiger A, Lenz P, Stiller C, Urtasun R (2013) Vision meets robotics: the Kitti dataset. Int J Robot Res 32:1231–1237

Golla T, Klein R (2015) Real-time point cloud compression. In: IEEE/RSJ international conference on intelligent robots and systems (IROS)

Gong Y, Sun F, Yuan J, Zhu W, Sun Q (2021) A two-level framework for place recognition with 3d lidar based on spatial relation graph. Pattern Recognit 120:108171

Guo J, Borges PV, Park C, Gawel A (2019) Local descriptor for robust place recognition using lidar intensity. IEEE Robot Autom Lett 4:1470–1477

Guo Y, Wang H, Hu Q, Liu H, Liu L, Bennamoun M (2020) Deep learning for 3d point clouds: a survey. IEEE Trans Pattern Anal Mach Intell 43:4338–4364

Habich T-L, Stuede M, Labbé M, Spindeldreier S (2021) Have I been here before? learning to close the loop with lidar data in graph-based slam. In: IEEE/ASME international conference on advanced intelligent mechatronics (AIM)

Han X-F, Feng Z-A, Sun S-J, Xiao G-Q (2023) 3D point cloud descriptors: state-of-the-art. Artif Intell Rev 56:12033–12083

Hess W, Kohler D, Rapp H, Andor D (2016) Real-time loop closure in 2D lidar SLAM. In: IEEE international conference on robotics and automation (ICRA)

He L, Wang X, Zhang H (2016) M2dp: a novel 3D point cloud descriptor and its application in loop closure detection. In: IEEE/RSJ international conference on intelligent robots and systems (IROS)

Hongming S, Qun Z, Hanchen L, Zhang X, Bailing T, Lei H (2022) A distributed approach for lidar-based relative state estimation of multi-UAV in GPS-denied environments. Chin J Aeronaut 35:59–69

Hou Z, Yan Y, Xu C, Kong H (2022) Hitpr: hierarchical transformer for place recognition in point cloud. In: International conference on robotics and automation (ICRA)

Huang T, Liu Y (2019) 3d point cloud geometry compression on deep learning. In: Proceedings of the 27th ACM international conference on multimedia

Hui L, Cheng M, Xie J, Yang J, Cheng M-M (2022) Efficient 3d point cloud feature learning for large-scale place recognition. IEEE Trans Image Process 31:1258–1270

Hui L, Yang H, Cheng M, Xie J, Yang J (2021) Pyramid point cloud transformer for large-scale place recognition. In: Proceedings of the IEEE/CVF international conference on computer vision (ICCV)

Ibrahim M, Akhtar N, Anwar S, Mian A (2023) Unloc: a universal localization method for autonomous vehicles using lidar, radar and/or camera input. arXiv:2307.00741

Ibrahim M, Akhtar N, Anwar S, Wise M, Mian A (2023) Slice transformer and self-supervised learning for 6dof localization in 3d point cloud maps. arXiv:2301.08957

Ibrahim M, Akhtar N, Anwar S, Wise M, Mian A (2023) Perth-WA localization dataset in 3D point cloud maps. IEEE DataPort. https://doi.org/10.21227/s2p2-2e66

Jiang J, Wang J, Wang P, Bao P, Chen Z (2020) Lipmatch: lidar point cloud plane based loop-closure. IEEE Robot Autom Lett 5:6861–6868

Kim G, Park B, Kim A (2019) 1-day learning, 1-year localization: long-term lidar localization using scan context image. IEEE Robot Autom Lett 4:1948–1955

Kim G, Choi S, Kim A (2021) Scan context++: structural place recognition robust to rotation and lateral variations in urban environments. IEEE Trans Robot 38:1856–1874

Kim G, Kim A (2018) Scan context: egocentric spatial descriptor for place recognition within 3D point cloud map. In: IEEE/RSJ international conference on intelligent robots and systems (IROS)

Kim G, Park YS, Cho Y, Jeong J, Kim A (2020) Mulran: multimodal range dataset for urban place recognition. In: IEEE International conference on robotics and automation (ICRA)

Kim G, Yun S, Kim J, Kim A (2022) Sc-lidar-slam: a front-end agnostic versatile lidar slam system. In: International conference on electronics, information, and communication (ICEIC)

Knights J, Moghadam P, Ramezani M, Sridharan S, Fookes C (2022) Incloud: incremental learning for point cloud place recognition. In: IEEE/RSJ international conference on intelligent robots and systems (IROS)

Knights J, Vidanapathirana K, Ramezani M, Sridharan S, Fookes C, Moghadam P (2023) Wild-places: a large-scale dataset for lidar place recognition in unstructured natural environments. In: IEEE international conference on robotics and automation (ICRA), pp 11322–11328

Knott E, Skolnik M (2008) Radar handbook. McGraw-Hill, New York

Kolmet M, Zhou Q, Ošep A, Leal-Taixé L (2022) Text2pos: text-to-point-cloud cross-modal localization. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition (CVPR)

Komorowski J (2021) Minkloc3d: point cloud based large-scale place recognition. In: Proceedings of the IEEE/CVF winter conference on applications of computer vision

Komorowski J (2022) Improving point cloud based place recognition with ranking-based loss and large batch training. In: International conference on pattern recognition (ICPR)

Komorowski J, Wysoczańska M, Trzcinski T (2021) Minkloc++: lidar and monocular image fusion for place recognition. In: International joint conference on neural networks (IJCNN)

Kong X, Yang X, Zhai G, Zhao X, Zeng X, Wang M, Liu Y, Li W, Wen F (2020) Semantic graph based place recognition for 3d point clouds. In: IEEE/RSJ international conference on intelligent robots and systems (IROS)

Kuan YW, Ee NO, Wei LS (2019) Comparative study of intel R200, Kinect v2, and primesense RGB-D sensors performance outdoors. IEEE Sens J 19:8741–8750

Kuang H, Chen X, Guadagnino T, Zimmerman N, Behley J, Stachniss C (2023) IR-MCL: implicit representation-based online global localization. IEEE Robot Autom Lett 8:1627–1634

Labbé M, Michaud F (2019) RTAB-Map as an open-source lidar and visual simultaneous localization and mapping library for large-scale and long-term online operation. J Field Robot 36:416–446

Lai H, Yin P, Scherer S (2022) Adafusion: visual-lidar fusion with adaptive weights for place recognition. IEEE Robot Autom Lett 7:12038–12045

LeCun Y, Bengio Y, Hinton G (2015) Deep learning. Nature 521:436–444

Lee AJ, Song S, Lim H, Lee W, Myung H (2023) $(lc)^2$: lidar-camera loop constraints for cross-modal place recognition. IEEE Robot Autom Lett 8:3589–3596

Li L, Kong X, Zhao X, Huang T, Li W, Wen F, Zhang H, Liu Y (2022) RINet: efficient 3d lidar-based place recognition using rotation invariant neural network. IEEE Robot Autom Lett 7:4321–4328

Li L, Ding W, Wen Y, Liang Y, Liu Y, Wan G (2023) A unified BEV model for joint learning of 3d local features and overlap estimation. arXiv:2302.14511

Li L, Kong X, Zhao X, Huang T, Li W, Wen F, Zhang H, Liu Y (2021) SSC: semantic scan context for large-scale place recognition. In: IEEE/RSJ international conference on intelligent robots and systems (IROS)

Li L, Kong X, Zhao X, Li W, Wen F, Zhang H, Liu Y (2021) Sa-loam: semantic-aided lidar slam with loop closure. In: IEEE international conference on robotics and automation (ICRA)

Lillesand T, Kiefer RW, Chipman J (2015) Remote sensing and image interpretation

Lin J, Zhang F (2019) A fast, complete, point cloud based loop closure for lidar odometry and mapping. arXiv:1909.11811

Li Y, Su P, Cao M, Chen H, Jiang X, Liu Y (2021) Semantic scan context: global semantic descriptor for lidar-based place recognition. In: IEEE international conference on real-time computing and robotics (RCAR)

Liu Z, Suo C, Zhou S, Xu F, Wei H, Chen W, Wang H, Liang X, Liu Y-H (2019) Seqlpd: sequence matching enhanced loop-closure detection based on large-scale point cloud description for self-driving vehicles. In: IEEE/RSJ international conference on intelligent robots and systems (IROS)

Liu Z, Zhou S, Suo C, Yin P, Chen W, Wang H, Li H, Liu Y-H (2019) Lpd-net: 3D point cloud learning for large-scale place recognition and environment analysis. In: Proceedings of the IEEE/CVF international conference on computer vision (ICCV)

Li Y, Zheng S, Yu Z, Yu B, Cao S-Y, Luo L, Shen H-L (2023) I2p-rec: recognizing images on large-scale point cloud maps through bird's eye view projections. arXiv:2303.01043

Lowe DG (2004) Distinctive image features from scale-invariant keypoints. Int J Comput Vis 60:91–110

Lowry S, Sünderhauf N, Newman P, Leonard JJ, Cox D, Corke P, Milford MJ (2015) Visual place recognition: a survey. IEEE Trans Robot 32:1–19

Lun R, Zhao W (2015) A survey of applications and human motion recognition with Microsoft Kinect. Int J Pattern Recognit Artif Intell 29:1555008

Luo L, Cao S-Y, Han B, Shen H-L, Li J (2021) Bvmatch: Lidar-based place recognition using bird's-eye view images. IEEE Robot Autom Lett 6:6076–6083

Luo L, Zheng S, Li Y, Fan Y, Yu B, Cao S-Y, Li J, Shen H-L (2023) Bevplace: learning lidar-based place recognition using bird's eye view images. In: Proceedings of the IEEE/CVF international conference on computer vision (ICCV), pp 8700–8709

Lu Y, Yang F, Chen F, Xie D (2020) Pic-net: point cloud and image collaboration network for large-scale place recognition. arXiv:2008.00658

Ma J, Zhang J, Xu J, Ai R, Gu W, Chen X (2022) OverlapTransformer: an efficient and yaw-angle-invariant transformer network for lidar-based place recognition. IEEE Robot Autom Lett 7:6958–6965

Ma J, Xiong G, Xu J, Chen X (2023) CVTNet: a cross-view transformer network for lidar-based place recognition in autonomous driving environments. IEEE Trans Ind Inform. https://doi.org/10.1109/TII.2023.3313635

Ma J, Chen X, Xu J, Xiong G (2023) SeqOT: a spatial-temporal transformer network for place recognition using sequential lidar data. IEEE Trans Ind Electron 70:8225–8234

Maddern W, Pascoe G, Linegar C, Newman P (2017) 1 year, 1000 km: The oxford Robotcar dataset. Int J Robot Res 36:3–15

Maffra F, Chen Z, Chli M (2018) Tolerant place recognition combining 2d and 3d information for uav navigation. In: IEEE international conference on robotics and automation (ICRA)

Magnusson M, Andreasson H, Nüchter A, Lilienthal AJ (2009) Automatic appearance-based loop detection from three-dimensional laser data using the normal distributions transform. J Field Robot 26:892–914

Magnusson M, Andreasson H, Nuchter A, Lilienthal AJ (2009) Appearance-based loop detection from 3d laser data using the normal distributions transform. In: IEEE international conference on robotics and automation (ICRA)

Masone C, Caputo B (2021) A survey on deep visual place recognition. IEEE Access 9:19516–19547

Minh D, Wang HX, Li YF, Nguyen TN (2022) Explainable artificial intelligence: a comprehensive review. Artif Intell Rev 55:3503–3568

Muhammad N, Lacroix S (2011) Loop closure detection using small-sized signatures from 3d lidar data. In: IEEE International symposium on safety, security, and rescue robotics

Olson E (2009) Recognizing places using spectrally clustered local matches. Robot Auton Syst 57:1157–1172

Pandey G, McBride JR, Eustice RM (2011) Ford campus vision and lidar data set. Int J Robot Res 30:1543–1552

Pan Y, Xu X, Li W, Cui Y, Wang Y, Xiong R (2021) Coral: colored structural representation for bi-modal place recognition. In: IEEE/RSJ international conference on intelligent robots and systems (IROS)

Patel B, Barfoot TD, Schoellig AP (2020) Visual localization with google earth images for robust global pose estimation of uavs. In: IEEE international conference on robotics and automation (ICRA)

Qi CR, Su H, Mo K, Guibas LJ (2017) Pointnet: deep learning on point sets for 3d classification and segmentation. In: Proceedings of the IEEE conference on computer vision and pattern recognition (CVPR)

Qi CR, Yi L, Su H, Guibas LJ (2017) Pointnet++: deep hierarchical feature learning on point sets in a metric space. In: Advances in neural information processing systems, vol 30

Qiao Z, Hu H, Shi W, Chen S, Liu Z, Wang H (2021) A registration-aided domain adaptation network for 3d point cloud based place recognition. In: IEEE/RSJ international conference on intelligent robots and systems (IROS)

Ramezani M, Khosoussi K, Catt G, Moghadam P, Williams J, Borges P, Pauling F, Kottege N (2022) Wildcat: online continuous-time 3d lidar-inertial slam. arXiv:2205.12595

Röhling T, Mack J, Schulz D (2015) A fast histogram-based similarity measure for detecting loop closures in 3-d lidar data. In: IEEE/RSJ international conference on intelligent robots and systems (IROS)

Rublee E, Rabaud V, Konolige K, Bradski G (2011) Orb: an efficient alternative to sift or surf. In: International conference on computer vision (ICCV)

Sánchez-Belenguer C, Ceriani S, Taddei P, Wolfart E, Sequeira V (2020) Global matching of point clouds for scan registration and loop detection. Robot Auton Syst 123:103324

Savva M, Kadian A, Maksymets O, Zhao Y, Wijmans E, Jain B, Straub J, Liu J, Koltun V, Malik J, et al. (2019) Habitat: a platform for embodied AI research. In: Proceedings of the IEEE/CVF international conference on computer vision (ICCV)

Scaramuzza D (2014) Omnidirectional camera. In: Ikeuchi K (eds), Computer Vision: A Reference Guide. ISBN: 978-0-387-30771-8. Springer

Schaupp L, Bürki M, Dubé R, Siegwart R, Cadena C (2019) Oreos: oriented recognition of 3D point clouds in outdoor scenarios. In: IEEE/RSJ international conference on intelligent robots and systems (IROS)

Shan T, Englot B, Duarte F, Ratti C, Rus D (2021) Robust place recognition using an imaging lidar. In: IEEE international conference on robotics and automation (ICRA)

Shi C, Chen X, Huang K, Xiao J, Lu H, Stachniss C (2021) Keypoint matching for point cloud registration using multiplex dynamic graph attention networks. IEEE Robot Autom Lett 6(4):8221–8228

Shi X, Chai Z, Zhou Y, Wu J, Xiong Z (2021) Global place recognition using an improved scan context for lidar-based localization system. In: IEEE/ASME international conference on advanced intelligent mechatronics (AIM)

Shi C, Chen X, Deng W, Lu H, Xiao J, Bin D (2023) RDMNet: reliable dense matching based point cloud registration for autonomous driving. In: IEEE Transactions on intelligent transportation systems

Shi C, Chen X, Xiao J, Dai B, Lu H (2023) Fast and accurate deep loop closing and relocalization for reliable lidar slam. arXiv:2309.08086

Steder B, Grisetti G, Burgard W (2010) Robust place recognition for 3d range data based on point features. In: IEEE international conference on robotics and automation (ICRA)

Steder B, Ruhnke M, Grzonka S, Burgard W (2011) Place recognition in 3d scans using a combination of bag of words and point feature based relative pose estimation. In: IEEE/RSJ international conference on intelligent robots and systems (IROS)

Sun Q, Liu H, He J, Fan Z, Du X (2020) Dagc: employing dual attention and graph convolution for point cloud based place recognition. In: Proceedings of the 2020 international conference on multimedia retrieval

Tang TY, De Martini D, Newman P (2021) Get to the point: Learning lidar place recognition and metric localisation using overhead imagery. In: Proceedings of robotics: science and systems

Tang TY, De Martini D, Wu S, Newman P (2021) Self-supervised learning for using overhead imagery as maps in outdoor range sensor localization. Int J Robot Res 40:1488–1509

Thrun S (2002) Probabilistic robotics. Commun ACM 45:52–57

Tinchev G, Penate-Sanchez A, Fallon M (2019) Learning to see the wood for the trees: deep laser localization in urban and natural environments on a CPU. IEEE Robot Autom Lett 4:1327–1334

Tinchev G, Nobili S, Fallon M (2018) Seeing the wood for the trees: reliable localization in urban and natural environments. In: IEEE/RSJ international conference on intelligent robots and systems (IROS)

Tomono M (2020) Loop detection for 3d lidar slam using segment-group matching. Adv Robot 34:1530–1544

Uy MA, Lee GH (2018) Pointnetvlad: deep point cloud based retrieval for large-scale place recognition. In: Proceedings of the IEEE conference on computer vision and pattern recognition (CVPR)

Vaswani A, Shazeer N, Parmar N, Uszkoreit J, Jones L, Gomez AN, Kaiser Ł, Polosukhin I (2017) Attention is all you need. In: Advances in neural information processing systems, vol 30

Vidanapathirana K, Moghadam P, Harwood B, Zhao M, Sridharan S, Fookes C (2021) Locus: lidar-based place recognition using spatiotemporal higher-order pooling. In: IEEE international conference on robotics and automation (ICRA)

Vidanapathirana K, Ramezani M, Moghadam P, Sridharan S, Fookes C (2022) Logg3d-net: locally guided global descriptor learning for 3d place recognition. In: International conference on robotics and automation (ICRA)

Vosselman G, Maas HG (eds) (2010) Airborne and terrestrial laser scanning

Waikhom L, Patgiri R (2022) A survey of graph neural networks in various learning paradigms: methods, applications, and challenges. Artif Intell Rev 56(7):6295–6364

Wandinger U (2005) In: Weitkamp C (ed), Introduction to lidar, pp 1–18. Springer, New York

Wang Q, Tan Y, Mei Z (2020) Computational methods of acquisition and processing of 3d point cloud data for construction applications. Arch Comput Methods Eng 27:479–499

Wang Z, Shen Y, Cai B, Saleem MT (2019) A brief review on loop closure detection with 3D point cloud. In: IEEE international conference on real-time computing and robotics (RCAR)

Wang Y, Sun Z, Xu C-Z, Sarma SE, Yang J, Kong H (2020) Lidar iris for loop-closure detection. In: IEEE/RSJ international conference on intelligent robots and systems (IROS)

Wang H, Wang C, Xie L (2020) Intensity scan context: coding intensity and geometry relations for loop closure detection. In: IEEE international conference on robotics and automation (ICRA)

Wang W, Zhao W, Wang X, Jin Z, Li Y, Runge T (2019) A low-cost simultaneous localization and mapping algorithm for last-mile indoor delivery. In: International conference on transportation information and safety (ICTIS)

Wasenmüller O, Stricker D (2016) Comparison of Kinect v1 and v2 depth images in terms of accuracy and precision. In: Computer vision–ACCV workshops, Taipei, Taiwan, November 20-24. Revised Selected Papers, Part II 13, pp 34–45

Wiesmann, L, Marcuzzi R, Stachniss C, Behley J (2022) Retriever: point cloud retrieval in compressed 3d maps. In: Proceedings of the IEEE international conference on robotics and automation (ICRA)

Wiesmann L, Milioto A, Chen X, Stachniss C, Behley J (2021) Deep compression for dense point cloud maps. IEEE Robot Autom Lett 6:2060–2067

Wietrzykowski J, Skrzypczyński P (2021) On the descriptive power of lidar intensity images for segment-based loop closing in 3-d slam. In: IEEE/RSJ international conference on intelligent robots and systems (IROS)

Witze A (2020) Nasa has launched the most ambitious mars rover ever built: here's what happens next. Nature 584:15–16

Xiang G, Huang Y, Yu J, Zhu M, Su J (2018) Intelligence evolution for service robot: an ADRC perspective. Control Theory Technol 16:324–335

Xiang H, Shi W, Fan W, Chen P, Bao S, Nie M (2021) Fastlcd: a fast and compact loop closure detection approach using 3d point cloud for indoor mobile mapping. Int J Appl Earth Observ Geoinf 102:102430

Xia Y, Xu Y, Li S, Wang R., Du, J., Cremers, D., Stilla, U.: Soe-net: A self-attention and orientation encoding network for point cloud based place recognition. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition (CVPR)

Xie S, Pan C, Peng Y, Liu K, Ying S (2020) Large-scale place recognition based on camera-lidar fused descriptor. Sensors 20:2870

Xu T-X, Guo Y-C, Lai Y-K, Zhang S-H (2021) Transloc3d: point cloud based large-scale place recognition using adaptive receptive fields. arXiv:2105.11605

Xu Y, Stilla U (2021) Toward building and civil infrastructure reconstruction from point clouds: a review on data and key techniques. IEEE J Select Top Appl Earth Obs Remote Sens 14:2857–2885

Xu X, Yin H, Chen Z, Li Y, Wang Y, Xiong R (2021) Disco: differentiable scan context with orientation. IEEE Robot Autom Lett 6:2791–2798

Ye T, Yan X, Wang S, Li Y, Zhou F (2022) An efficient 3-d point cloud place recognition approach based on feature point extraction and transformer. IEEE Trans Instrum Meas 71:1–9

Yin P, Wang F, Egorov A, Hou J, Jia Z, Han J (2021) Fast sequence-matching enhanced viewpoint-invariant 3-d place recognition. IEEE Trans Ind Electron 69:2127–2135

Yin P, Xu L, Feng Z, Egorov A, Li B (2021) Pse-match: a viewpoint-free place recognition method with parallel semantic embedding. IEEE Trans Intell Transp Syst 23:11249–11260

Yin P, Xu L, Zhang J, Choset H (2021) Fusionvlad: a multi-view deep fusion networks for viewpoint-free 3d place recognition. IEEE Robot Autom Lett 6:2304–2310

Yin H, Xu X, Wang Y, Xiong R (2021) Radar-to-lidar: heterogeneous place recognition via joint learning. Front Robot AI 8:661199

Yin H, Tang L, Ding X, Wang Y, Xiong R (2018) Locnet: global localization in 3d point clouds for mobile vehicles. In: IEEE intelligent vehicles symposium (IV), pp 728–733

Yin P, Wang F, Egorov A, Hou J, Zhang J, Choset H (2020) Seqspherevlad: sequence matching enhanced orientation-invariant place recognition. In: IEEE/RSJ international conference on intelligent robots and systems (IROS)

Yin P, Xu L, Liu Z, Li L, Salman H, He Y, Xu W, Wang H, Choset H (2018) Stabilize an unsupervised feature learning for lidar-based place recognition. In: IEEE/RSJ international conference on intelligent robots and systems (IROS)

Yin H, Xu X, Lu S, Chen X, Xiong R, Shen S, Stachniss C, Wang Y (2023) A survey on global lidar localization. arXiv:2302.07433

Yin P, Zhao S, Cisneros I, Abuduweili A, Huang G, Milford M, Liu C, Choset H, Scherer S (2022) General place recognition survey: towards the real-world autonomy age. arXiv:2209.04497

Yudin D, Solomentsev Y, Musaev R, Staroverov A, Panov AI (2023) Hpointloc: point-based indoor place recognition using synthetic RGB-D images. In: Neural information processing: 29th international conference

Zaffar M, Garg S, Milford M, Kooij J, Flynn D, McDonald-Maier K, Ehsan S (2021) Vpr-bench: an open-source visual place recognition evaluation framework with quantifiable viewpoint and appearance change. Int J Comput Vis 129:2136–2174

Zaganidis A, Zerntev A, Duckett T, Cielniak G (2019) Semantically assisted loop closure in slam using NDT histograms. In: IEEE/RSJ international conference on intelligent robots and systems (IROS)

Zennaro S (2014) Evaluation of Microsoft Kinect 360 and Microsoft Kinect One for robotics and computer vision applications

Zhang X, Wang L, Su Y (2021) Visual place recognition: a survey from deep learning perspective. Pattern Recognit 113:107760

Zhang L, Ghosh BK (2000) Line segment based map building and localization using 2d laser rangefinder. In: IEEE international conference on robotics and automation (ICRA). Symposia Proceedings (Cat. No. 00CH37065)

Zhang W, Xiao C (2019) Pcan: 3d attention map learning using contextual information for point cloud based retrieval. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition (CVPR)

Zhou Z, Zhao C, Adolfsson D, Su S, Gao Y, Duckett T, Sun L (2021) Ndt-transformer: large-scale 3d point cloud localisation using the normal distribution transform representation. In: IEEE international conference on robotics and automation (ICRA)

Zhuang Y, Jiang N, Hu H, Yan F (2012) 3-d-laser-based scene measurement and place recognition for mobile robots in dynamic indoor environments. IEEE Trans Instrum Meas 62:438–450

Zhu Y, Ma Y, Chen L, Liu C, Ye M, Li L (2020) Gosmatch: graph-of-semantics matching for detecting loop closures in 3D lidar data. In: IEEE/RSJ International conference on intelligent robots and systems (IROS)

Zimmerman N, Guadagnino T, Chen X, Behley J, Stachniss C (2023) Long-term localization using semantic cues in floor plan maps. IEEE Robot Autom Lett 8:176–183

Żywanowski K, Banaszczyk A, Nowicki MR, Komorowski J (2021) MinkLoc3D-SI: 3D lidar place recognition with sparse convolutions, spherical coordinates, and intensity. IEEE Robot Autom Lett 7:1079–1086

Żywanowski K, Banaszczyk A, Nowicki MR (2020) Comparison of camera-based and 3d lidar-based place recognition across weather conditions. In: International conference on control, automation, robotics and vision (ICARCV)

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Authors and Affiliations

Kan Luo[1,2] · Hongshan Yu[2] · Xieyuanli Chen[3] · Zhengeng Yang[2,4] · Jingwen Wang[2] · Panfei Cheng[2] · Ajmal Mian[5]

✉ Hongshan Yu
yuhongshancn@hotmail.com

Kan Luo
yscan@hnu.edu.cn

Xieyuanli Chen
xieyuanli.chen@nudt.edu.cn

Zhengeng Yang
yzg050215@163.com

Jingwen Wang
774536095@qq.com

Panfei Cheng
chengpf@hnu.edu.cn

Ajmal Mian
ajmal.mian@uwa.edu.au

[1] Science Teaching and Research Section, Changsha Normal University, Changsha 410100, China

[2] National Engineering Laboratory for Robot Visual Perception and Control Technology, College of Electrical and Information Engineering, Hunan University, Changsha 410082, China

[3] College of Intelligence Science and Technology, National University of Defense Technology, Changsha 410003, China

[4] College of Engineering and Design, Hunan Normal University, Changsha 410081, China

[5] Department of Computer Science, The University of Western Australia, Perth, WA 6009, Australia