PARTIAL 3D OBJECT RETRIEVAL AND COMPLETENESS EVALUATION FOR URBAN STREET SCENE

Yan Guo[1]*, Chenglu Wen*[1]*, Xiaotian Sun*[1]*, Cheng Wang*[1]*, Jonathan Li*[1,2]

1 Fujian Key Laboratory of Sensing and Computing for Smart Cities, Xiamen University, Xiamen, Fujian 361005, China

2 Department of Geography and Environmental Management, University of Waterloo, Waterloo, Ontario N2L 3G1, Canada

clwen@xmu.edu.cn

ABSTRACT

3D objects detected from real-world data are usually incomplete in different degrees. Objects with different degrees of incompleteness should be treated and processed separately. This paper proposes a framework for partial 3D object retrieval and completeness evaluation in an urban street scene based on mobile laser scanning (MLS) point cloud data. The framework consists of three parts. A deep learning method is first used to detect objects from 3D point cloud data. Then, for each detected object, the most similar object in the reference dataset, which contains complete objects, is obtained by a partial 3D shape retrieval method. Last, a completeness evaluation of the detected object is conducted by calculating the completeness index that reflects the integrity of the detected object, and a missing part prediction is given to guide further completion. The proposed framework is validated on the public dataset KITTI and our own point cloud dataset. The experiment includes 3D detection, the partial 3D shape retrieval, and the completeness evaluation. Results show the good performance of the object detection and partial shape retrieval, also a reasonable evaluation of objects completeness.

Index Terms— partial 3D object retrieval, completeness evaluation, mobile laser scanning (MLS), point cloud

1. INTRODUCTION

Nowadays, the development of technologies such as self-driving, intelligent transportation systems (ITS) is heating up. The primary task of these technologies is to percept the road scene. With the increasing importance attached to three-dimensional (3D) data, which nowadays can be easily acquired using range sensors like Light Detection and Ranging (LiDAR), it is possible to percept road environment easily via these 3D point cloud data.

Different from synthetic data, 3D objects detected from the real world data are sometimes incomplete or partial because of occlusion between objects, sensor limitations, or weather conditions. Objects completion can be applied after object detection, to provide better information or augment datasets. However, most of the current 3D shape completion methods complete every fragmentary target without consideration of their completeness [1], which may be applicable when you know their forms before completing them. In some cases, such as in urban street scenes, object completing without consideration sometimes leads to unreasonable results. When large chunks of an object are missing, it is barely possible to recognize what it is without prior knowledge. Suppose that there is a sample with only a few points, which originally belongs to a truck on the street, the completion network directly completes it into a bus according to a pre-trained model, which is unreasonable. So, if the data of an object has very few points, it should be discarded or recollected, but not completed. Thus, it is necessary to develop a completeness evaluation of the detected object to avoid unreasonable completion and provide information for further completion.

This paper proposes a framework for 3D object completeness evaluation. 3D objects are firstly detected from 3D point cloud data of the urban street scene. Then a most similar and complete object for the detected one is retrieved. Finally, the completeness of the object is automatically evaluated both globally and locally. The results of evaluation include metrics which reflect the completeness of the object and a missing part prediction to guide further completion. Completeness evaluation of the 3D object can avoid unreasonable completion results for the detected objects and provide some prior knowledge for completion.

2. METHODOLOGY

The proposed framework contains the following three main parts as (1) 3D objects detection network based on PointRCNN [2]; (2) partial 3D shape retrieval based on SO-Net [3]; and (3) completeness evaluation metrics for detected objects. The architecture of the proposed framework is detailed in Fig. 1.

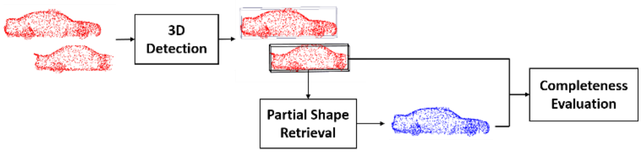

Fig.1. Proposed framework

2.1. 3D object detection using PointRCNN

We first detect 3D objects from the point cloud using PointRCNN [2] and evaluate their completeness. PointRCNN is a 3D object detection network by using only point cloud as input. Instead of generating proposals from RGB images, PointRCNN directly generates 3D proposals from the point cloud in a bottom-up manner via segmentation the point cloud of the whole scene into foreground points and background. After generating the proposals, a box regression is appended to regress 3D bounding box locations and refining the box locations and orientation based on the proposals by region pooling. Fig.2 is a schematic diagram of 3D objects detection from a single frame of point cloud data.

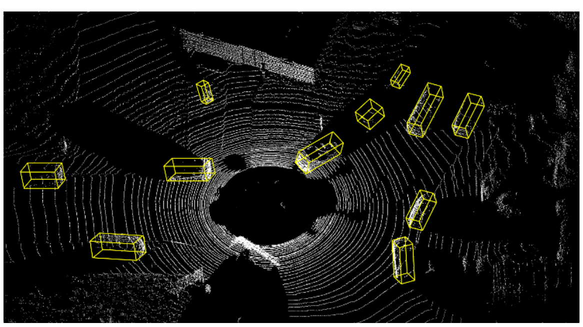

Fig.2. Schematic diagram of 3D objects detection from point clouds

2.2. Partial 3D shape retrieval

SO-Net [3] is a permutation invariant architecture for deep learning on unordered point clouds, it models the spatial distribution of point cloud by building a Self-Organizing Map (SOM), which is trained unsupervised and competitively to produce a low-dimensional, discretized representation of the input space. Point clouds are converted into SOM node features and a global feature that can be applied to shape retrieval. First, SO-Net constructs a SOM point set with the size of $M \times M$, and updates SOM while training to simulate the distribution of the input point cloud, i.e. Fig 3. After given the output of the SOM, SO-Net search for the k nearest neighbors(kNN) on the SOM nodes S for each point p_i. Then, each p_i is normalized into k points by subtraction with its associated nodes. The resulting kN normalized points are forwarded into a series of fully connected layers to extract individual point features.

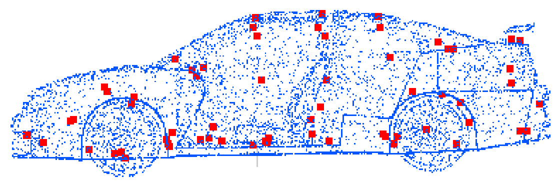

Fig.3. Example of a SOM training result for a car.

For the retrieval task, given a query shape and a shape library, the similarity between the query and the candidates can be computed as their feature vector distances.

2.3. Completeness evaluation

The completeness of a detected object is evaluated by calculating the distance between its corresponding retrieval sample and itself (distance between two point sets). Two metrics, retrieval recall, and Chamfer Distance are used to evaluate the object completeness as below.

2.3.1. Retrieval recall

Retrieval recall is a measure of relevance in an information retrieval scenario and could be transformed into our method as a metric to represent the similarity between the detected object and the complete sample [4]. Let $\boldsymbol{S}$ be the complete sample point set, and $\boldsymbol{P}$ be the detected point set being evaluated. For a complete sample point $\boldsymbol{s} \in \boldsymbol{S}$, its distance to the detected set $\boldsymbol{P}$ is defined as:

$$d_{s \to P} = \min_{s \in S} |s - p| \quad (1)$$

Also, the retrieval recall r_d of the detected point set for a threshold d_t is defined as:

$$r_d = \frac{1}{|S|} \sum_{s \in S} [d_{s \to P} < d_t] \quad (2)$$

where $r_d \in [0,1]$ and can be regarded as a percentum. A threshold value $\boldsymbol{R}$ is set manually to determine whether the object should be completed or just be dropped. If $r_d < R$, it indicates this incomplete object does not have enough information to be completed, it is supposed to be dropped or recollected. If $r_d > R$, it indicates this object could be completed.

2.3.2. 3D Chamfer Distance

Chamfer Distance (CD) is a function to calculate the distance between two sets [5]. We extend it into 3D as a completeness metric. The 3D CD metric, d_{CD}, is defined as:

$$d_{CD}(S,P) = \sum_{s\in S} \min_{p\in P} |s-p|^2 + \sum_{p\in P} \min_{s\in S} |s-p|^2 \quad (3)$$

For each point, the calculation of CD finds the nearest neighbor in the other set and sums the squared distances up, which intuitively reflects the similarity between two sets. A threshold $\boldsymbol{D}$ is also set to help make a decision. Different from $\boldsymbol{R}$ in retrieval recall, $\boldsymbol{D}$ is set depending on the unit of measure in a point set.

2.3.3 Local integrity evaluation

To individually evaluate the integrity of each incomplete sub-box, we divide the bounding box of an object into eight sub-boxes uniformly to get a prediction of the missing part's position. In summary, the output of completeness evaluation consists of the following two parts: (1) value metrics of completeness that determine if this detected sample is usable and worth completing; and (2) a position prediction of missing part.

3. RESULTS AND DISCUSSIONS

3.1 Experiments

We first trained and validated the PointRCNN [2] on the 3D object detection benchmark of KITTI dataset [6], these results provide incomplete 3D objects detected from a real-world urban street scene. Then, retrieve a complete object for each detected object by partial 3D shape retrieval based on SO-Net. Finally, we evaluated the object completeness by calculating and comparing both retrieval recall and CD. Besides, instructed by the evaluation results, we completed some 3D objects by a completion network [1] to verify the necessity of the completeness evaluation before object completion.

3.2 Results and conclusion

3.2.1 Partial 3D object retrieval results

The visualization results of the retrieval task are shown in Figure 4. The incomplete blue cars in the top are the queries, and the red cars below are the corresponding top three of retrieval return. It can be seen intuitively that the complete models from the retrieval task is consistent with the incomplete model to a certain extent and can be seen as a complete model reference for the defective sample.

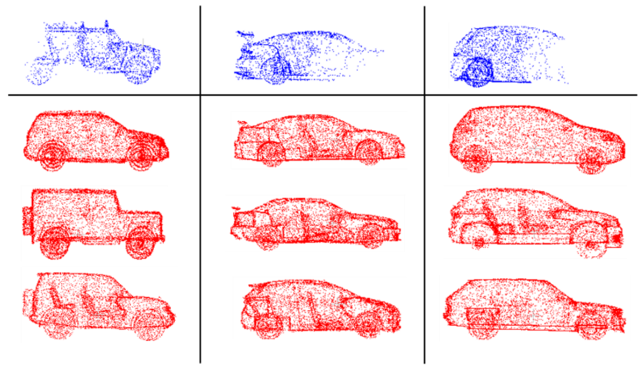

Fig.4. Qualitative result of a partial 3D object retrieval. Blue: query. Red: top 3 retrieved shapes ordered by feature similarity.

3.2.2 Completion evaluated results

Some examples of incomplete objects being evaluated are shown in Fig.5. The blue object is the complete sample. The middle row of red samples are incomplete objects being evaluated, and the objects in the top row are the corresponding ground truth. The bottom row shows from a different perspective view to better demonstrate the missing part of the objects. Some results of completeness evaluation are given in Table.1.

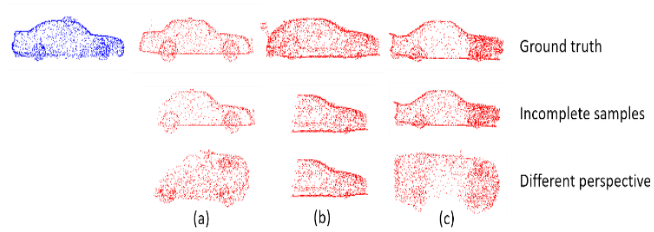

Fig.5. Illustration of three incomplete objects (a), (b), (c) and a complete sample (blue).

Metrics	(a)	(b)	(c)
Recall	0.591308	0.538574	0.874023
CD	18.198436	35.529631	2.560999

Table.1. The retrieval recall and CD of the objects in Fig.5.

As shown in Table.1, for objects with different degree of incompleteness, the results of retrieval recall and CD are different. For recall, the higher value means the detected object is with higher completeness level. As for CD, the value shows the distance between two point sets (detected object point set and complete reference point set), so the lower value means the detected object is with higher completeness level. Compare to CD, recall is more intuitive to evaluate the completeness of an object because it can be regarded as a percentage. CD has a limit in the multi-scale presentation of an object. Some samples with different range of r_d are collected (Fig.6). With different r_d, the completeness of an object is different. For those incomplete samples with a very low r_d (for example smaller than 0.3), it indicates they are not suitable to be completed.

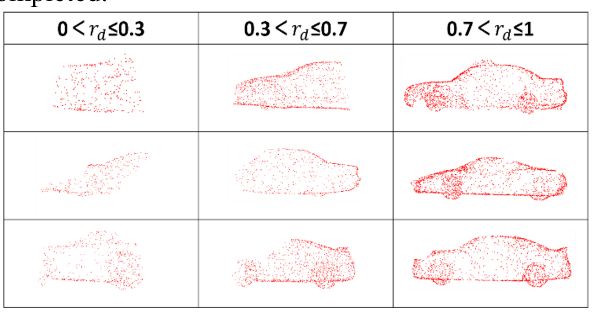

Fig.6. Object completeness at different values of r_d .

REFERENCES

[1] Stutz, David, and Andreas Geiger. "Learning 3D Shape Completion from Laser Scan Data with Weak Supervision." *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 2018.

[2] Shi, Shaoshuai, Xiaogang Wang, and Hongsheng Li. "PointRCNN: 3D Object Proposal Generation and Detection from Point Cloud." *arXiv preprint arXiv:*1812.04244, 2018.

[3] Li, Jiaxin, Ben M. Chen, and Gim Hee Lee. "So-net: Self-organizing network for point cloud analysis." *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2018.

[4] Knapitsch, Arno, et al. "Tanks and temples: Benchmarking large-scale scene reconstruction." *ACM Transactions on Graphics (ToG)* 36.4 (2017): 78.

[5] Fan, Haoqiang, Hao Su, and Leonidas Guibas. "A point set generation network for 3d object reconstruction from a single image." *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. IEEE, 2017.

[6] Geiger, Andreas, et al. "Vision meets robotics: The KITTI dataset." *The International Journal of Robotics Research* 32.11 (2013): 1231-1237.