Proceedings of CECNet 2022
A.J. Tallón-Ballesteros (Ed.)

doi:10.3233/FAIA220566

3D Point Cloud and BIM Component Retrieval for Subway Stations via Deep Learning

Yongfei SHI [a], Wei YE [c], Bo QU [a], Hong JIA [c], Zhanying FU [a], Xinming LV [b], Chenglu WEN [c] and Weiquan LIU [c,1]

[a] *Xiamen Rail Transit Group Co., Ltd., Xiamen, China*
[b] *Xiamen Metro Technology Innovation Technology Co., Ltd., Xiamen, China*
[c] *Fujian Key Laboratory of Sensing and Computing for Smart Cities, School of Informatics, Xiamen University, Xiamen, China*

Abstract. It is urgent to digitize the subway equipment to detect the changes in the components in the subway station in time. In this paper, we use the 3D point cloud of the subway station as a benchmark, compare it with the completed BIM model of the subway station, find out the changes in the components of the subway station, and retrieve 3D point cloud and BIM model components. First, we obtained 3D point cloud and BIM model components in Xiamen Metro Station. Second, we labeled 140 pairs of matched 3D point cloud components and BIM model component point cloud. Third, we constructed a Siamese network which embedded triplet loss to learn the feature descriptors of 3D point cloud and BIM model components, and then retrieve 3D point cloud components to BIM model components. Experimental results show that our proposed method realize the retrieval of 3D point cloud components to BIM model components in the subway station environment.

Keywords. 3D point cloud, BIM, component retrieval, subway station

1. Introduction

Urban rail transit is the backbone of the urban public transportation system, with a huge investment scale, huge construction projects, and complex operation and management, which puts forward higher requirements for the digital construction of urban rail transit information. The proper management of data is a key business enabler and booster for companies [1]. The Building Information Modeling (BIM) methodology was proposed to unify projects around a Digital Twin of the information necessary for collaboration [2]. BIM model is the core data in the digital construction of urban rail transit informatization. It runs through the whole construction project management of feasibility analysis, investment decision-making, design, construction and completion [3]. However, BIM models cannot display the dynamic change information of the building.

Traditional methods usually use manual methods to manage subway equipment. In addition, manual management of subway equipment requires a lot of manpower and time,

[1]Corresponding Author: Weiquan LIU, E-mail: wqliu@xmu.edu.cn.

and it is difficult to form digital subway management. However, there is a problem of inconsistent data representation between 3D point cloud and BIM model. It is necessary to unify the 3D point cloud and BIM model into the same data format. In this paper, we regard the data of BIM model as mesh data, and adopt the method of uniform surface sampling to convert the BIM model into the form of 3D point cloud data. Therefore, the retrieval problem of 3D point cloud and BIM model components becomes the retrieval problem of 3D point cloud components and BIM model components point cloud.

2. Research Objective

In this paper, we aim at the problem of digital management at the equipment component level of the subway station. First, we compare the difference between the 3D point cloud of subway station and the 3D point cloud data generated by the subway BIM model, and detect the areas of difference. Second, we construct a Siamese network which embedded triplet loss, to learn the consistent feature descriptors of the 3D point cloud and BIM model components. Experiments show that our proposed method realizes the retrieval of 3D point cloud components to BIM model components in the subway station.

The contributions of this paper are summarized as follows:

(1) We propose a Siamese network to jointly learn cross-domain descriptors of 3D point cloud components and BIM model components in the subway station environment. Meanwhile, the triplet loss is embedded to balance the consistency of the 3D point cloud and BIM model component feature descriptors.

(2) The local cross-domain feature descriptor learned by the proposed Siamese network has been applied in 3D point cloud components and BIM model retrieval, and has been demonstrated in the subway station environment.

3. Related Work

3.1. 3D feature descriptors

3D feature descriptors can be divided into handcrafted and deep learning based methods. Handcrafted 3D feature descriptors defined by geometric relationships between points, such as PFH [4] which describe the k-neighborhood geometric properties of points by parameterizing the spatial differences between query points and neighbor points and forming a multidimensional histogram, FPFH [5] which calculates the simplified point feature histogram (SPFH) of each point in the k-neighborhood of the query point separately, and then weights all SPFHs into the final fast point feature histogram through a formula.

Deep learning based methods use raw point clouds as input and output local 3D feature descriptorssuch as PointNet [6], PointNet++ [7], Paconv [8] and TopoSeg [9]. Some deep learning neural networks perform very well in extracting 3D feature descriptors, such as PerfectMatch [10] and D3Feat [11]. Furthermore, methods for extracting 3D feature descriptions have emerged from new perspectives. PVRCNN [12] combines the advantages of point-based and voxel-based methods to extract 3D feature descriptors more effectively and improve the performance of 3D object detection. PCTNet [13] uses Transformer [14] to encode the input point cloud into a new high-dimensional feature space.

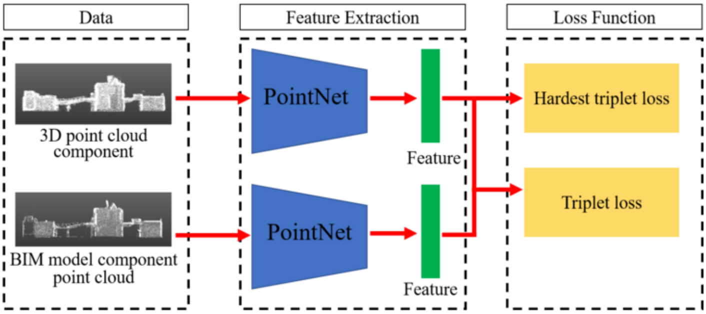

Figure 1. The constructed network structure of our method.

3.2. Deep similarity learning networks

Siamese network is commonly used for deep similarity learning networks, it uses two neural networks with shared weights to learn feature descriptors. The two neural networks convert the inputs into a vector respectively, the similarity is obtained by judging the distance. Metric learning networks achieve good performance by choosing to learn a nonlinear function to measure the similarity between feature descriptors, but require a lot of time and cost to calculate the similarity between feature descriptors. There are some Siamese networks for image patch matching based on feature retrieval such as MatchNet [15], SiamAM-Net [16], H-Net [17], Y-Net [18] and 2D3D-MVPNet [19].

4. Method

4.1. Network Structure

The constructed Siamese network structure is a two-tower structure, as shown in Figure 1. Each branch receives one type of data, resulting in a common feature descriptor. Subway 3D point cloud and BIM model point cloud belong to different domains of data, and it is difficult to match the two on the original data. Therefore, it is more reasonable to use feature descriptors based on different data for data matching. We use the Siamese network to process data from different sources and extract the cross-domain feature descriptors for 3D point cloud and BIM model components point cloud.

Each branch of the Siamese network adopts the Pointnet [6] structure. The parameters of the two branches are independent, which enables the branches to adapt to different data feature extraction. The structural parameters of Pointnet are: $conv1d(1,64,1) - conv1d(1,64,1) - conv1d(1,64,1) - conv1d(1,128,1) - conv1d(1,1024,1) - maxpool - fc(1024,128)$, where $conv1d(k,d,s)$ is a one-dimensional convolution with convolution kernel k, output dimension d, and stride s, maxpool represents the maximum pooling operation, $fc(p,q)$ represents a fully connected layer with input dimension p and output dimension q. Based on this structure, each branch receives a point cloud containing 8192 points and output a common feature descriptor with 128-dimensional.

4.2. Loss Function

To ensure that the matching network can distinguish matching samples from non-matching samples, we construct positive samples and negative samples for network training. Positive samples and negative samples constitute the constraints of the loss function. In this paper, triplet loss is used as a constraint, as shown in the Eq. 1, where p and c are the feature descriptors obtained by the subway 3D point cloud component P and the BIM component point cloud C after passing through the twin network, and $d(p,c)$ is the feature descriptor between the feature descriptors.

$$L(p,c) = \sum_{i=1,n} \max\left\{0, 0.2 + d(p,c) - d\left(p, c_{negative}\right)\right\} \tag{1}$$

We uses two methods to construct negative samples at the same time, namely, random sampling method and hardest sample method to construct negative samples. The random sampling method is to randomly sample a BIM model component in the range other than the matched BIM model component, in each pair of matching pair data of subway 3D point cloud components. Then the 3D point cloud component and the BIM model component form as a negative sample. The negative samples and positive samples obtained by this sampling method are substituted into triplet loss.

Difficult sample method, which finds samples that are indistinguishable by the network as negative samples in the batch of positive samples, thereby improving the network's ability to distinguish difficult samples. The specific process, for the training process with the number of batch samples n, first, based on n pairs of cross-dimensional feature vectors, construct an $n \times n$ distance matrix D, which is expressed as follows:

$$D = \begin{pmatrix} d_{1,1} & \cdots & d_{1,n} \\ \vdots & \ddots & \vdots \\ d_{n,1} & \cdots & d_{n,n} \end{pmatrix} \tag{2}$$

Second, the negative sample construction is performed. The sample sampling strategy selects the second nearest neighbor as a negative sample. For example, the positive sample (p_1, c_1) constructs the negative sample process: Assume that $d(p_2, c_1)$ and $d(p_1, c_4)$ in the distance matrix D are the minimum distances outside the positive sample; Compare $d(p_1, c_1)$ and $d(p_1, c_4)$, select the minimum value as the distance similarity corresponding to the negative sample, assuming $d(p_1, c_4) > d(p_2, c_1)$, then select p_2 as the negative sample of the sample (p_1, c_1).

Therefore, the triplet loss can be written as:

$$L(p,c) = \frac{1}{n} \sum_{i=1,n} \max\left\{0, 0.2 + d\left(p_i, c_i\right) - \min\left[d\left(p_i, c_{j_{\min}}\right), d\left(p_{k_{\min}}, c_i\right)\right]\right\} \tag{3}$$

4.3. Training strategy

The training process constructs positive samples and negative samples at the same time, and optimizes the network parameters through the ternary loss proposed in Section 4.2. During the training process, the batch size of the sample is 8, and for each pair of positive samples of the 3D point cloud components, a total of two negative samples are obtained through the random sampling and difficult sample sampling methods proposed in Section 4.2, and are used in triplet loss.

The experiments are performed on a hardware device with NVIDIA RTX3090 GPU with 24GB memory. The learning parameters are set to 100 epochs, the optimizer SGD is the optimizer, the momentum parameter is 0.9, the learning rate is 0.001, and it is reduced by 0.05

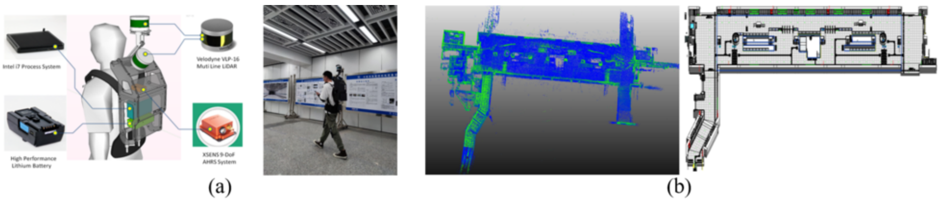

Figure 2. The backpack mobile LiDAR scanning equipment.

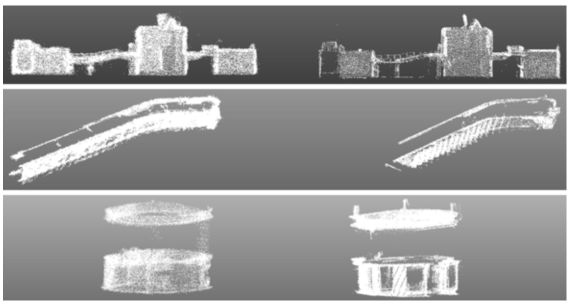

Figure 3. The matching 3D point cloud component and BIM model component point cloud. Left: 3D point cloud components. Right: BIM model component point cloud .

5. Experiments

5.1. Dataset

We use the backpack mobile LiDAR scanning equipment developed by the SCSC Laboratory of Xiamen University to collect 3D point cloud of subway stations, as shown in Figure 2 (a). We collected the 3D point cloud and the corresponding BIM model from Houcun Station, Shuangshi Station, and Huojuyuan Station of Xiamen Metro Line 3. Figure 2 (b) shows the 3D point cloud and the corresponding BIM model of Houcun Station of Xiamen Metro Line 3.

Specifically, we generated 3D point cloud from BIM model components of subway stations. Based on the acquired 3D point cloud and point cloud of BIM model, 140 pairs of matching 3D point cloud components and BIM model component point cloud were obtained by manual annotation. Figure 3 shows three pairs of matching 3D point cloud components and BIM model component point cloud. In this paper, the 140 pairs of matching samples are used for the experiment, among which 126 pairs of samples are used as the training set of deep learning neural network, and 14 pairs of samples are used as the test set. The training set and test set are guaranteed not to cross each other.

5.2. Difference detection between 3D point cloud and BIM model of subway station

We convert the BIM model into 3D point cloud data and perform registration processing with the real-time 3D point cloud of the subway station. We perform the same down-sampling process on the point could of the BIM model and the 3D point cloud of the subway station, so that the minimum distance between all points in the 3D point cloud

Figure 4. Left: 3D point cloud of subway station. Right: 3D point cloud generated by BIM model.

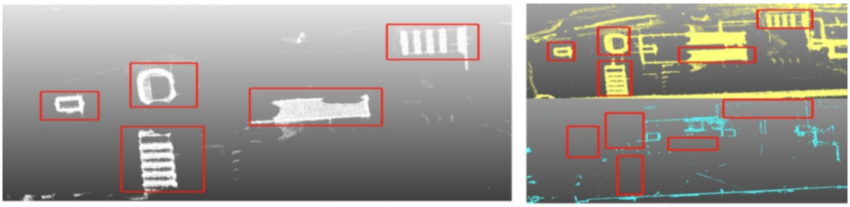

Figure 5. Left: The results of difference detection between 3D point cloud and BIM model of subway station. Upper right: 3D point cloud. Lower right: 3D point cloud generated by BIM model.

data is $2cm$. After down-sampling processing, the ceiling and ground of the completed BIM model point cloud and the 3D point cloud are cut at the same time, and their internal component structure information is retained, so that the point cloud segmentation difference comparison algorithm [20] in PCL can detect and compare the areas of difference.

After processing the 3D point cloud of Xiamen Metro Line 3 Houcun Station Hall and the BIM model point cloud of Xiamen Metro Line 3 Houcun Station Hall, the results are shown in Figure 4. Using point cloud segmentation difference comparison algorithm to compare the point clouds above, and the detected difference area is shown in Figure 5.

5.3. 3D Point Cloud - BIM Component Retrieval

We use TOP1 retrieval accuracy and TOP5 retrieval accuracy as evaluation indexes to measure the retrieval results of 3D point cloud components and BIM model component point cloud. Among them, the TOP1 retrieval accuracy is defined as the proportion of the number of correct matches in the first result of retrieval confidence ranking in all queries, and the TOP5 retrieval accuracy is defined as the proportion of the number of correct matches in the TOP5 results of retrieval confidence ranking in all queries.

The TOP1 retrieval accuracy and TOP5 retrieval accuracy are $Accuracy_{TOP1} = \frac{k_1}{n} * 100\%$ and $Accuracy_{TOP5} = \frac{k_2}{n} * 100\%$, where, k_1 is the number of correct matches in the TOP1 retrieved confidence ranking result, k_2 is the number of correct matches in the TOP5 retrieved confidence ranking result, and n is the total number of queries.

Using our method, the final TOP1 retrieval accuracy in the test set is 78.57%, and the TOP5 retrieval accuracy is 92.85%. Figure 6 shows the visualization results of TOP5 retrieval of 3D point cloud to BIM model point cloud components. The BIM model point cloud components that are correctly retrieved are marked by red boxes.

From the experimental retrieval results in Figure 6, it can be found that our method successfully realizes the retrieval of 3D point cloud components to BIM model component point cloud. The first to third rows of Figure 6 show the correct TOP1 retrieval

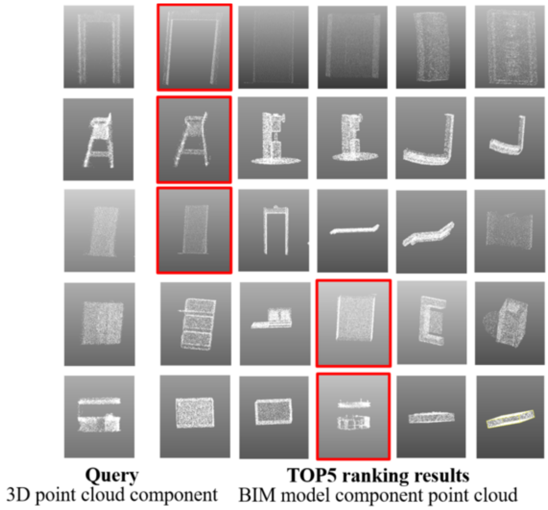

Figure 6. TOP5 ranking results. The queries are 3D point cloud components, and the ground truths and correct retrieval results of BIM model point cloud components are labeled with the red bounding boxes.

results, while the fourth and fifth rows show the wrong TOP1 retrieval results, but the correct TOP5 retrieval results. Although the fourth and fifth rows of Figure 6 are wrong TOP1 retrieval results, from the visualization results, we can find that the shape structure of BIM model component point cloud retrieved by our method is still similar to the query 3D point cloud components. In all the retrieval results, the shape and structure of BIM model component point cloud are similar to the query 3D point cloud components, which demonstrates the robustness of our proposed method.

6. Conclusion

In this paper, aiming at the problem of equipment parts-level digital management of subway stations, we compared the real-time acquired 3D point cloud of subway stations with the BIM model of subway stations to detect the area of change, and then detect the parts of difference, so as to update the BIM model components. To achieve this goal, we construct a Siamese network to extract consistent feature descriptors based on the obtained 3D point cloud components and BIM model components of subway stations. The experimental results demonstrate that the proposed method realizes the retrieval of 3D point cloud components to BIM model components in the subway station environment, and provide digital support for the digital management of equipment components in subway stations.

Acknowledgement
This work is supported by the China Postdoctoral Science Foundation (No.2021M690 094); the FuXiaQuan National Independent Innovation Demonstration Zone Collaborative Innovation Platform (No.3502ZCQXT2021003) and the open fund of PDL (WDZC20215250113).

References

[1] Bifulco I, Cirillo S, Esposito C, et al. An intelligent system for focused crawling from Big Data sources[J]. Expert Systems with Applications, 2021, 184: 115560.
[2] Schiavi B, Havard V, Beddiar K, et al. BIM data flow architecture with AR/VR technologies: Use cases in architecture, engineering and construction[J]. Automation in Construction, 2022, 134: 104054.
[3] Chen B, Jiang S, Qi L, et al. Design and implementation of quantity calculation method based on BIM data[J]. Sustainability, 2022, 14(13): 7797.
[4] Rusu R B, Blodow N, Marton Z C, et al. Aligning point cloud views using persistent feature histograms[C]//IEEE/RSJ International Conference on Intelligent Robots and Systems. 2008: 3384-3391.
[5] Rusu R B, Bradski G, Thibaux R, et al. Fast 3d recognition and pose using the viewpoint feature histogram[C]//IEEE/RSJ International Conference on Intelligent Robots and Systems. 2010: 2155-2162.
[6] Qi C R, Su H, Mo K, et al. Pointnet: Deep learning on point sets for 3d classification and segmentation[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern recognition (CVPR). 2017: 652-660.
[7] Qi C R, Yi L, Su H, et al. Pointnet++: Deep hierarchical feature learning on point sets in a metric space[J]. Advances in Neural Information Processing Systems (NeurIPS), 2017, 30: 5099-5108.
[8] Xu M, Ding R, Zhao H, et al. Paconv: Position adaptive convolution with dynamic kernel assembling on point clouds[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). 2021: 3173-3182.
[9] Liu W, Guo H, Zhang W, et al. TopoSeg: Topology-aware Segmentation for Point Clouds[C]//Proceedings of the International Joint Conference on Artificial Intelligence (IJCAI). 2022: 1201-1208.
[10] Gojcic Z, Zhou C, Wegner J D, et al. The perfect match: 3d point cloud matching with smoothed densities[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). 2019: 5545-5554.
[11] Bai X, Luo Z, Zhou L, et al. D3feat: Joint learning of dense detection and description of 3d local features[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). 2020: 6359-6367.
[12] Shi S, Guo C, Jiang L, et al. Pv-rcnn: Point-voxel feature set abstraction for 3d object detection[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). 2020: 10529-10538.
[13] Guo M H, Cai J X, Liu Z N, et al. Pct: Point cloud transformer[J]. Computational Visual Media, 2021, 7(2): 187-199.
[14] Vaswani A, Shazeer N, Parmar N, et al. Attention is all you need[J]. Advances in Neural Information Processing Systems (NeurIPS), 2017, 30.
[15] Han X, Leung T, Jia Y, et al. Matchnet: Unifying feature and metric learning for patch-based matching[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR). 2015: 3279-3286.
[16] Liu W, Wang C, Bian X, et al. Learning to match ground camera image and uav 3-d model-rendered image based on siamese network with attention mechanism[J]. IEEE Geoscience and Remote Sensing Letters, 2019, 17(9): 1608-1612.
[17] Liu W, Shen X, Wang C, et al. H-Net: Neural Network for Cross-domain Image Patch Matching[C]//Proceedings of the International Joint Conference on Artificial Intelligence (IJCAI). 2018: 856-863.
[18] Liu W, Wang C, Chen S, et al. Y-Net: Learning Domain Robust Feature Representation for ground camera image and large-scale image-based point cloud registration[J]. Information Sciences, 2021, 581: 655-677.
[19] Lai B, Liu W, Wang C, et al. 2D3D-MVPNet: Learning cross-domain feature descriptors for 2D-3D matching based on multi-view projections of point clouds[J]. Applied Intelligence, 2022: 1-16.
[20] Rusu R B, Cousins S. 3d is here: Point cloud library (pcl)[C]//Proceedings of the IEEE International Conference on Robotics and Automation. 2011: 1-4.