Recognition of wave-dissipating blocks from large-scale measured point cloud using deep learning

○Yajun XU, Satoshi KANAI, Hiroaki DATE (Hokkaido University)
Tomoaki SANO, Takuya TERANISHI (Alpha Hydraulic Engineering Consultants, Co.Ltd.)

Abstract

A method to recognize 6D poses of wave-dissipating blocks from 3D point clouds measured by photogrammetry is proposed in the paper. The model-based object recognition approach was adopted. To ensure the accuracy, the recognition process was designed to be composed of the deep-learning-based instance segmentation and descriptor-based 6D pose estimation. Fast Point Cloud Cluster (FPCC) was originally proposed for the instance segmentation. The performance of the recognition was verified and discussed.

1. Introduction

In ports and harbors, many concrete wave-dissipating blocks are installed to prevent erosions caused by the waves. As time goes by, the blocks gradually sink and even ablate. Therefore, as shown in **Fig. 1**, periodical repair works need to be carried out where new supplemental blocks are stacked on top of the existing blocks until the planned raising height. Before the works, constructors must estimate the number of blocks to be supplemented as precisely as possible. Recently, the constructors can easily capture the surface of existing blocks as 3D point clouds using drone-based photogrammetry and narrow-multi-beam(NMB) sonar. However, because the individual 6D poses of the existing blocks are not precisely identified in the point clouds currently, the number of blocks needed cannot be accurately estimated.

To solve the issue, we propose a method to recognize 6D stack-up poses of individual wave-dissipating blocks from measured 3D point clouds. Because the nominal block shape is identical and given as a CAD model, the model-based object recognition approach was adopted. To ensure the accuracy and efficiency, the recognition process was composed of the deep-learning-based instance segmentation and descriptor-based 6D pose estimation. The performance of the recognition was verified quantitatively and qualitatively.

2. Problem Statement and Approach

Fig. 2 illustrates two problems to be solved in our study. First, the block stack-up poses in the existing upper level should be recognized from large-scale dense point clouds measured by photogrammetry and NMB sonar. Second, plausible 3D stack-up poses and the number of supplemental blocks between the existing upper level and planned raising height should be estimated with non-overlap constraint. This report focuses only on the first problem, i.e., the recognition of existing block stack-up poses from point clouds measured by photogrammetry.

In view of the huge advantages of convolutional neural networks (CNNs) in segmentation [1] and poor performance in pose estimation [2], the recognition process of existing block stack-up poses is divided into two stages as shown in **Fig. 3**. First, in the block instance segmentation, the point cloud of a single block is segmented from the whole point clouds by CNNs. Second, in the block pose estimation, the 6D pose of an individual block is estimated from each segmented point cloud using the classical descriptor-based algorithms (Point Pair Feature (PPF) [3] and ICP [4]) and a CAD model of the block.

3. Block Pose Estimation Process

3.1 Block instance segmentation by FPCC

Fast Point Cloud Cluster Net (FPCC-Net), a type of CNN, is newly proposed for the block instance segmentation in this study. FPCC-Net has a modified structure of the CNN for instance segmentation [1]. As shown in **Fig.4,** FPCC-Net is composed of a

Fig. 1 Block stacking work and the measured point clouds

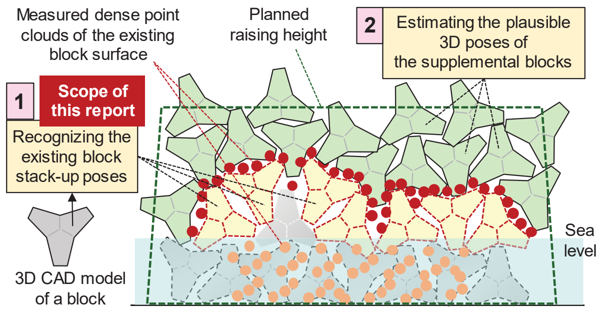


Fig. 2 The problems to be solved in this study

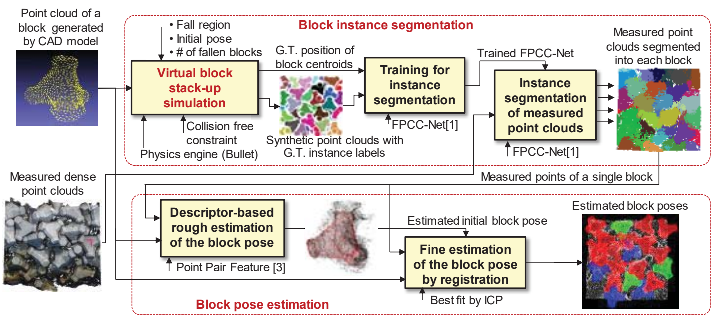


Fig. 3 The pipeline of the proposed block pose estimation

point-wise feature extractor and two branches. The point-wise feature extractor is a CNN that has the semantic segmentation structure of DGCNN [5]. The extracted features are sent to the point-wise embedded feature branch and center score branch, respectively. And the points in the same instance have similar features, while points in the different instances have relatively different features. The center score branch predicts the probability that each point is placed at the center of the object.

In the inference phase, non-maximum suppression is used to find the point with the highest score for each target object as a reference point for clustering. Then the distance between the remaining points and the reference point were calculated. The remaining points are clustered with the nearest center point in terms of the feature distance.

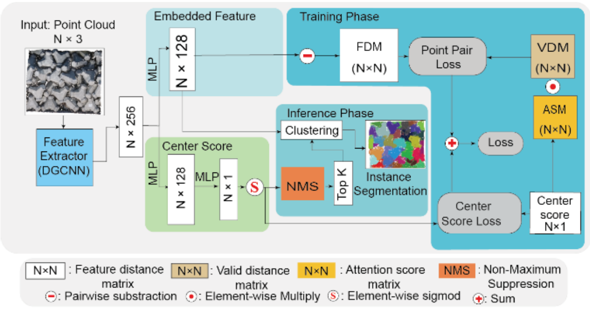


Fig. 4 Network architecture of FPCC-Net

3.2 Training by synthetic point cloud scenes

CNNs should usually be trained by a bunch of manually - segmented point cloud scenes of actual block stack-up scenes. However, it is practically impossible to prepare large amount of high-accuracy segmented point clouds of the actual scenes. To address the issue, many different collision-free block stack-up scenes are virtually generated in a computer simulation using a block CAD model and Bullet engine [6]. In each scene, as shown in **Fig. 5,** 50 blocks free-fell at random locations in a 10m by 10m region, and the synthetic point clouds with instance labels are sampled on top of the stack-up blocks. Our training data set consists of the synthetic point clouds of 500 scenes.

3.3 Block pose estimation by PPF and ICP

For each instance after the segmentation, if the number of point clouds in it is greater than 3000, it is considered as a candidate for pose estimation. As shown in **Fig. 3**, PPF descriptor [3] are evaluated at every point both of the scene point cloud and sampled points on the block CAD model, and the initial 6D pose of the block is estimated. Then, ICP [4] is performed to refine the pose.

4 Case study

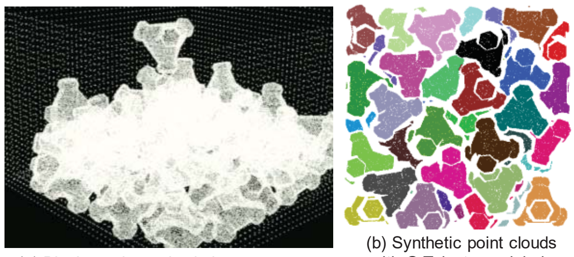
4.1 Instance segmentation accuracies

The point clouds of wave-dissipating blocks were captured by UAV-based photogrammetry from Sawara port in Hokkaido. Three validation regions with 10m by 10m as shown in **Fig. 6(a)** were sampled in it. Each region contains about 180,000 points, and 5-ton clinger blocks are stacked. The ground truth instance labels were made manually in these three regions.

An instance segmentation result using FPCC-Net is visualized in **Fig.6(b)-(c). Fig. 6(b)** indicates the center score. The red represents higher score region. **Fig. 6(c)** is the estimated block center found by non-maximum suppression**.** The final instance segmentation is visualized in **Fig. 6(d).** The segmentation accuracy was evaluated using the metric of the average precision (AP) with an IoU threshold of 0.5. As indicated in **Table 1**, the accuracies in all validation regions achieved 85-92%. It indicates that the proposed FPCC-Net exhibits the good ability enough for block instance segmentation.

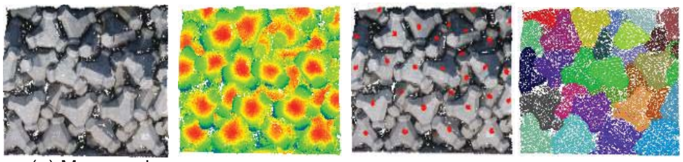
4.2 Pose estimation error

Given the fact that determining ground truth 6D poses of all blocks in scenes is a time-consuming and ambiguous task. Thus, an indirect method is used to evaluate the quality of the pose estimation. After the pose estimation, the scene was reconstructed by placing the block CAD model in those poses. **Fig. 7(a)** is an example of the point cloud of a validation region. Then the nearest distance between each measured point and the model is evaluated as the pose estimation error. **Fig.7(b)** indicates the estimated block poses of **Fig. 7(a),** and **Fig.7(c)** shows its pose estimation error. The smaller the error, the greener the color. Finally, the average nearest distance between each measured point and the model in the whole scene was evaluated as a pose estimation error. The errors in ten sampled regions are summarized in **Table 2**. All errors went below 0.1m, which exhibited the excellent ability of the proposed block pose estimation.



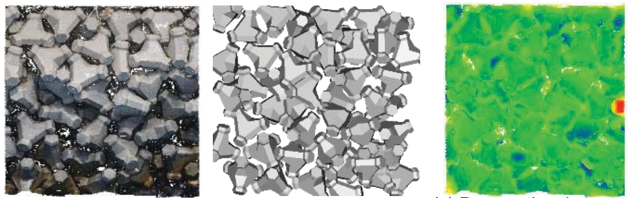
(a) Block stack-up simulation scene (b) Synthetic point clouds with G.T. instance labels

Fig. 5 Virtual block stack-up simulation



(a) Measured point cloud (b) Center score (c) Estimated block center (d) Instance segmentation

Fig. 6 An example of block instance segmentation



(a) Measured point cloud (b) Reconstructed scenes of estimated block poses (c) Pose estimation error (Green: Error <0.1m)

Fig. 7 An example of block pose estimation

Table 1 Instance segmentation accuracies in the validation [%]

Region no.	1	2	3
$AP_{0.5}$ [%]	92.3	90.9	84.6

Table 2 Block pose estimation error in different regions [m]

Region no.	1	2	3	4	5
Pose estimation error	0.06	0.04	0.04	0.03	0.03
Region no.	6	7	8	9	10
Pose estimation error	0.02	0.04	0.03	0.03	0.03

5 Summary

A method to recognize 6D poses of wave-dissipating blocks from 3D point clouds measured by photogrammetry was proposed based on deep-learning and shape descriptor. The validation results exhibited the excellent instance segmentation and pose estimation performances in block recognition. At present, only the segmentation on multi-instance but single-category scene has been implemented. In the future, we will implement the segmentation on multi-category and multi-instance scene.

References

[1] Y. Xu, et al., "A Convolutional Neural Network for Point Cloud Instance Segmentation in Cluttered Scene Trained by Synthetic Data Without Color," IEEE Access, 8, pp.70262-70269, (2020).

[2] Y. Xiang, et al., "PoseCNN: A convolutional neural network for 6d object pose estimation in cluttered scenes," arXiv:1711.00199, (2017).

[3] M. Drost, et al., "Model globally, match locally: Efficient and robust 3D object recognition," Proc. IEEE CVPR 2010, pp. 998-1005, (2010).

[4] P. J. Besl and N. D. McKay, "A method for registration of 3-D shapes," IEEE Trans. PAMI, 14(2), pp.239-256, (1992).

[5] Y. Wang, et al., "Dynamic graph cnn for learning on point clouds," ACM Trans. on Graphics, 38(5), pp.1-12, (2019).

[6] C.Erwin, and Y. Bai, "Pybullet, a python module for physics simulation for games, robotics and machine learning,"https://pybullet.org/wordpress/(2016).