Title	A study on stacked object recognition and stacking operation planning combining 3D point cloud representation, deep-learning and physics engine [an abstract of dissertation and a summary of dissertation review]
Author(s)	許, 雅俊
Citation	北海道大学. 博士(情報科学) 甲第15552号
Issue Date	2023-03-23
Doc URL	http://hdl.handle.net/2115/89552
Rights(URL)	https://creativecommons.org/licenses/by/4.0/
Type	theses (doctoral - abstract and summary of review)
Additional Information	There are other files related to this item in HUSCAP. Check the above URL.
File Information	Yajun_Xu_abstract.pdf (論文内容の要旨)

学 位 論 文 内 容 の 要 旨

博士の専攻分野の名称　博士（情報科学）　氏名　許　雅俊

学 位 論 文 題 名

A study on stacked object recognition and stacking operation planning combining 3D point cloud representation, deep-learning and physics engine
（3 次元点群表現, 深層学習および物理エンジンを用いた積み上げ物体認識と積み上げ作業計画に関する研究）

Technically, three-dimensional (3D) data which provides more decadent geometric, shape, and scale information than 2D data, make it easier for machines to understand and interact with their surrounding environment. Typical 3D data include depth images, point clouds, meshes, and voxels. Point clouds are widely used in various fields, such as robotics, autonomous driving, and civil engineering, to preserve the original geometric information in 3D space without discretized approximation.

In some realistic scenes, many artificial objects are stacked on each other. For instance, in a robotic bin-picking scene, wherein heavily piled-up parts occlude each other; on the coast, a large number of wave-dissipating blocks are stacked together to protect the embankment. Recognizing stably and accurately individual objects in these cluttered object-stacked scenes remains an unsolved problem in 3D object recognition research. Planning optimal poses of new objects stacked onto the existing stacked objects to address engineering requirements is an even more considerable challenge in optimization and task planning research.

This dissertation first designs a 3D instance segmentation framework for stacked objects scenes using a deep neural network and applies it to the precise 3D object pose estimation and object type classification. Then the dissertation develops a system that simulates object stacking using a physics engine and deep learning to plan the stacking operation of new additional objects based on the individual stacked objects that have been recognized.

Several deep learning-based 3D recognition frameworks for indoor scenes from point clouds have been proposed recently. However, the deep learning of 3D point clouds still faces several significant technical challenges, such as efficient and reliable data annotation, the memory required to process large-scale point clouds, and time-consuming processing. This study proposes a novel fast point cloud clustering-based deep neural network, called FPCC, for the instances segmentation of the point cloud of stacked objects. FPCC simultaneously predicts the geometrical similarity of points and the likelihood of being centroids. Based on the predicted results, this study designs a novel clustering algorithm that can quickly generate the final instance segmentation results. Experimental results on public datasets of robotic bin-picking scenarios showed that the proposed method outperformed the current state-of-the-art segmentation methods in precision and processing speed.

Then, this study extends the application of this 3D instance segmentation scheme to the recognition of wave-dissipating blocks, a structural unit of breakwaters. The recognition consists of three main steps: point cloud instance segmentation of the blocks, pose estimation, and classification. For this purpose, a point cloud feature extractor is newly designed to replace the original feature extractor of FPCC, which can process more points faster with the same computational overhead. The new feature extractor employs an attention-pooling mechanism, which allows the neural network to learn richer local information. Afterward, the block-wise 6D pose is estimated using a three-dimensional feature descriptor, point cloud registration, and CAD models of blocks. Finally, the type of each segmented block is classified using model registration results. The performances of wave-dissipating block instance segmentation, block pose estimation, and block type classification for large-scale point cloud scenes captured by UAV and multi-beam echo sounder demonstrated the effectiveness of the proposed approach.

Finally, based on the recognized results of wave-dissipating blocks, this dissertation develops a system to simulate the block stacking plan utilizing a physics engine and deep learning. It can predict the additional block amounts and their stacking poses and provides pre-visualization of their stacking operations. Deep learning was used to estimate the additional block poses that better fit the stacked blocks. The simulation was applied to an actual block-stacking operation in a local port at Hokkaido. The final construction results in the real world verified the accuracy and usefulness of the simulation.

To sum up, this dissertation makes three significant contributions to object recognition and object stacking simulation. The first proposed a fast framework for point cloud instance segmentation called FPCC to realize faster and more reliable robotic bin picking. The second improved FPCC and its use for instance segmentation of stacked wave-dissipating block scenes. Combined with pose estimation, this enables us to accurately retrieve most of the blocks at the instance level from point clouds of realistic cluttered 3D scenes. The third was developing a simulation system for mimicking the additional object stacking operations on the existing stacked objects. It provided the functions of customizable pre-visualization and optimal object stacking operation planning according to different engineering requirements, such as supplementary works of wave-dissipating blocks.