A 2D-image-AI-based 3D Detection Method Using Continuous Orthogonal View Projections

Chaoyu Dong[1,3], Zhaoheng Shi[1], Hong Luo[1]

1 Singapore Institute of Manufacturing Technology (SIMTech), Agency for Science, Technology and Research (A*STAR), 5 CleanTech Loop #01-01, CleanTech Two Block B, Singapore 636732, Republic of Singapore

2 College of Computing and Data Science, Nanyang Technological University, 50 Nanyang Avenue, Singapore 639798, Republic of Singapore

Corresponding Author / Email: dongcy@simtech.a-star.edu.sg, TEL: +65-97716329

KEYWORDS: Point cloud, 3D, Precision measurement, Prefabricated Prefinished Volumetric Construction

There is a growing need to support 3D measurement to enhance efficiency in quality control. Currently, building model generation primarily relies on manual measurements to determine the properties of architectural, mechanical, electrical, and piping elements. While advancements in 3D scanning technology now enable the scanning of existing buildings, converting this scan data into usable models still requires significant human involvement due to challenges such as large data sizes, occlusion, and the presence of various structural elements. As a result, a more efficient method is needed to streamline 3D object detection and measurement from such complex datasets to enable fully automated scan-to-model generation. This work proposes an accelerated 3D component measurement strategy that improves efficiency by projecting 3D points with RGB information onto multiple continuously moving orthogonal 2D views, effectively reducing data complexity. The projected 2D images are then used for object detection training. Based on the object types and positions detected in multiple orthogonal 2D views, objects are segmented and re-projected back into the 3D space to enable 3D object modeling and measurement. The exact dimensions of the objects are determined through a multi-shape 3D geometry extraction process using the detected 3D bounding boxes. This approach significantly enhances the automation and accuracy of 3D measurement, making scan-to-model generation more efficient for various applications.

NOMENCLATURE

AM = Additive Manufacturing
MEP = mechanical, electrical and piping
PPVC = Prefabricated Prefinished Volumetric Construction
PBU = Prefabricated Bathroom Unit

1. Introduction

The 3D point cloud has become a widely adopted format for representing spatial environments. Leveraging point cloud data from 3D scanners, [1] integrated multiple scales to extract features from complex scenes, enabling effective classification and segmentation of spatial objects. For RGB-D scans in both indoor and outdoor settings, [2] employed 2D and 3D deep learning techniques to generate 3D bounding boxes for object detection. In the context of autonomous driving [3], the point cloud analysis was organized into two stages, proposal generation and refinement, to learn spatial features [4]. Beyond indoor scanning and autonomous driving, 3D point clouds have also been applied to corrugated surface inspection [5], offering a robust representation that complements traditional 2D imaging methods [9].

Various deep learning architectures can be employed to process 3D point cloud data, including convolutional neural networks [7], multi-layer perceptron [8], and transformer [9]. However, regardless of the chosen network, directly processing 3D point cloud data is computationally intensive. This is due to the large number of points in the space, each requiring recognition, which significantly impacts detection efficiency.

Efficient 3D object measurement can accelerate various applications. For example, there is an increasing demand to support existing buildings that have only 2D design blueprints for effective maintenance [10]. However, generating building models often involves manual measurement and identification of architectural and MEP elements. On the other hand, advancements in 3D scanning technology have made it feasible to scan existing buildings, producing 3D point clouds enriched with RGB information [2]. Incorporating efficient 3D object measurement techniques can significantly enhance the processing of large-scale point cloud data, improving overall workflow efficiency.

To address this challenge, this work introduces a 2D-image-AI-based 3D detection method designed to enable efficient 3D object measurement. The approach consists of four key steps: exterior extraction, interior projection, 3D detection, and dimension measurement.

The structure of this work is organized as follows: Section 2 presents the 2D-image-AI-based 3D detection method. Section 3 outlines the experiments performed for 3D assessment. Section 4 concludes the study with key findings and insights.

2. 2D-image-AI-based 3D Detection Method

To achieve efficient 3D measurements, the proposed strategy begins by decoupling the 3D point cloud into exterior and interior components. This is followed by interior projection and 3D object detection, culminating in the final dimension measurement.

2.1 Exterior Extraction

When a 3D point cloud is used to represent a large space or object, planar patterns are the most prevalent, accounting for the majority of points.

By applying iterative plane fitting, the exterior 3D point clouds can be sequentially decomposed into individual planes based on the point density. Gradually removing these exterior planes from the complete 3D point cloud significantly reduces the total number of points, thereby enabling more efficient interior projection.

2.2 Interior Projection

To illustrate using the Y-Z plane projection, consider that the differences between the maximum and minimum values of the point cloud in the X, Y, and Z directions are 500 mm, 600 mm, and 700 mm, respectively. Consequently, the projection image on the Y-Z plane measures 700 × 600. Along the X-axis, which is perpendicular to the Y-Z plane, 500 planes can be generated using a segmentation step of 1 mm. This dynamic projection process is depicted in Figure 2.

In complex environments, target objects are often obscured by other elements during projection. To address this challenge, the orthogonal projection view will be adaptively adjusted, ensuring that all target objects are effectively captured in the projection images.

An object is inferred to be the same across two orthogonal projection views if it receives identical labeling from the AI-trained model and exhibits nearly identical 2D bounding box dimensions in both projections.

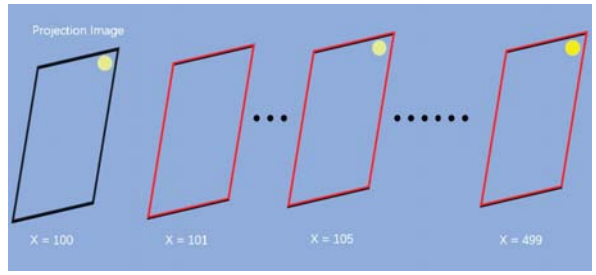


Figure 1. Moving projection in the Y-Z plane.

2.3 3D Detection

To identify target objects, each projected 2D image is processed using YOLO [11] for object detection. Based on the detected object types and positions across multiple orthogonal 2D views, the objects are segmented and re-projected into 3D space, enabling the creation of 3D object models.

One challenge of object detection using projected images is the false detection caused by object shadows. Shadows can lead to two types of false positives: bounding boxes with no actual object inside or objects being incorrectly detected on walls. To address the first type, the number of 3D points within the bounding box is counted; if this number is insufficient, the detection is classified as false and removed. For shadow-related false detections on walls, the bounding boxes often have disproportionately large dimensions. These are filtered out based on their size. The filtering process is illustrated in Figure 2.

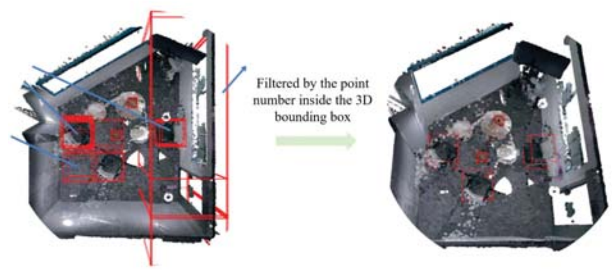


Figure 2. Shadow filtering via point number within 3D bounding boxes.

2.4 3D Geometry Determination

With the detected 3D bounding boxes, the exact object dimensions are determined through a multi-shape 3D geometry extraction process. The workflow begins by normalizing the enclosed point clouds within the bounding box based on the farthest distance from the origin. The normalized point clouds are then denoised using a Gaussian distribution to remove outlier points, ensuring precise measurements. A sequential shape extraction approach is employed: starting with planes, followed by spheres, and then cylinders. Iterative fitting is conducted for each shape, with fitted point clouds removed after each step. This allows for the computation of comprehensive dimensions for objects composed of multiple geometric shapes. The complete workflow of the proposed 2D-image-AI-based 3D detection method, starting from 3D point cloud data, is summarized in Figure 3.

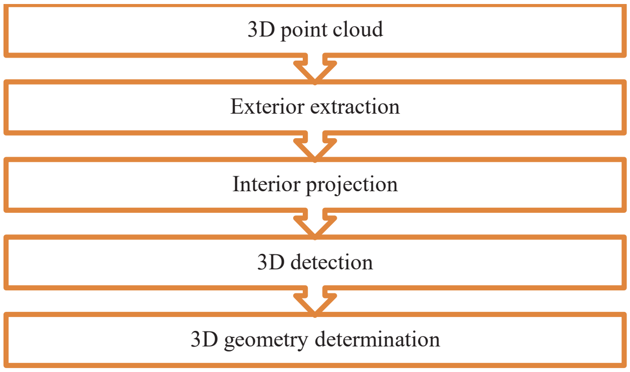


Fig. 3 2D-image-AI-based 3D detection workflow.

3. Experiment and Analysis

The proposed 2D-image-AI-based 3D detection method is validated through an indoor space scan using the Leica BLK360. The scanning locations and the generated 3D point cloud are shown in Figure 4. Subsequently, a 3D point cloud representing the PBU is used to demonstrate the general applicability of the proposed method.

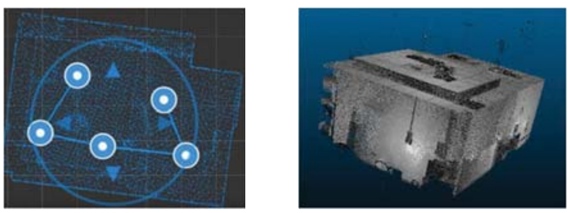

Scanning locations 3D point cloud

Fig. 4 Scanning locations and 3D point cloud.

3.1 Implementation Process

The complete implementation process is illustrated in Figure 5, starting from the input 3D point cloud. Through exterior extraction, the 6-side surroundings are removed from the entire dataset, significantly reducing the number of points. Internal components, such as the fire extinguisher, are then detected in the projection images. The detected 2D objects are used to formulate the 3D bounding boxes. Finally, 3D geometry determination is conducted for plane, sphere, and cylinder fitting. For instance, the detected coordinates of the fire extinguisher's center are [-2.365, -3.727, -0.057], with a radius of 68mm. This radius result matches that obtained using commercial software GOM from ZEISS [12].

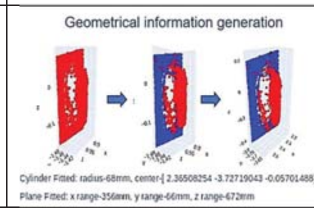
Input	
Exterior extraction	
Interior projection	
3D detection	
3D geometry determination	
Output	

Fig. 5 Implementation process of proposed method.

3.2 Other Applications

In addition to the indoor space, the PBU is also utilized to demonstrate the effectiveness of the proposed approach. Using the 3D point cloud shown in Figure 6, the developed method successfully decouples the exteriors and interiors, ultimately generating a mesh model for the construction unit. Furthermore, a user interface, as depicted in Figure 7, has been developed to enable data visualization and interaction, facilitating a more intuitive exploration of the results.

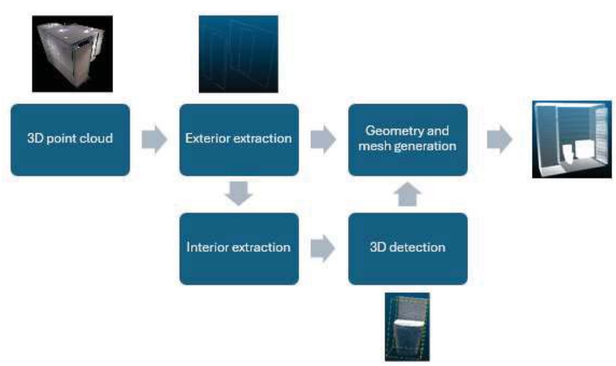


Fig. 6 Application of the method in construction.

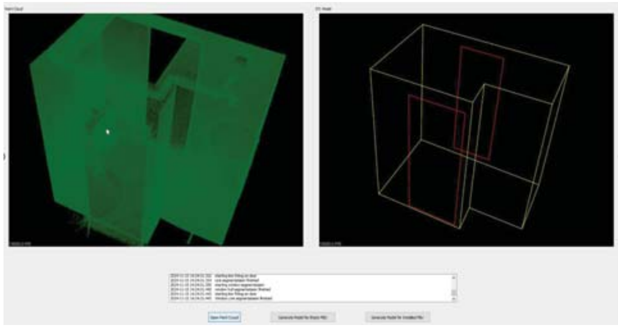


Fig. 7 User interface of the developed method.

3.3 Efficiency Advantage

To further evaluate the efficiency of the proposed method, a point cloud dataset containing 3,887,089 points is used. A comparison is made between PointNet++ and our proposed 2D-image-AI-based 3D detection method. The processing time and hardware configurations are summarized in Table 1.

As shown in Table 1, the 2D-image-AI-based 3D detection method can process the same point cloud data much faster, even when using less powerful hardware. This is because the proposed method converts the 3D data processing into a 2D domain, which significantly reduces the time and dependency on high-end CPUs and GPUs.

Table 1 Accuracy assessment with feature components

Method	Point number	Processing time (second)	CPU	GPU
PointNet++ [1]	3887089	80	Intel® Xeon® Gold 5320	NVIDIA RTX A6000 (with 48 GB memory)
Proposed method	3887089	14	Intel(R) Core (TM) i7-10750H	NVIDIA GeForce RTX 2080 Super (with 8GB memory)

4. Conclusions

A 3D component measurement strategy is proposed that involves projecting 3D points with RGB information onto multiple continuously moving orthogonal 2D views. These projected 2D images are then utilized for object detection. By analyzing the object types and positions detected across multiple orthogonal 2D views, the objects are segmented and subsequently re-projected back into the 3D space. This approach integrates 2D detection within the 3D measurement process, thereby enhancing the efficiency of 3D measurement.

ACKNOWLEDGEMENT

This research is supported by A*STAR under its RIE2025 Manufacturing, Trade and Connectivity (MTC) Industry Alignment Fund- Pre-Positioning (IAF-PP) (Award M21K1a0104).

REFERENCES

1. Qi, C.R., Yi, L., Su, H. and Guibas, L.J., "Pointnet++: Deep Hierarchical Feature Learning on Point Sets in A Metric Space," Advances in Neural Information Processing Systems, Vol. 30, 2017.
2. Qi, C.R., Liu, W., Wu, C., Su, H. and Guibas, L.J., "Frustum Pointnets for 3D Object Detection from RGB-D Data," In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp.918-927, 2018.
3. Geiger, A., Lenz, P. and Urtasun, R., "Are We Ready for Autonomous Driving? the KITTI Vision Benchmark Suite," In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp.3354-3361, 2012.
4. Shi, S., Wang, X. and Li, H., "Pointrcnn: 3D Object Proposal Generation and Detection from Point Cloud," In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 770-779, 2019.
5. Dong, C., Shi, Z. and Luo, H., "LIDAR-based Automated 3D Inspection System for Corrugated Surface Defects," In Proceedings of the Asian Society for Precision Engg. and Nanotechnology, 2022.
6. Dong, C., Liu, T., Cheng, F. and Qian, K. "X-ray Artificial Intelligence Identification System for Agricultural Products In Proceedings of the Asian Society for Precision Engg. and Nanotechnology, 2022.
7. Shi, S., Wang, Z., Shi, J., Wang, X. and Li, H., "From Points to Parts: 3D Object Detection From Point Cloud with Part-aware and Part-aggregation Network," IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 43, No. 8, pp.2647-2664.
8. Hao, N., "3D Object Detection from Point Cloud Based on Deep Learning." Wireless Communications and Mobile Computing, Vol. 1, pP.6228797, 2022.
9. Dong, C., Tan, A.W.W., Liu, T., Cheng, F. and Qian, K., "Transformer-based Smart Inspection for Agricultural Products via X-ray Images," In International Conference on Optical and Photonic Engineering, Vol. 13069, pp.322-327, SPIE, 2024.
10. Volk, R., Stengel, J. and Schultmann, F., "Building Information Modeling (BIM) for Existing Buildings—Literature Review and Future Needs," Automation in construction, Vol. 38, pp.109-127, 2014.
11. Dong, C., Putra, J.S., and Malcolm, A. A., "Artificial Intelligence Defect Detection Robustness in Real-time Non-Destructive Testing of Metal Surfaces," In 5th Singapore International NDT Conference & Exhibition, 2025, accepted.
12. GOM, https://www.zeiss.com/metrology/en/software/metrology-software.html.