2024 IEEE/ASME International Conference on Advanced Intelligent Mechatronics (AIM)
July 15-19, 2024. Boston, MA, USA

Efficient Color Point Cloud Reconstruction with Robotic Arm-Guided Multiview Fusion

Udaka A. Manawadu and Naruse Keitaro

Abstract— **This research addresses the significant challenge of noise in point cloud (PC) data, which undermines the accuracy of object recognition and pose estimation. We introduce a novel methodology that leverages systematic robotic camera movements for multiview PC acquisition, aimed at enhancing reconstruction accuracy in noisy environments. Using a Jaco robot arm outfitted with a Realsense D435 RGB-D camera, PCs of a globe valve are captured from multiple angles, focusing on a 60° area with 15° intervals. This setup results in a dataset of 81 PCs per iteration, with a total of three noisy PC datasets collected for analysis. The PCs are merged at 15° and 30° intervals, using the Color Iterative Closest Point (ICP) algorithm and refined through downsampling. The method is evaluated by measuring position and orientation accuracy using the Random sample consensus (RANSAC) algorithm across 216 instances-108 each at 15° and 30° intervals. The proposed methodology enhances the accuracy of pose estimation by 93% in both intervals, reducing mean errors in position to 2.35 mm and in orientation to 18.4°. This significant improvement underscores the effectiveness of our approach in mitigating noise in PC data for more precise object recognition and pose estimation applications.**

I. INTRODUCTION

Point clouds (PCs) are indispensable for object recognition and pose estimation, providing rich three-dimensional (3D) representations, making them essential for accurately discerning object identities and determining their spatial orientations in real-world environments [1], [2]. The quality of the PC is crucial for achieving accurate object recognition and pose estimation [2]. Noise in PCs presents significant obstacles to the accuracy of object recognition and pose estimation. Therefore, reconstructing PCs is crucial for mitigating the effects of noise and enhancing the quality of data, ultimately leading to more precise object recognition and pose estimation, especially in noisy environments [3], [4].

Over the past decade, the use of color PCs for 3D object recognition and pose estimation has gained popularity, thanks to the widespread availability of the necessary devices and the ease with which objects can be recognized from color PCs [1], [5], [6]. Color details in PCs crucially enhance object recognition by providing nuanced information absent in monochrome data. In recent times, the domain of factory automation has seen the integration of RGB-D cameras with robot manipulators for object handling tasks with various methods employed for object recognition and pose estimation across different scenarios [1]. Geometric errors in 3D reconstructions present significant hurdles in faithfully representing scanned shapes, arising from factors such as data noise, imperfect sampling, and algorithmic limitations [3]. In previous research, our team encountered significant issues with PC noise, which prompted the development of a novel system for object recognition and pose estimation in factory automation, as described in [6]. This system employs an RGB-D camera mounted on a robotic arm to enhance accuracy through pose estimation from multiple viewpoints. While it performs well in controlled environments, its effectiveness is compromised in noisier settings. Consequently, the current research explores Multi-View PC acquisition using an RGB-D camera as a promising solution, particularly when combined with a camera attached to a robotic arm.

Obtaining PCs from various viewpoints has become a popular method for 3D PC restructuring in recent years. Takimoto *et al.* suggest a technique to enhance the accuracy of 3D reconstructions by merging PCs captured from various viewpoints with structured light [7]. The authors used a Kinect RGB-D sensor to capture multiple PCs from different angles and applied the Iterative Closest Point (ICP) algorithm to generate reconstructions over large areas. Wei *et al.* proposed a method for automatic identification and autonomous sorting of cylindrical parts in cluttered scenes using monocular vision 3D reconstruction [8]. The research aimed to solve the challenge of obtaining a globally optimal solution for pose estimation. The method involved capturing images of parts, extracting color signatures, using Random sample consensus (RANSAC) and Remote Closest Point (RCP) algorithms, and completing fine registration with Levenberg-Marquardt-ICP. The experimental results demonstrated higher pose estimation accuracy than other methods, with minimal position and orientation errors. Sang *et al.* propose a novel RGB-D reconstruction method that addresses challenges in surface modeling and image formation [9]. Their approach leverages multi-view uncalibrated photometric stereo and gradient Signed Distance Function (SDF) to achieve high-quality reconstructions of fine-scaled geometry, albedo, lighting, and camera tracking. Extensive evaluation of synthetic and real-world datasets demonstrates their effectiveness.

In contrast to prior studies that rely on static RGB-D camera positions for acquiring multiview 3D data, our proposed approach introduces a novel methodology employing a robotic arm-guided RGB-D camera system that allows to capture of 3D data from multiple viewpoints. This addresses the limitations associated with static configurations, as identified in our previous research [6]. Our work highlights

This research was performed by the commissioned research fund provided by F-REI (JPFR23010102)

Udaka A. Manawadu and Naruse Keitaro are with the Graduate School of Computer Science and Engineering, University of Aizu, Fukushima, Japan
d8232111@u-aizu.ac.jp, naruse@u-aizu.ac.jp

the challenges arising from geometric errors in 3D reconstructions derived from PC data, despite the demonstrated effectiveness of algorithms like ICP for PC registration and alignment. Notably, despite the growing prominence of PC technology, the literature lacks a comprehensive exploration of multiview PC integration, specifically the integration of PCs collected by systematic camera movements. Our study aims to fill this gap by investigating the potential benefits of integrating information from multiple viewpoints, contributing to more accurate and comprehensive scene representations.

The ICP algorithm is widely employed as a method for registering two sets of 3D points or surfaces [10]. The color ICP algorithm surpasses traditional ICP by reducing search time, enhancing accuracy with color-based point selection, enabling faster convergence, and improving reconstruction accuracy [11]. To extend its applicability to RGB-D data, Color-ICP can leverage color and geometric information to establish correspondences between points in the PCs being registered [12]–[14]. The integration of color information proves beneficial in disambiguating correspondences and refining alignment accuracy. This becomes particularly valuable in scenarios where geometric features alone are insufficient for constraining optimization, such as on smooth surfaces. One of the primary objectives of this research is to enhance the accuracy of pose estimation derived from PC reconstruction. The RANSAC algorithm, widely utilized as a 3D pose estimation method, is central to our approach [5], [6], [15]. In our specific use case, we conduct the verification of pose estimation accuracy within the framework of the RANSAC algorithm.

In response to the prevalent challenge of noise in PC data, our research proposes a novel approach to reconstructing PC data through multiview acquisition involving systematic camera movements. A globe valve will be used as the target object in this research. To execute this methodology, we use a Jaco robot arm equipped with 6 Degrees of Freedom (DOF), housing a Realsense D435 RGB-D camera atop its frame. A testing environment with artificial noise is used to collect data. A large number of PCs are collected from various angles by manually operating the robotic arm around the target object. Subsequently, the captured PCs from different angles are merged using the Color ICP algorithm. A downsampling method is used to smooth the merged PC. The accuracy of the resulting PC will be evaluated by measuring the positional and orientational accuracy of the final reconstruction, which is referred to as "efficient" in this research. Different PCs from different angles are merged to analyze the most accurate angles. The key contributions of this research include:

- Reconstructing a large number of noisy PC data gathered from systematic robotic camera movements using the introduced methodology.
- Investigating the accuracy of reconstructed PC integration in data obtained through systematic robotic camera movements.
- Comparing the accuracy of PC reconstruction using data merged at different angles.

II. METODOLOGY

A. Testing Environment

To acquire color PC data with added noise, the authors devised an artificial environment, illustrated in Figure 1. Artificial noise was introduced using a torch. The reflection produced by the torch's light beam introduces noise into the PC. Data were collected both horizontally (yaw axis from the target) and vertically (pitch axis from the target) relative to the target object, as shown by the yellow and pink curved arrows in Figure 1. Suppose the current position of Figure 1 is directly facing the target object, the acquired PC is labeled as *pitch000-yaw000*. In Figure 1, "A" represents the placement of the torch, "B" signifies the target object (a globe valve), "C" denotes the Intel Realsense D435 RGB-D camera, and"D" represents the JACO arm.

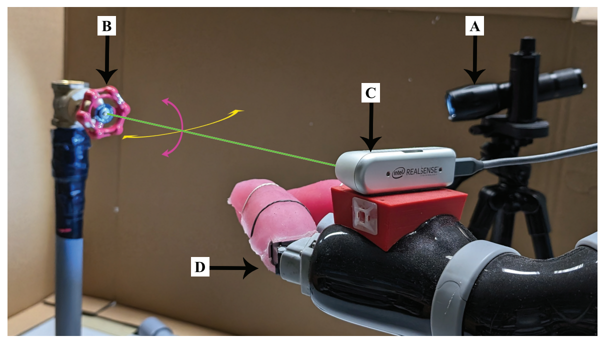

Fig. 1. Noise data collecting environment. The yellow and pink curved arrows indicate the movement directions of the robot arm while the green line shows the direction of the object from the camera.

B. Proposed System Architecture

The proposed system architecture, depicted in six stages in Figure 2, involves collecting noisy PC data using systematic movements of a robot arm, as shown in Figure 1. This is followed by the application of post-processing filters to lessen the PC density. The process continues with color-based region-growing segmentation and filtering to isolate the target object from noise. Color PC registration is then executed using the Color ICP algorithm, with subsequent downsampling to further reduce PC density. Finally, the RANSAC algorithm verifies the accuracy of the final PC.

1) Move robot arm to pre-programmed positions: The data-gathering process is designed to capture the target object using an RGB-D camera mounted on the Jaco robot arm with 6 DOF as in Figure 1. A globe valve becomes the target object in this research. As shown in Figures 3 on the left, the area 60 ° to the left, right, up, and down from the target object's middle was selected for data collection. This area was chosen to maximize the RGB-D camera's field of view while capturing the most critical aspects of the globe valve. Data collection is performed from the central axis of the object, employing a 15-degree interval between each instance, as depicted in Figure 3 on the right. The 15-degree

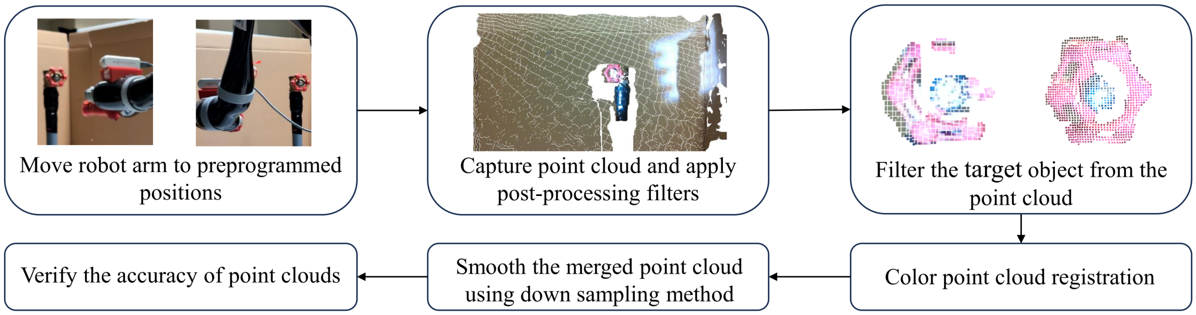

Fig. 2. The proposed reconstruction method with six stages.

interval between each instance provides detailed coverage without redundancy when PC merging. This systematic approach ensures the acquisition of diverse and detailed PC information. A total of 3 noisy PC data sets were collected in this research.

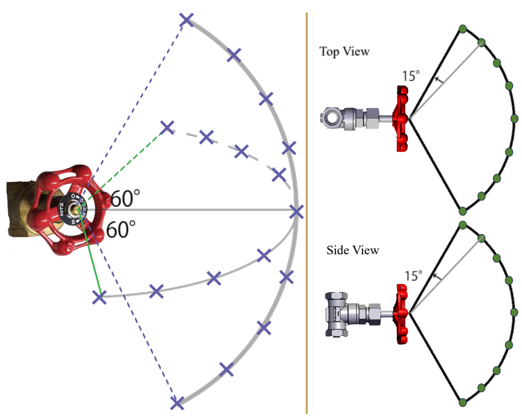

Fig. 3. Left: Data were gathered horizontally and vertically from the target object. For the target object, an area of 60 degrees to the left, right, up, and down was considered. Right: A total of 81 data points were considered, with 15-degree intervals at each point.

2) Capture PCs and apply post-processing filters: This research uses Intel Realsense D435 as the depth camera [16]. Before applying the object recognition part post-processing is required to enhance the quality and accuracy of the depth data captured by the sensors [17]. As in Figure 2, smooth alpha and smooth delta are used to control the amount of smoothing applied to the depth data, while hole filling is employed to fill in gaps or "holes" in the depth map.

3) Filter the target object from the PC: After applying post-processing filters, a color-based region-growing segmentation algorithm is used to segment the color scene PC [18]. Subsequently, a color-filtering method is employed to isolate the target object from the segmentations. For improved filtering, RGB data of the PC is converted to the HSV (Hue, Saturation, Value) color space, which offers a more intuitive representation of colors [19]. Figure 2 displays examples of globe valve handles after being filtered using these algorithms.

4) Color PC registration: Color ICP algorithm is used for PC registration which uses both geometric and color information to establish correspondences between points in the PCs [12]. Let P and Q denote two colored PCs, where each point $p_i \in P$ is characterized by its 3D coordinates (x_i, y_i, z_i) and associated RGB color values (R_i, G_i, B_i). Similarly, each point $q_j \in Q$ is described by analogous attributes. The objective of Color ICP is to determine the transformation T comprising rotations and translations that minimize the distance between corresponding points in P and Q. The optimization function for the registration process is formulated as follows:

$$E(T) = (1-\delta)E_{col}(T) + \delta E_{geo}(T), \tag{1}$$

where $E_{\text{geo}}(T)$ represents the geometric error and $E_{\text{col}}(T)$ the color error. The geometric error is defined as the sum of squared distances between corresponding points in the transformed PC P and Q:

$$E_{\text{geo}}(T) = \sum_i \sum_j \omega_{ij} \cdot \|T \cdot p_i - q_j\|^2, \tag{2}$$

and the color error is the sum of squared differences in color values between matched points:

$$E_{\text{col}}(T) = \sum_i \sum_j \omega_{ij} \cdot \|\text{color}(p_i) - \text{color}(q_j)\|^2, \tag{3}$$

Here, ω_{ij} are weights assigned based on the confidence in each correspondence, and δ is a parameter between 0 and 1 that balances the influence of geometric versus color data. The parameter δ is empirically determined to optimize registration accuracy for specific applications. The Color ICP algorithm iteratively refines the transformation T by minimizing the combined error. The process continues until convergence, which is determined when the change in error between successive iterations falls below a predefined threshold, indicating a stable solution. This iterative approach ensures that both color and geometric alignments contribute effectively to the final registered PC.

5) Smooth the PC using downsampling method: Voxel grid downsampling was used in this research to smooth the PCs [20]. This technique ensures a more uniform distribution of points, and in our case, it reduces the density of the PC by averaging the points within each voxel. Suppose the selected voxel is s_x units wide, s_y units long, and s_z units tall. From

Equation 4, it can be found which voxel a point belongs to as follows,

$$\text{Voxel Index}(x, y, z) = \left(\left\lfloor \frac{x}{s_x} \right\rfloor, \left\lfloor \frac{y}{s_y} \right\rfloor, \left\lfloor \frac{z}{s_z} \right\rfloor \right). \quad (4)$$

For each voxel, if there are multiple points inside, the average position becomes the representative point for that voxel. Suppose the representative point of the voxel is denoted as V,

$$V(x_v, y_v, z_v) = \frac{1}{N} \sum_{p=1}^{N} (x_p, y_p, z_p), \quad (5)$$

where N is the number of points within the voxel. In this research, a voxel of s_x = 1.5 mm, s_y = 1.5 mm, and s_z = 5 mm was selected.

6) Verify PCs: The RANSAC algorithm is utilized to verify the accuracy of the PC, focusing on position and orientation. RANSAC is particularly effective in environments with high noise levels and outliers, typical of industrial settings. Specifically, for the globe valve, the RANSAC circle algorithm facilitates accurate pose estimation of the round-shaped handle [6], [21]. By robustly distinguishing inliers from outliers, RANSAC ensures that position accuracy assessments are based on the most reliable data, enhancing the precision of the estimated pose. Using three randomly sampled points, p_1, p_2, and p_3, on the globe valve handle, the normal n and the center c_1 of the handle are calculated using Equations 6 and 7 as follows:

$$n_1 = \frac{(p_2 - p_1) \times (p_3 - p_1)}{\|(p_2 - p_1) \times (p_3 - p_1)\|}, \quad (6)$$

$$c_1 = [n_1 \quad p_2 - p_1 \quad p_3 - p_1]^{-T} \begin{bmatrix} n_1^T p_1 \\ (p_2 - p_1)^T (p_1 + p_2)/2 \\ (p_3 - p_1)^T (p_1 + p_3)/2 \end{bmatrix}, \quad (7)$$

The position error of the globe valve can be calculated as,

$$Error_{pos} = \|C_{true} - c_1\|, \quad (8)$$

where C_{true} is the true center, manually selected in each PC. The orientation accuracy is calculated by the angular difference of the normal vector given by the RANSAC plane algorithm and $n1$ given by Equation 6. The algorithm identifies the dominant plane by analyzing the coefficients a,b, and c, which represent the plane fitted to the PC. Given that the target PC is approximated as a plane, these coefficients are used to compute the normal vector (n_2) of the plane. This approach ensures accurate orientation measurements by aligning the derived plane's orientation with the true geometrical configuration of the object, thereby enhancing the reliability of the pose estimation. If true orientation considered as the orientation given by n_2, then Orientation Error (E_{ori}) can be obtained as,

$$E_{ori} = \arccos \left(\frac{\mathbf{n}_1 \cdot \mathbf{n}_2}{\|\mathbf{n}_1\| \|\mathbf{n}_2\|} \right). \quad (9)$$

Finally, the accuracy improvement (%) was calculated using this algorithm,

$$\text{Improvement} = \frac{\text{Initial Error} - \text{Final Error}}{\text{Initial Error}} \times 100\%. \quad (10)$$

The accuracy of both position and orientation was calculated.

III. Results and Discussion

From the collected noise data batch, nine positions are selected for 3D reconstruction, each spaced 30 degrees apart from the next. 15° and 30° PC merging were considered in this research. Each PC comprises four instances to be merged: two instances with PCs differing by 15 and 30 degrees horizontally (yaw axis) and two instances with PCs differing by 15 and 30 degrees vertically (pitch axis). Figure 4 illustrates the step-by-step results provided by the proposed methodology for four PCs. In total, 36 instances of merging (9 x 4) were considered from one dataset. This research considered 216 instances from the three noisy datasets to measure the accuracy of the proposed methodology comprising 108 instances for 15° and 108 instances for 30°.

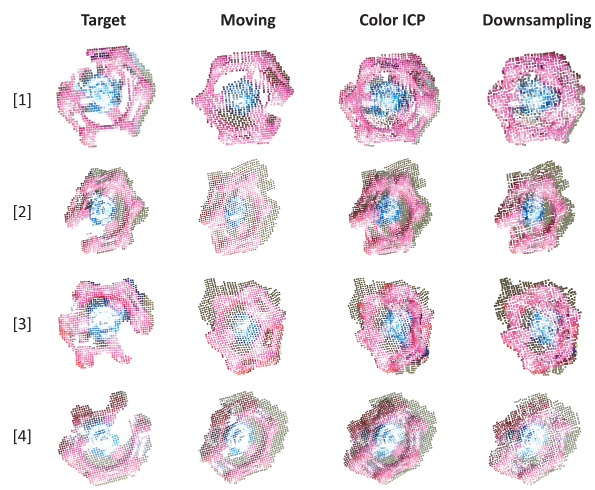

Fig. 4. Applying the color ICP algorithm and then downsampling it to noisy data for 30° instances. Illustrations No. 1: the merging of pitch000-yaw000 data into pitch000-yaw330, No. 2: pitch000-yaw030 merged into the pitch000-yaw060PC, No. 3: the merging of pitch030-yaw000 into pitch030-yaw330, and No. 4: pitch330-yaw000 merged into pitch330-yaw030.

From Figure 4, it is evident that the noise data of the target PC can be restructured using the proposed methodology. The incomplete points in the target PC were filled after merging using color ICP. The integration of color ICP facilitates significant densification of the merged PCs, imbuing them with a rich abundance of data points. However, such heightened density necessitates a subsequent downsampling process to refine and smoothen the PC. The refined PC exhibits a more refined and organized structure, with individual data points arranged more uniformly, thereby enhancing the overall perceptual quality. Moreover, the downsampling process effectively mitigates the previously observed thickness of the PC, resulting in a smoother and more visually appealing representation of the underlying data.

TABLE I

THE POSITION AND ORIENTATION RESULTS FOR 30° AND 15° MERGING.

Error	30°								15°							
	Position prediction (mm)				Orientation prediction (°)				Position prediction (mm)				Orientation prediction (°)			
	Mean	Min	Max	SD	Mean	Min	Max	SD	Mean	Min	Max	SD	Mean	Min	Max	SD
Initial	37.6	4.6	191	71	8.53	3.81	13.3	3.7	30.7	4.79	191	63.2	9.59	2.9	15.6	3.86
C-ICP	4.88	1.8	8.4	1.67	7.89	2.5	13.5	3.11	5.16	2.61	30.8	5.23	7.22	1.1	14.2	3.53
DS	2.49	1.2	5.0	0.60	7.65	2.5	13.4	3.23	2.22	1.06	5.08	0.93	7.1	1	13.5	3.34
Improv	93.3%	73%	97.3%	99%	10.3%	-	-	12.7%	92.8%	77.7%	97%	98%	26.5%	-	-	13%

Fig. 5 illustrates the results produced by the RANSAC algorithm. In this depiction, the inliers utilized to predict the position of the PC in the RANSAC algorithm are represented by green circles. Notably, the inliers utilized for determining position and orientation in the "Target" PC do not form complete circles, indicating potential inaccuracies in position and orientation estimation. Upon examination of Figure 5, it becomes apparent that there is an enhancement in the prediction of the position within the final PC (post-downsampling) in contrast to the initially captured PC (target). Here, the midpoint of the valve serves as the reference true position. To assess the accuracy of the position estimation, we calculate the error by measuring the disparity between the current position indicated by the endpoint of the green line in Figure 5, and the estimated position derived using Equation 7.

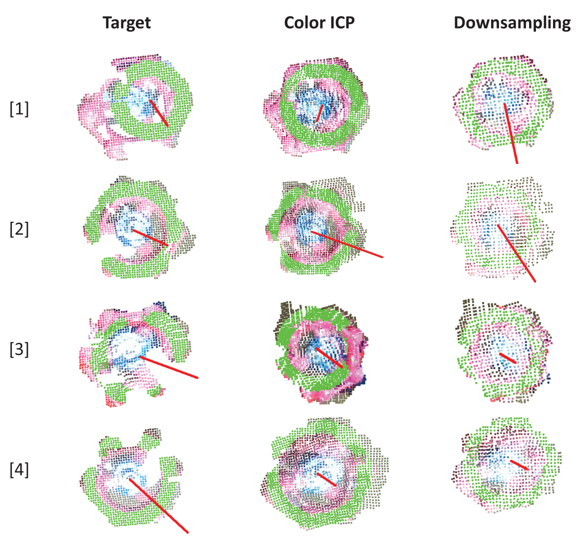

Fig. 5. Calculated position for original, color ICP, and downsampling for the same samples in Figure 4.

Table I displays the PC dataset results, where "C-ICP" refers to Color ICP PC, "DS" to downsampled PC, and "Improv" to the percentage (%) improvement calculated by Equation 10. For the 30° interval, the mean position error was markedly reduced from an initial 37.6 mm to 4.88mm post-Color-ICP and further to 2.49mm following downsampling, reflecting a 93.3% improvement from the baseline. This trend was consistent in the 15° data, where mean position errors decreased from 30.7 mm to 5.16 mm and then to 2.22 mm, achieving a 92.8% reduction. Orientation errors also exhibited notable enhancements; however, the degree of improvement was less pronounced compared to positional accuracy. Specifically, the mean orientation error for the 30° merged data decreased from 8.53° initially to 7.65° after all enhancements, while the 15° data saw a reduction from 9.59° to 7.1°. These findings underscore the efficacy of downsampling in refining the precision of PCs. Given the objective of manipulating the globe valve using a robotic hand, an error of less than 5 mm is deemed acceptable for accurate position prediction. From Figure 6, it can be seen that more than 91% of the 30° errors and more than 98% of the 15° errors are less than 5 mm.

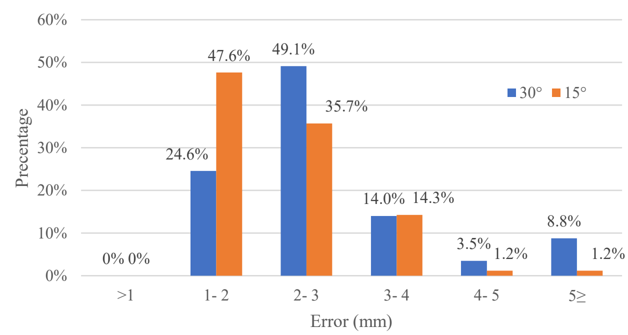

Fig. 6. Error distribution for position prediction

Figure 7 illustrates the orientation difference between the reference orientation provided by the RANSAC plane and the predicted orientation provided by the RANSAC circle. It is noteworthy that there is minimal orientation discrepancy between the initial (target) and final (downsampling) PCs in both scenarios (15° and 30° merging), a fact supported by the results presented in Table I. Initially, the mean error and standard deviation for orientation prediction are 8.53 ° and 3.7°, respectively, for 30° merging. These values undergo reductions of 10.3% and 12.7%, resulting in a final PC mean orientation error of 7.65° and a standard deviation of 3.23°. However, in the case of 15° merging, the accuracy improvement increases by 26.5% and 13% for mean error and standard deviation, respectively. This may be attributed to the smaller angle difference of merged PCs at 15° compared to 30°. The improvement in orientation accuracy is not substantial with this proposed methodology. This could be attributed to the high density of merged PCs from color ICP, potentially leading to orientation errors in some merged PCs.

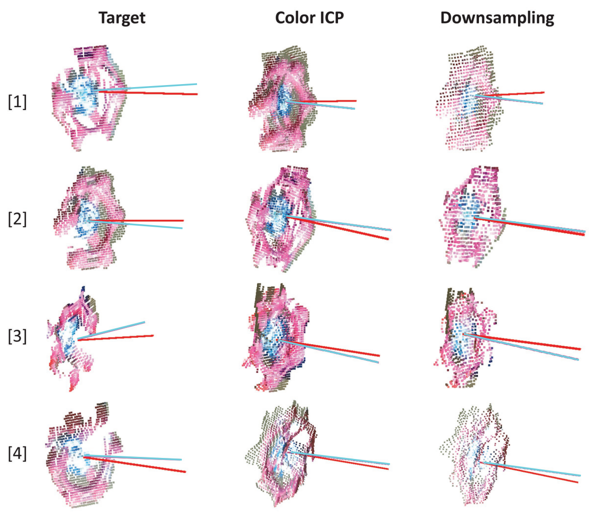

Fig. 7. Calculated orientation for original, color ICP, and downsampling. The light blue line depicts the reference orientation, while the red line represents the predicted orientation.

IV. CONCLUSIONS AND FUTURE WORKS

In this study, a novel process for reconstructing noisy PC data was proposed, leveraging color ICP algorithms, subsequent downsampling, and PC capturing with a systematic robotic arm with an RGB-D camera. Through the analysis of 216 instances across three datasets, encompassing both 15° and 30° PC merging scenarios, significant improvements in both position and orientation accuracy were achieved. For position prediction, mean errors were reduced by approximately 93%, with final mean errors below 5 mm, meeting the criteria for accurate position estimation in applications such as robotic manipulation in valves. Meanwhile, orientation accuracy improvements varied, with smaller angle differences leading to more substantial enhancements. Despite minimal improvements in orientation accuracy, the proposed methodology demonstrated its efficacy in refining noisy PC data.

The merging process sometimes results in a denser PC, which decreases the orientation accuracy of the final output. It was observed that the color ICP algorithm struggles to produce a refined output when there are significant differences in the angles of the two merging PCs. To address this challenge, enhancing the color ICP merging algorithm by incorporating position and orientation data from the robot arm could be implemented in future research. Computational efficiency of the approach evaluating the execution time and comparing it with other methods will be done in the next stage.

REFERENCES

[1] Guoguang Du, Kai Wang, Shiguo Lian, and Kaiyong Zhao. Vision-based robotic grasping from object localization, object pose estimation to grasp estimation for parallel grippers: a review. *Artificial Intelligence Review*, 54(3):1677–1734, 2021.

[2] Andreas ten Pas, Marcus Gualtieri, Kate Saenko, and Robert Platt. Grasp pose detection in point clouds. *The International Journal of Robotics Research*, 36(13-14):1455–1473, 2017.

[3] Matthew Berger, Andrea Tagliasacchi, Lee M Seversky, Pierre Alliez, Gael Guennebaud, Joshua A Levine, Andrei Sharf, and Claudio T Silva. A survey of surface reconstruction from point clouds. In *Computer graphics forum*, volume 36, pages 301–329. Wiley Online Library, 2017.

[4] Matthew Berger, Andrea Tagliasacchi, Lee M. Seversky, Pierre Alliez, Joshua A. Levine, Andrei Sharf, and Claudio T. Silva. State of the art in surface reconstruction from point clouds. *35th Annual Conference of the European Association for Computer Graphics, Eurographics 2014 - State of the Art Reports*, 2014.

[5] Chi-Yi Tsai and Shu-Hsiang Tsai. Simultaneous 3d object recognition and pose estimation based on rgb-d images. *IEEE Access*, 6:28859–28869, 2018.

[6] Udaka A. Manawadu, Shishiki Keito, and Naruse Keitaro. Object recognition and pose estimation from rgb-d data using active sensing. In *2022 IEEE/ASME International Conference on Advanced Intelligent Mechatronics (AIM)*, pages 165–170, 2022.

[7] Rogério Yugo Takimoto, Marcos de Sales Guerra Tsuzuki, Renato Vogelaar, Thiago de Castro Martins, André Kubagawa Sato, Yuma Iwao, Toshiyuki Gotoh, and Seiichiro Kagei. 3d reconstruction and multiple point cloud registration using a low precision rgb-d sensor. *Mechatronics*, 35:11–22, 2016.

[8] Kun Wei, Yong Dai, and Bingyin Ren. Automatic identification and autonomous sorting of cylindrical parts in cluttered scene based on monocular vision 3d reconstruction. *Sensor Review*, 39(6):763–775, 2019.

[9] Lu Sang, Björn Häfner, Xingxing Zuo, and Daniel Cremers. High-quality rgb-d reconstruction via multi-view uncalibrated photometric stereo and gradient-sdf. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)*, pages 3106–3115, January 2023.

[10] P.J. Besl and Neil D. McKay. A method for registration of 3-d shapes. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 14(2):239–256, 1992.

[11] Ji Hoon Joung, Kwang Ho An, Jung Won Kang, Myung Jin Chung, and Wonpil Yu. 3d environment reconstruction using modified color icp algorithm by fusion of a camera and a 3d laser range finder. In *2009 IEEE/RSJ International Conference on Intelligent Robots and Systems*, pages 3082–3088, 2009.

[12] Jaesik Park, Qian-Yi Zhou, and Vladlen Koltun. Colored point cloud registration revisited. In *2017 IEEE International Conference on Computer Vision (ICCV)*, pages 143–152, 2017.

[13] Yang Yang, Weile Chen, Muyi Wang, Dexing Zhong, and Shaoyi Du. Color point cloud registration based on supervoxel correspondence. *IEEE Access*, 8:7362–7372, 2020.

[14] Weile Chen, Yang Yang, Dandan Fan, Zhuo Chen, and Qian Kou. Registration of color point cloud by combining with color moments information. In *2018 IEEE International Conference on Systems, Man, and Cybernetics (SMC)*, pages 2102–2108, 2018.

[15] Martin A. Fischler and Robert C. Bolles. Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography. *Commun. ACM*, 24(6):381–395, jun 1981.

[16] Leonid Keselman, John Iselin Woodfill, Anders Grunnet-Jepsen, and Achintya Bhowmik. Intel realsense stereoscopic depth cameras. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, July 2017.

[17] Anders Grunnet-Jepsen and Dave Tong. Depth post-processing for intel® realsense™ d400 depth cameras. *New Technologies Group, Intel Corporation*, 3, 2018.

[18] Jun Tang. A color image segmentation algorithm based on region growing. In *2010 2nd International Conference on Computer Engineering and Technology*, volume 6, pages V6–634–V6–637, 2010.

[19] Guichao Lin, Yunchao Tang, Xiangjun Zou, Juntao Xiong, and Yamei Fang. Color-, depth-, and shape-based 3d fruit detection. *Precision Agriculture*, 21:1–17, 2020.

[20] Xian-Feng Han, Jesse S. Jin, Ming-Jie Wang, Wei Jiang, Lei Gao, and Liping Xiao. A review of algorithms for filtering the 3d point cloud. *Signal Processing: Image Communication*, 57:103–112, 2017.

[21] Jeongsoo Lim, Inho Lee, Inwook Shim, Hyobin Jung, Hyun Min Joe, Hyoin Bae, Okkee Sim, Jaesung Oh, Taejin Jung, Seunghak Shin, et al. Robot system of drc-hubo+ and control strategy of team kaist in darpa robotics challenge finals. *Journal of Field Robotics*, 34(4):802–829, 2017.