One-Shot Accurate Cylinder Pose Estimation From Point Cloud Data With Density-Based Geometric Clustering

Ayato Ohashi and Keitaro Naruse

Abstract-In recent years, the advent of machine learning technologies has significantly increased interest in factory automation (FA). Pose estimation, a crucial process in binpicking within FA, has been explored extensively by researchers worldwide. This technology has applications in various fields including computer graphics (CG), virtual and augmented reality (VR/AR), and robotics. This paper specifically addresses the pose estimation of cylinders using a single point cloud (PC), a challenging problem due to potential ambiguities when the PC captures both the base and side of a cylinder, which can significantly impact the accuracy of pose estimation. To address this, we propose a geometric density-based clustering approach centered on the cylinder axis as the critical feature. Our method involves three steps: first, performing probability density estimation using two Gaussian spheres based on the normals and cross-products of the PC, applying directional kernel density estimation (DKDE). Second, choosing the dominant aspect either the base or the side through a point-topoint matching process to estimate the center point. Finally, conducting aspect clustering using an in-out circle created by cosine similarities to utilize the estimated cylinder axis. The center point is then determined either as the average of the PC or by the least square circle fitting, depending on the dominant aspect identified. Our approach demonstrates precise one-shot pose estimation results using a single PC.

I. INTRODUCTION

Bin-picking is a crucial task in factory automation (FA), significantly impacting the efficiency of assembly lines. It involves three key steps: object detection, pose estimation, and path planning. Initially, target objects captured by an RGB-D camera are identified. Subsequently, the orientation and position of these objects are estimated. Finally, they are grasped by a robotic arm and placed in a designated area. The advent of deep learning technologies has facilitated these tasks; however, challenges such as dynamic environments, real-time processing, and precise operation persist. This research focuses on developing an accurate pose estimation system to address the need for precise operations.

In pose estimation, a combination of color, depth, or point cloud (PC) information is typically utilized. PCs are defined as point-wise data representations in a Cartesian coordinate system, including color data from RGB-D cameras. A traditional method, the Iterative Closest Point (ICP) [1], is often used for pose estimation and has been widely applied in simultaneous localization and mapping (SLAM) to address registration issues, effectively matching point-wise data to a predefined target object. Despite its high accuracy, ICP requires considerable computation time. Alternatively, color information can be utilized to detect key points from 2D images and match them with known object models, such as 3D CAD models [2]. Moreover, convolutional neural network (CNN) [3] and PointNet [4] based methods represent the latest trends in machine learning technologies for pose estimation.

Our research aims to accurately estimate the pose of cylinders using a single PC from an arbitrary camera viewpoint, a notably challenging problem due to the variability in PC appearance. Our approach differs from previous works as our PCs may include both the base and side of a cylinder. The dominance of these features varies depending on the viewpoint, which complicates accurate pose estimation using conventional methods that typically rely on side PCs. Our method addresses cylinders of varying dimensions, from thin, long pipes to thick, short discs, ensuring our approach is not limited by specific dimensional ratios. To address these challenges, we propose a three-step solution: firstly, conducting aspect clustering between the base and side; secondly, selecting the dominant aspect to accurately estimate the center point; and finally, estimating the cylinder axis using the dominant cluster. Notably, this research does not use color information or machine learning techniques due to the need for extensive modifications of color values and the substantial computation required for data annotation, augmentation, and training.

II. PREVIOUS WORKS

A cylinder is one of the primitive shapes frequently used in manufacturing industries. There are a lot of previous works concerning pose estimation of cylinders, and they are generally divided into two categories in terms of whether to use surface normals [5]–[7]. In this section, methods that utilize a normal vector at each point are termed "Normal-Based Methods," while others are called "Non-Normal-Based Methods." Our approach falls into the former category. Commonly employed techniques in cylinder pose estimation include principal component analysis (PCA), random sample consensus (RANSAC), Hough transform (HT), and least squares (LS), with the LS circle fitting method utilized in our research for center point estimation.

A. Normal-Based-Methods

The method for extracting cylinders from cylindrical objects was proposed early on [5]. It involves two main steps: initially, projecting a normal at each point of a cylinder

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

Ayato Ohashi and Keitaro Naruse are with the Graduate School of Computer Science and Engineering, University of Aizu, Fukushima, Japan m5261148@u-aizu.ac.jp, naruse@u-aizu.ac.jp

onto a unit sphere termed Gaussian mapping is performed, which is beneficial for characterizing object shapes. Secondly, a cylinder axis is decided to make a random subset of normal by RANSAC in the Gaussian map. Another approach simplifies the 5D HT into 2D and 3D to manage the computational complexity, where normals are projected onto a semi-sphere, and great circles are computed iteratively from the intersection points with the origin [6]. The point that intersects the most between all great circles is estimated as the cylinder axis. For shape detections, the method with the underlying cylinder feature is adapted [8]. Minimal subsets of normals are estimated by RANSAC and the orientation is equal to the cross product between normal from the side of cylinder. The score function is employed as well to recognize each shape because it is expected to be other primitive shapes such as planes, spheres, and cones. Multiple cylinder extractions at the same time are presented with region-growing clustering [7]. As the covariance matrix of normal on the Gaussian sphere, the eigenvector of PCA is computed and the smallest eigenvector can be regarded as the cylinder orientation because it is obvious if normals are from the side of the cylinder, they are supposed to be on the arc of a cylinder.

B. Non-Normal-Based Methods

For cylindrical object detections in facilities, a circle slicebased approach is presented with HT [9]. As the cylinder orientation is pre-defined as the x, y, or z-axis, the circle slicing procedure with the known circle radius is conducted along with the direction. Hough voting is used for the approximation of a circle. Finally, cylindrical objects are extracted. A robust cylinder fitting method is proposed with a robust PCA (RPCA) method [10]. The robust covariance matrix of a PC is calculated to solve the weakness toward outliers or noise in the conventional PCA. Thus, the cylinder position and direction are found accordingly. A fast cylinder shapes matching method against the large-scale PC is proposed [11]. curvatures are calculated with PCA and kneighbors. The curvature is the feature that the shape of the PC shows whether to be curved or straight line. For robust cylinder estimation in PCs, the slicing-based approach is proposed [12] like the previous work [10]. Candidate cylinder axes and ellipses are generated by slicing the PC with RANSAC. Next, the cylinder similarity function is used to extract the most possible cylinder out of all candidates.

III. APPROACH

In the proposed approach, the geometric principle concerning the cylinder axis provides solutions for pose estimation. The cylinder axis can be described using two properties consistently. Firstly, the unit normal of the base PC n^b can describe the orientation as shown in Eq. 1. Secondly, the unit cross product of the side PC n^s can describe the orientation as shown in Eq. 2. Hence, it is straightforward that the normal from the base and the cross product between two normal from the side are equal Eq. 3. Fig.1 shows the illustration of the following Eq. 1 to 3.

$$\boldsymbol{n} = \boldsymbol{n}_k^b, \tag{1}$$

$$\boldsymbol{n} = \boldsymbol{c}^{s} = \frac{\boldsymbol{n}_{i}^{s} \times \boldsymbol{n}_{j}^{s}}{\|\boldsymbol{n}_{i}^{s} \times \boldsymbol{n}_{j}^{s}\|} \ (i \neq j), \tag{2}$$

$$\boldsymbol{n}^{b} = \boldsymbol{c}^{s}.$$
 (3)



Fig. 1. Two Gaussian spheres are shown via Gaussian mapping. (a) normals from the base PC. (b) normals and the cross-product from the side PC.

This principle allows for the estimation of a cylinder axis using a PC obtained from an RGB-D camera at an arbitrary viewpoint. From this property, two ideas are considered for the estimate of an accurate center point. Firstly, the aspect clustering is conducted using the cylinder axis as the key point. Secondly, the dominant aspect is chosen out of the base and side to analyze the cylinder axis from normals and cross products. If the cylinder axis derived from normals is dominant, the base aspect is expected to be predominant, and vice versa. Hence, normals correspond to the base aspect and cross-products to the side aspect, based on the cylinder axis.

The approach is organized as follows. Firstly, the PC of a cylinder is used as the input in Fig. 2(a). Two Gaussian spheres of normals and cross-products are calculated in Fig. 2(b). Each density sphere from two Gaussian spheres is estimated with directional kernel density estimation (DKDE) [13] in Fig. 2(c). Based on these spheres, points with high density in one density sphere and the corresponding aspect are chosen with the point-to-point matching process in Fig. 2(c). To use the estimated cylinder axis from the sphere, the aspect clustering is performed in Fig. 2(d), and the result is shown in Fig. 2(e). Finally, the center point is estimated with the dominant cluster.

A. Gaussian Sphere Estimate

This step involves calculating two Gaussian spheres representing normals and cross products, centered at the origin o and normalized to a unit norm. Firstly, outward normals which have the outside orientation are calculated. As a conventional approach, a normal flipping process is performed to ensure the normals are outward-facing [14]. However, there is the case that normals can not be flipped appropriately due to the multicollinearity problem to estimate



Fig. 2. Flowchart of the proposed clustering approach: (a) PC of a cylinder, (b) Gaussian spheres of normals and cross products, (c) Density estimate with DKDE and choice of dominant aspect, (d) Aspect clustering with in-out circle, (e) PC after clustering.

normals using singular value decomposition (SVD) from kneighbors, especially when the alignment of neighbors is nearly collinear. To overcome this problem, the adaptive kneighbors algorithm is employed.

The density-based Spatial Clustering of Applications with Noise (DBSCAN) algorithm is adapted to discard sparse side points that can cause the above problem [15]. The radius of the sphere and the number of points within the radius are predefined as thresholds. Moreover, a cluster with the largest number of points is only used for the normal estimate. Secondly, the initial normal is estimated as the eigenvector v_3 corresponding to the smallest eigenvalue λ_3 from SVD. As the matrix of PC of the cylinder is M, SVD is described using three matrices that are left singular vectors U, singular values Σ and right singular vectors V as shown in Eq. 4,

$$M = U\Sigma V. \tag{4}$$

Adaptive k-neighbors algorithm from kd-tree that iteratively updates k neighborhood size (initial k=10) and reestimates a normal if the contribution rate of the principal component is close to one to avoid the situation that the second and third largest eigenvalues are close to zero in Eq. 5,

$$k = \begin{cases} k+1, & \text{if } \frac{\lambda_1}{\lambda_1 + \lambda_2 + \lambda_3} \approx 1\\ k, & \text{otherwise.} \end{cases}$$
(5)

Thirdly, the conventional normal flipping process [14] is performed to obtain the outward normal consistently if the inner product is a negative value. Supposing p_i is the point of the cylinder, the process is shown in Eq. 6. Therefore, the outward normal is estimated in Fig. 3(b) compared with Fig. 3(a),

$$\boldsymbol{n}_{i} = \begin{cases} -\boldsymbol{n}_{i}, & \text{if } -\boldsymbol{p}_{i} \cdot \boldsymbol{n}_{i} < 0\\ \boldsymbol{n}_{i}, & \text{otherwise.} \end{cases}$$
(6)

Additionally, 10,000 cross products c_i are calculated using two normals, excluding the normal itself, selected through a



Fig. 3. Difference of normal orientations is shown. (a) Normals from ordinary *k*-neighbors. (b) Normals from adaptive *k*-neighbors.

random sampling process. This method is chosen to optimize computation time.

B. Density Estimate

If the raw normal and cross product are directly used for cylinder axis seeking, points which do not describe the cylinder axis can be estimated wrong as the cylinder axis. Thus, normal density sphere (NDS) and cross product density sphere (CDS) using Gaussian spheres of normals and cross products are created with DKDE [13] for an effective representation. Additionally, the density range is divided into five intervals, as depicted in Fig. 4. The unit vectors x with three-dimensional coordinates on a sphere can be described by a Cartesian coordinate system,

$$\Omega_2 = \{ \boldsymbol{x} \in \mathbb{R}^3 : \| \boldsymbol{x} \| = x_1^2 + x_2^2 + x_3^2 \}.$$
(7)

Let $x_1, x_2, ..., x_n \in \Omega_2$ be unit vectors by a probability density function f(x) on Ω_2 ,

$$\hat{f}(x) = \frac{c_h(K)}{n} \sum_{i=1}^n K\left(\frac{1 - \boldsymbol{x}^T \boldsymbol{x}_i}{h^2}\right),\tag{8}$$

where h is the bandwidth only to take positive value $[0, \infty]$ for an approximation of the density. K is the directional

kernel function $K(r) = e^{-r}$ that describes monotonically decreasing and $c_h(K)$ is a normalizing constant.



Fig. 4. Density sphere and color bar of density interval.

C. Aspect Choice

As the way of choosing the dominant aspect, each distribution of density spheres and maximum density will be analyzed as necessary. For seeking the distribution of the cylinder axis, the point-to-point matching process between NDS and CDS is performed. In the process, the dominant aspect and yellow points, have the highest density interval, in one density sphere only to describe cylinder axis are chosen to check the number of yellow points of NDS corresponding to CDS. Additionally, suppose yellow points in CDS are chosen. In that case, true yellow points that refer to the correct orientation of the cylinder are chosen to distinguish the opposite one for the cylinder axis estimate using Eq. 9 as v_f is the cylinder axis from the front viewpoint. However, the cylinder orientation can not be found in the parameter of θ is almost the same as 90 degrees. In this case, the known correct orientation is used,

$$\boldsymbol{p}_{y}^{true} = \begin{cases} \boldsymbol{p}_{y}^{i}, & \text{if } -\boldsymbol{v}_{c} \cdot \boldsymbol{p}_{y}^{i} > 0\\ \emptyset, & \text{otherwise.} \end{cases}$$
(9)

In special cases where there are no corresponding yellow points, the cylinder axis distribution might appear exclusively on one of the density spheres. Here, a comparison of each sphere's maximum density is conducted.

Fig. 5 illustrates two informative features concerning the maximum density. Firstly, the maximum density of NDS gradually decreases as the number of side points is large because the distribution varies the concentrated to scattered in NDS. On the other hand, the maximum density of CDS gradually increases because the distribution varies from the scattered to the concentrated in CDS. Secondly, the magnitude relationship of these maximum densities is flipped by the ratio of base and side. Therefore, the aspect with the larger maximum density is chosen as the dominant one. Three cases are shown as follows.

- 1) Case A: Yellow points in NDS and base PC
- 2) Case B: Yellow points in CDS and side PC

 Case C: Case A or B based on the maximum density comparison



Fig. 5. Change of density distributions of NDS and CDS.

D. Aspect Clustering

Firstly, the cylinder axis is estimated as the point with the maximum density in the selected yellow points. For the aspect clustering, the in-out circle is computed using the estimated cylinder axis and the point with minimum density in yellow points. If normal is within the in-out circle, the corresponding PC is classified as the base cluster. Otherwise, the corresponding PC is classified as the side cluster.

E. Center Point Estimate

One of two existing methods depending on the dominant aspect is used for the center point estimate. If the base cluster is chosen to use, the average of the base cluster is calculated based on the assumption that the base cluster will be within the base. In contrast, if the side cluster is chosen to use, the LS circle fitting is performed after projecting the side cluster on a plane formed by the farthest point from the estimated cylinder axis and cylinder axis because of making use of the feature that the side cluster is on the arc of a cylinder.

IV. EVALUATION

A. Experimental Setting

As illustrated in Fig. 6, the experimental setup involves a cylinder model and an RGB-D camera for point cloud (PC) acquisition, using Choreonoid [16], a robot simulator. The camera is positioned to rotate around the center point of the cylinder, which serves as the center of rotation. For the purpose of evaluating the clustering model, the base of the cylinder is painted red, and the side is painted blue. Since a simulator is used, there is no inherent measurement error in the camera's capture process. To assess the robustness of the proposed method against noise, Gaussian noise is added to the acquired PC. We explore the accuracy of the

method under various conditions by adjusting the following parameters:

- θ ∈ {0, 5, 10, ..., 90} : Degree from the initial camera position by 5 [deg]
- $H \in \{10, 50, 100, 500\}$: Height of cylinder [mm]
- $R \in \{25, 50, 75, 100\}$: Radius of cylinder [deg]
- $D \in \{500\}$: Distance between camera and cylinder [mm]
- $\sigma \in \{0, 0.1, 0.5, 1\}$: Gaussian noise along with the width of the camera coordinate system [mm]



Fig. 6. Illustration of experimental setting in Choreonoid.

B. Evaluation Metrics

For the evaluation of pose estimation, two key metrics are utilized: cosine similarity for the cylinder axis and Euclidean distance for the center point. These metrics assess the error between the estimated values and the ground truth. In assessing the clustering model, three metrics derived from the confusion matrix are employed: accuracy score, precision score, and recall score. These metrics help determine the effectiveness and characteristics of the clustering model.

V. RESULT AND DISCUSSION

There are sixty-four results of pose estimation and clustering. Then they are divided into accurate or inaccurate results concerning each pose estimation accuracy. Fiftynine results are accurate at all angles. However, under some conditions, $R = \{25\}$, $H = \{10\}$, $\sigma = \{0, 0.1, 0.5, 1\}$ or $R = \{25\}$, $H = \{50\}$, $\sigma = \{1\}$, pose estimation results are inaccurate at any angle.

A. Results

According to the results depicted at the top row of Fig. 7, the orientation accuracy is exceptionally high, as indicated by the cosine similarity values being close to one, suggesting minimal error in orientation estimation. The center point estimation is also highly accurate, with the Euclidean distance typically within the expected radius. However, the distance can suddenly be large at 85 degrees although the result is still accurate. The accuracy score can be inaccurate depending on

the strength of Gaussian noise (σ). As the number of side points is large, the accuracy gradually changes to be accurate as well. The precision score shows a quite high accuracy consistently. Recall score is similar to the way of change with accuracy score.

Conversely, the bottom of Fig. 7 illustrates inaccurate results. Here, the cosine similarity falls below 0.5 at angles around 80 to 85 degrees, indicating significant errors in orientation. The Euclidean distance exceeds the radius, reflecting inaccuracies in center point estimation. Additionally, as the angle approaches degrees known for less accurate pose estimation, the clustering results' accuracy declines sharply.

B. Discussion

In clustering results, the factor of the high precision score is written as follows. The base cluster only contains base points because a strict point with the minimum density in yellow points is used for the calculation of the in-out circle. The side cluster contains side points and a part of edge points which have normals with the ambiguous orientation formed by base and side points. Furthermore, the ratio of edge points is basically small. Thus, the precision score is close to one. However, if the height of the cylinder is extremely small, the precision score will be inaccurate because the ratio of edge points will be large.

In pose estimation results, if a small arc or small size of base or side points is obtained, the algorithm will not be useful because the density distribution is inaccurate. Thus, the maximum density comparison is unreliable. In Fig. 8, Case A is wrong chosen although the density distribution does not describe the cylinder axis. In other words, The accurate pose estimation is performed if these contexts are satisfied.

VI. CONCLUSION

The one-shot pose estimation approach with geometric density-based clustering was proposed only using a single PC data. The pose estimation result is accurate under most cylinder dimensions. Then it turned out that the clustering model has the feature to show a high precision score if the result is accurate. As future works, two tasks will be conducted in the future. Firstly, multiple cylinder pose estimation will be conducted. It will be necessary to detect each cylinder first before adapting the proposed approach. Otherwise, the density distribution in the density sphere will be inaccurate. Secondly, the approach will be used against a real cylindrical object. In real situations, noise is not necessarily followed by Gaussian distribution. Thus, the process of noise smoothing or removal can be quite important.

REFERENCES

- Paul J. Besl and Neil D. McKay. Method for registration of 3-D shapes. In Paul S. Schenker, editor, *Sensor Fusion IV: Control Paradigms and Data Structures*, volume 1611, pages 586 – 606. International Society for Optics and Photonics, SPIE, 1992.
- [2] Mathieu Aubry, Daniel Maturana, Alexei A. Efros, Bryan C. Russell, and Josef Sivic. Seeing 3d chairs: Exemplar part-based 2d-3d alignment using a large dataset of cad models. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (CVPR), June 2014.



Fig. 7. The top row shows accurate results of pose estimation and clustering, and the bottom row shows inaccurate results. (a) Cylinder axis: Cosine similarity, (b) Center point: Euclidean distance, (c) Clustering: Accuracy score, (d) Clustering: Precision score, (e) Clustering: Recall score



Fig. 8. Inaccurate NDS and CDS are shown with each corresponding PC. The top is from $R = \{25\}$, $H = \{10\}$, $\sigma = \{0\}$, and the bottom is from $R = \{25\}$, $H = \{50\}$, $\sigma = \{1\}$

- [3] Yu Xiang, Tanner Schmidt, Venkatraman Narayanan, and Dieter Fox. Posecnn: A convolutional neural network for 6d object pose estimation in cluttered scenes. arXiv preprint arXiv:1711.00199, 2017.
- [4] Charles R. Qi, Hao Su, Kaichun Mo, and Leonidas J. Guibas. Pointnet: Deep learning on point sets for 3d classification and segmentation. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), July 2017.
- [5] Thomas Chaperon and Francois Goulette. Extracting cylinders in full 3D data using a random sampling method and the Gaussian image. In Vision Modeling and Visualization Conference 2001 (VMV-01), Stuttgart, Germany, November 2001.
- [6] Tahir Rabbani and Frank Van Den Heuvel. Efficient hough transform for automatic detection of cylinders in point clouds. *Isprs Wg Iii/3*, *Iii/4*, 3:60–65, 2005.
- [7] Trung-Thien Tran, Van-Toan Cao, and Denis Laurendeau. Extraction

of cylinders and estimation of their parameters from point clouds. *Computers & Graphics*, 46:345–357, 2015. Shape Modeling International 2014.

- [8] R. Schnabel, R. Wahl, and R. Klein. Efficient ransac for point-cloud shape detection. *Computer Graphics Forum*, 26(2):214–226, 2007.
- [9] Rui Figueiredo, Atabak Dehban, Plinio Moreno, Alexandre Bernardino, José Santos-Victor, and Helder Araújo. A robust and efficient framework for fast cylinder detection. *Robotics and Autonomous Systems*, 117:17–28, 2019.
- [10] Mahmoud Fouad Ahmed, Carl T. Haas, and Ralph Haas. Automatic detection of cylindrical objects in built facilities. *Journal of Computing in Civil Engineering*, 28(3):04014009, 2014.
- [11] A. Nurunnabi, Y. Sadahiro, and R. Lindenbergh. Robust cylinder fitting in three-dimensional point cloud data. *The International Archives* of the Photogrammetry, Remote Sensing and Spatial Information Sciences, XLII-1/W1:63–70, 2017.
- [12] Young-Hoon Jin and Won-Hyung Lee. Fast cylinder shape matching using random sample consensus in large scale point cloud. *Applied Sciences*, 9(5), 2019.
- [13] ZD Bai, C Radhakrishna Rao, and LC Zhao. Kernel estimators of density function of directional data. In C.R. Rao and M.M. Rao, editors, *Multivariate Statistics and Probability*, pages 24–39. Academic Press, 1989.
- [14] Radu Bogdan Rusu. Semantic 3d object maps for everyday manipulation in human living environments. *KI-Künstliche Intelligenz*, 24:345– 348, 2010.
- [15] Martin Ester, Hans-Peter Kriegel, Jörg Sander, Xiaowei Xu, et al. A density-based algorithm for discovering clusters in large spatial databases with noise. In *kdd*, volume 96, pages 226–231, 1996.
- [16] Shin'ichiro Nakaoka, Shizuko Hattori, Fumio Kanehiro, Shuuji Kajita, and Hirohisa Hirukawa. Constraint-based dynamics simulator for humanoid robots with shock absorbing mechanisms. In 2007 IEEE/RSJ International Conference on Intelligent Robots and Systems, pages 3641–3647, 2007.