

Three-dimensional Morphological Reconstruction of Millimeter-Scale Soft Robots based on Stereo Vision

Tian-Ao Ren, Wenyan Liu, Tao Zhang, Lei Zhao, Hongliang Ren and Jiewen Lai*

Abstract—Continuum robots can be miniaturized to just a few millimeters in diameter. Among these, notched tubular continuum robots (NTCR) show great potential in many robotic surgery applications. Existing robotic modeling methods focus on kinematics and dynamics but neglect the robot's morphology—a significant factor that can expand the research landscape of continuum robots, especially for those with asymmetric continuum structures. This paper proposes a stereo vision-based method for the three-dimensional morphological reconstruction of millimeter-scale NTCRs. The method employs two oppositely located stationary binocular cameras to capture the point cloud of the NTCR, then utilizes predefined geometry as a reference for the KD tree method to relocate the capture point clouds, resulting in a morphologically correct NTCR despite the low-quality raw point cloud collection. The method has been proved feasible for an NTCR with a 3.5 mm diameter, capturing 14 out of 16 notch features, with the measurements generally centered around the standard of 1.5 mm, demonstrating the capability of revealing morphological details. Our proposed method paves the way for 3D morphological reconstruction of millimeter-scale soft robots for further self-modeling study.

I. INTRODUCTION

Soft robots, due to their inherent flexibility and compliance, necessitate innovative sensing methodologies distinct from those used in rigid robots. Vision-based approaches have become increasingly prevalent for measuring various characteristics of soft robots, as traditional sensing methods like joint motor encoders and optical markers are not entirely applicable [1]–[4]. These conventional methods require direct physical interaction with the highly deformable robotic structure, potentially altering its motion and accuracy [5].

Several vision-based methods have been developed to overcome these challenges. For example, Gu *et al.* [4] introduced a neural network-based shape estimation approach that achieves a high precision with an error margin of 2.91 mm given a soft parallel robot with three 175 mm length actuators. This method demonstrates robustness under

This work is supported in part by the IdeaBooster Fund Award under grant 3230391; in part by the Faculty Direct Fund under grant 4055213; in part by the MSc Admission Scholarship of the Department of Electronic Engineering, CUHK; in part by the Hong Kong Research Grants Council (RGC) Collaborative Research Fund under grant CRF C4026-21GF; and in part by the Hong Kong Research Grants Council (RGC) Research Impact Fund under grant R4020-22.

T.-A. Ren is with the Department of Mechanical Engineering, Stanford University, CA, USA. He was with the CUHK Shenzhen Research Institute, Shenzhen, China.

W. Liu, T. Zhang, H. Ren, and J. Lai are with the Department of Electronic Engineering, The Chinese University of Hong Kong, Shatin, NT, Hong Kong, China.

L. Zhao is with the College of Computer Science and Electronic Engineering, Hunan University, Changsha, China. She is also with the CUHK Shenzhen Research Institute, Shenzhen, China.

*Corresponding author. E-mail: jwlai@ee.cuhk.edu.hk

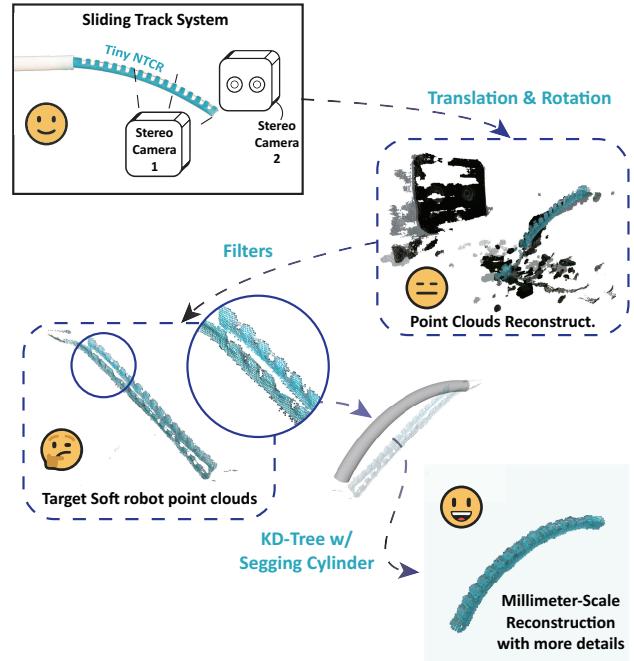


Fig. 1. Overview of the 3D-morphological reconstruction and optimization pipeline for millimeter-scale soft robots. It can be seen that the preliminary reconstructed robot demonstrates two flat faces of point clouds due to the low resolution of stereo camera systems. After applying our method, a refined soft robot with more details and patterns can be obtained.

uncertain observation conditions, making it a significant advancement for real-time shape estimation of soft parallel robots. Zheng *et al.* [6] proposed a convolutional neural network (CNN) regression approach for online markerless key point estimation of deformable robots. Their method outperforms state-of-the-art techniques by up to 4.5% in estimation accuracy, proving effective on various deformable structures such as soft robotic arms and soft robotic fish. Additionally, Lee *et al.* [7] explored vision-based key point estimation for deformable wires in wire harness assembly automation, effectively addressing challenges posed by ultra-thin objects. In [8], the depth vision-based key points estimation is used as the closed-loop positional feedback to enhance the continuum robot's distal tip trajectory tracking, showing the reliability of using the vision-based system for the soft robot's control.

Despite these advancements, the primary focus of current research has been on key point estimation [9], 2D/3D shape [10]–[13] or even tactile [14] and force estimation [15] rather than comprehensive morphological reconstruction. While

these methods are effective for specific tasks, they do not fully capture the detailed morphology of soft robots. This limitation presents a significant gap in the current research landscape, as a detailed morphological understanding is crucial for many applications, particularly in robot's self-modeling [16].

Current related work has been focusing on shape sensing on soft robots. For instance, Zheng *et al.* [6] proposed a CNN-based key point estimation method that excels in accuracy, but is not designed for detailed morphological reconstruction. Lee *et al.* [7] addressed challenges with deformable wires but did not extend their work to the comprehensive morphological reconstruction required for complex soft robots. Our method, by contrast, aims to fill these gaps by providing a practical approach to the full 3D reconstruction of soft robots, addressing the unique challenges associated with their small-scale and complex geometries. The work by Chen *et al.* highlights the benefits of robots creating accurate 3D representations of their morphology and kinematics [16]. Similarly, in [17], the tendon-sheath-driven system can enhance search and rescue operations by enabling precise shape reconstruction. These studies collectively advance robotic autonomy and functionality through innovative 3D reconstruction techniques. Moreover, Sim2Real transfer learning for soft robots can benefit from comprehensive high-resolution simulated data to minimize the reality gap—the discrepancy between simulated models and real-world scenarios [18], and employing generated 3D point clouds that could reflect the real-world geometry can potentially enhance the performance and applicability of the existing Sim2Real framework. However, such tools are inaccessible at the current stage.

To address this gap, we propose a novel stereo vision-based 3D morphological reconstruction method, particularly for millimeter-scale notched tubular continuum robots (NTCR). Our approach employs two oppositely located stationary binocular cameras to capture the 3D point cloud of the NTCR at a very close distance. Due to the limited camera resolution and small size of the NTCR, the captured point clouds are not organized enough to reveal the true morphology of our target robot. By utilizing predefined geometry as a reference and applying the KD tree method to relocate the captured raw point clouds, these scattered point clouds can be more organized to reveal the robot's morphology as well as the pattern. Figure 1 provides an overview of the proposed method.

Specifically, this work contributes a novel stereo vision-based 3D morphological reconstruction method for small notched tubular continuum robots in point clouds form. By leveraging the predefined geometry, we propose a point clouds relocating strategy that involves filtering and KD tree algorithms, streamlining a point cloud-based morphology reconstruction protocol for small continuum robots with a diameter below 3.5 mm. Our method has been rigorously verified with experiments, showcasing the performance in detail morphological reconstruction in comparison to the raw point clouds. Our method provides an alternative to not using expensive point cloud-capturing systems for small objects,

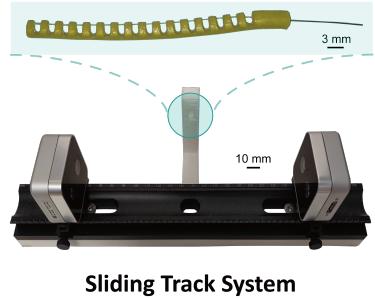


Fig. 2. The experimental setup with the sliding track system and dual Intel RealSense D405 cameras.

which could be used to model other small robots or objects with not just shapes but delicate details.

II. METHODOLOGY

A. Capturing Raw 3D Point Clouds

To capture the raw 3D point clouds for the NTCR, as shown in Fig. 2, we built a dual-stereo vision system by utilizing two Intel RealSense D405 cameras. These two cameras are arranged at a calibrated baseline distance to capture the opposite perspectives of the object (i.e., the small continuum robot). Although the system can become more explicit by using multiple cameras [16], we opt for the minimal required visions to reduce the data redundancy. These cameras generate depth maps, which provide critical spatial information about the object's shape and rough appearance. The depth Z of a point P can be derived from the disparity d between the corresponding points in the left and right camera images, i.e.,

$$d = |X_L - X_R|, \quad (1)$$

$$Z = \frac{bf}{d}, \quad (2)$$

where,

- b is the baseline distance between the two cameras;
- f is the focal length of the cameras;
- d is the disparity.

After that, the depth maps undergo the processing steps, including intrinsic and extrinsic camera calibration to correct distortions and accurately map the 3D coordinates. Subsequently, point cloud data is extracted from the depth maps, forming the basis for further refinement. This structured approach ensures precise 3D reconstruction, which is essential for applications requiring high-fidelity models [19].

B. Conversion and Filtering

The conversion and filtering stage is pivotal for refining raw data obtained from 3D scanning to produce accurate 3D models. Depth maps captured by the stereo vision system are first converted into point cloud data [20]. The transformation from pixel coordinates (u, v) to camera coordinates (X_c, Y_c, Z_c) is given by the intrinsic camera matrix K :

$$\begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = K \begin{bmatrix} X_c/Z_c \\ Y_c/Z_c \\ 1 \end{bmatrix}, \quad (3)$$

where K is the intrinsic camera matrix defined by

$$K = \begin{bmatrix} f_x & 0 & c_x \\ 0 & f_y & c_y \\ 0 & 0 & 1 \end{bmatrix}. \quad (4)$$

This conversion involves transforming 2D pixel coordinates into 3D spatial coordinates using parameters from the calibrated cameras. Once the point clouds are generated, they undergo several filtering processes to enhance data quality. Here, Statistical Outlier Removal (SOR) [21] is utilized to identify and exclude points that significantly deviate from their neighbors, thus preserving data integrity. For each point, the mean distance μ to all its neighbors is computed. Points are considered outliers if their mean distance exceeds a threshold $\mu + \alpha\sigma$, where σ is the standard deviation of distances:

$$\delta_{\text{Threshold}} = \mu + \alpha\sigma. \quad (5)$$

On top of that, three conditional filters are applied to retain points meeting specific criteria, such as (1) spatial coordinates and (2) color values. To further smooth the data, the (3) Moving Least Squares (MLS) filter is employed, fitting local polynomial surfaces to the point cloud to reduce noise while maintaining geometric details. Notably, the MLS filter smooths the point cloud by fitting a polynomial to the points within a neighborhood. The fitted function J minimizes the weighted sum of squared distances from the points to the polynomial surface, viz.,

$$J = \sum_{p \in \mathbb{N}_{>0}} w(\|p - x\|^2) \|f(x) - p\|^2, \quad (6)$$

where $w(\|p - x\|^2)$ is a weight function that decreases with distance.

C. Triangular Patch Construction

Triangular patch construction is essential for converting filtered point cloud data into a continuous and coherent 3D mesh [22]. This process involves identifying neighboring points within the point cloud and connecting them to form triangles, thereby creating a mesh that represents the object's surface. Poisson surface reconstruction [23] is a prevalent technique used in this step, which involves solving the Poisson equation to balance surface smoothness with fidelity to the original point data. The key equations is given by

$$\Delta\chi = \nabla \cdot \mathbf{V}, \quad (7)$$

where, Δ is the Laplace operator, measuring how a function diverges from its average value; χ is the implicit function representing the surface; $\nabla \cdot$ denotes the divergence operator; and \mathbf{V} is the vector field derived from the point cloud data. And the energy function $E(\chi)$ to be minimized is:

$$E(\chi) = \int \|\nabla\chi - \mathbf{V}\|^2 dp, \quad (8)$$

where, $E(\chi)$ is the energy function representing the total error between the gradient of χ and \mathbf{V} ; $\|\nabla\chi - \mathbf{V}\|^2$ is the squared norm of the difference, representing the local error at each point; and dp is the differential element of the domain.

This method generates a ‘watertight’ surface by interpreting the point cloud as samples of an implicit function, effectively filling gaps and ensuring continuity. Poisson reconstruction is particularly adept at handling noisy and incomplete data, as it robustly reconstructs the underlying surface by considering the normal vectors of the points. The resulting triangular mesh accurately depicts the robot's geometry, providing a high-quality 3D model suitable for visualization, and detailed analysis.

III. EXPERIMENT

To achieve accurate 3D reconstruction of millimeter-scale NTCR, a comprehensive experimental setup was devised. This section outlines the detailed experimental setup, the comparison of depth cameras used, and the preliminary results obtained from the scanning and reconstruction process.

A. Experiment Setup

The experimental setup for 3D scanning and reconstruction of millimeter-scale NTCR involves a custom-designed sliding track system equipped with two Intel RealSense D405 depth cameras. The cameras are mounted on adjustable slides that can move along the track, facilitating the calibration and optimization of their relative positions. A 3D-printed rack is placed in between the cameras that allows the NTCR to maintain a fixed position at the robot base, and this base frame can be referred to as the reference frame. The base features a vertical post with small holes for anchoring the robot, ensuring its stability and preventing movement during scanning. This arrangement is crucial for maintaining consistent reference coordinates throughout the scanning process, particularly when the robot's posture is altered by manipulating an internal control wire. As such, a point cloud data collection platform is established. The setup is shown in Figure 2. This setup ensures that the reference coordinates of the robot remain consistent, even when its configuration is altered by manipulating an internal control wire.

B. Depth Cameras Comparison and Calibration

To determine the optimal depth camera for accurate 3D reconstruction, both the Intel RealSense D435i and D405 cameras were evaluated. The cameras were assessed based on their ability to capture detailed point clouds of the same NTCR within a diameter of 3.5 mm (with 1.5-mm width notches), focusing on measurables such as noise levels, resolution, and the clarity of structural features.

The D435i camera exhibited certain limitations, producing point clouds with significant noise and incomplete details, failing to clearly distinguish the robot from the surrounding environment. Conversely, the D405 camera demonstrated better performance, generating point clouds with clear edges, well-defined boundaries, and high-resolution details, particularly capturing the grooves and intricate structures of the robot. The comparison can be seen in Fig. 3. However, according to the technical specification, the depth accuracy of D405 is below 2% at a range of 50 cm, meaning that it could result in an error of 10 mm. Although its performance has been the best among other depth cameras like Helios

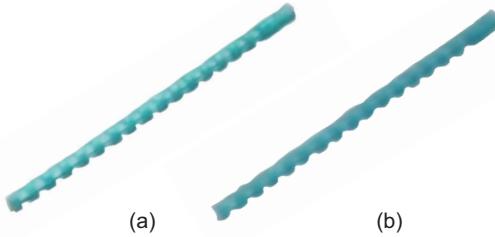


Fig. 3. Comparison of Point clouds (with RGB information) captured by Intel RealSense (a) D405 and (b) D435i.

2, Azure Kinect, and RealSense D435i [24], a 10 mm error in depth estimation would produce poor-quality point clouds for miniature objects. Based on this comparative analysis, the D405 camera was selected for our point cloud data collection platform due to its ability to produce the best depth data with minimal noise.

The extrinsic parameters, which include the rotation matrix R and the translation vector t , transform coordinates from the world coordinate system to the camera coordinate system:

$$\begin{bmatrix} X_c \\ Y_c \\ Z_c \end{bmatrix} = R \begin{bmatrix} X_w \\ Y_w \\ Z_w \end{bmatrix} + t, \quad (9)$$

where R and t can be easily obtained thanks to the adjustable experimental setup as introduced in Sec. III-A.

C. Preliminary Results

Preliminary scanning was performed using our setup with two cameras positioned at an optimal distance of 90 mm from the target continuum robot. Depth maps were simultaneously captured from two perspectives, and point cloud models were generated from each camera's depth data. These point clouds were then aligned and merged using a motion matrix that incorporated the cameras' relative positions and orientations using Eq. (9). Initial data filtering involved removing noise and outliers from the point clouds using Statistical Outlier Removal (SOR) [20], [25] and conditional filters as introduced in Sec. II-B. These filtering techniques significantly improved the data quality. Subsequently, Poisson surface reconstruction was applied to the filtered point clouds to create a continuous and coherent 3D mesh.

The preliminary results, shown in Figure 4, demonstrated that our point cloud data collection platform and processing pipeline could effectively capture and reconstruct the detailed geometry of the millimeter-scale soft robot. The resultant 3D model accurately represented the robot's basic structural features, including its curvature and fine notches/grooves. However, it can also be noticed that the preliminary result shows two separated point cloud sets as they were captured by two cameras in the opposite direction. The cameras' insufficient accuracy in the depth direction also contributes to such a result. The preliminary reconstructed point clouds fail to represent the cylindrical appearance of the tiny robot, which awaits further optimization.

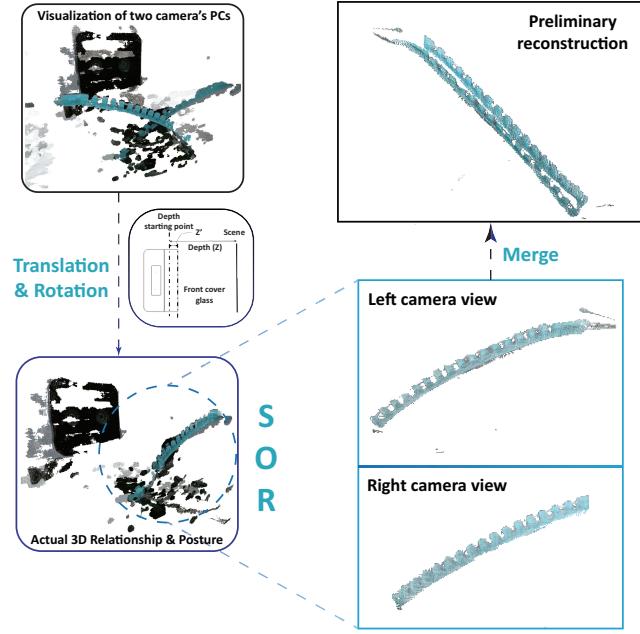


Fig. 4. Preliminary 3D reconstruction results showing the detailed geometry of the NTCR.

IV. OPTIMIZATION

Additional optimization techniques were employed to further refine the 3D reconstruction of the NTCR. This section details the use of KD-Tree for efficient point cloud processing and presents the results with analysis and discussion.

A. KD-Tree

The KD-Tree (K-Dimensional Tree) [26] is a spatial data structure that facilitates efficient nearest neighbor searches, which is useful for processing and optimizing large point cloud datasets. In this study, the KD-Tree was utilized to refine the alignment and accuracy of the reconstructed preliminary NTCR's 3D models. In this method, the Euclidean distance $d(p, q)$ between two points p and q in 3D space is defined as:

$$d(p, q) = \sqrt{(p_x - q_x)^2 + (p_y - q_y)^2 + (p_z - q_z)^2}. \quad (10)$$

To integrate the KD-Tree into our workflow, the following steps were undertaken:

a) Construction: The KD-Tree was constructed using the filtered point cloud data obtained from the initial 3D reconstruction. Each point in the point cloud was inserted as a node in the KD-Tree. This structure enabled efficient spatial querying, which is essential for subsequent processing steps.

b) Nearest Neighbor Search: For each point representing the current scan of the NTCR, the KD-Tree was employed to find the nearest corresponding point in the reference point cloud representing the ideal geometric model [27]. The nearest neighbor search minimizes the $d(p, q)$ between the matched points.

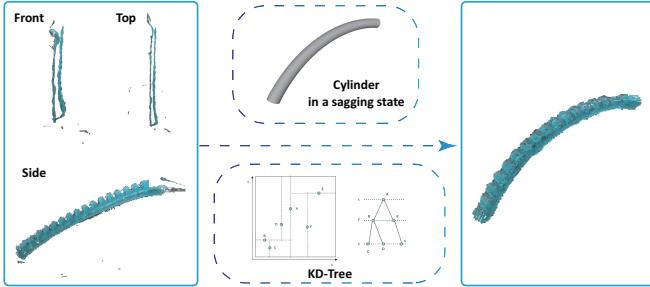


Fig. 5. Optimized 3D model of the NTCR after KD-Tree-based alignment and refinement. It can be seen that, compared to the preliminary result, the optimized model has no flat surfaces due to the re-alignment of the point clouds. The notch patterns are clearer.

c) Point Cloud Alignment: Using the nearest neighbor pairs identified through the KD-Tree, an iterative closest point (ICP) algorithm [28] was applied to align the target point cloud to the reference point cloud. The ICP algorithm aims to minimize the difference between corresponding points in two point clouds [29]. It iteratively solves for the transformation matrix T as

$$T = \arg \min_T \sum_i \|p_i - (Rq_i + t)\|^2 \quad (11)$$

where, p_i denotes the points in the target point cloud; and q_i denotes the corresponding points in the reference point cloud. This alignment involved transforming the target points to minimize the overall distance to their corresponding reference points, improving the coherence of the 3D model.

d) Iterative Refinement: The alignment process was refined through multiple iterations, continuously updating the KD-Tree with the latest transformed point cloud data. This iterative approach ensured that the alignment errors were progressively reduced, resulting in a highly accurate final model.

Figure 5 demonstrates the optimized 3D model of the example NTCR following KD-Tree-based alignment and refinement, showcasing the improved structural details and accuracy.

B. Analysis & Discussion

We compare the point density consistency (PDC) [20] between two point clouds: one from our final optimized algorithm (Point Cloud 1) and another from the non-optimized algorithm before KD-Tree mapping (Point Cloud 2). The metric PDC measures the uniformity of point distribution in a 3D point cloud. It quantifies how evenly the points are spread throughout the desired grid/voxel space. PDC can be computed by

$$\lambda_{\text{PDC}} = \frac{\sigma_{\text{PDC}}}{\mu_{\text{PDC}}}, \quad (12)$$

where σ_{PDC} and μ_{PDC} are the standard deviation of point counts per voxel and the mean of point counts per voxel, respectively. Here, the voxel size for point density calculation is set to 0.5 mm. The mean density and σ_{PDC} for the point clouds are summarized in the following Table I. The optimized algorithm generates point clouds with a higher

TABLE I
POINT DENSITY CONSISTENCY ANALYSIS

Point Clouds	μ_{PDC}	σ_{PDC}	λ_{PDC}
Point Clouds 1 (Optimized)	1.5489	3.9636	2.5590
Point Clouds 2 (Non-optimized)	0.9582	2.2759	2.3752

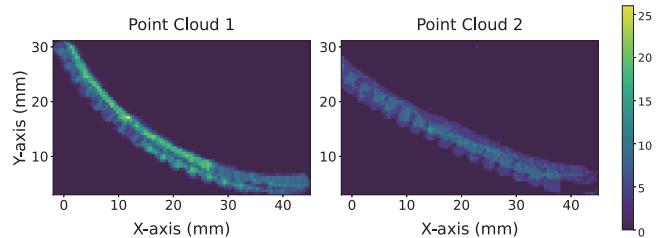


Fig. 6. Point density heatmaps for Point Cloud 1 (Optimized) and Point Cloud 2 (Non-optimized). The color bar represents point density per grid cell (2D), with higher values (color bar, unit: number of points) indicating denser regions.

average point density μ_{PDC} , which contributes to a more detailed model. However, the increased standard deviation suggests that there is more variability in point placement σ_{PDC} . Depending on the application, this variability can either be beneficial (capturing more detailed features) or detrimental (introducing noise). Further refinement of the optimization algorithm could aim to balance high mean density with lower variability, enhancing both the accuracy and consistency of the point cloud.

Our algorithm improves the point cloud density distribution without significantly increasing the computational load λ , thereby capturing more necessary details and enhancing the overall quality of the 3D model. Our result underscores the need for both high point density and uniform distribution to improve point cloud data quality. The following heatmaps, as shown in Fig. 6 illustrate the point density distribution for both point clouds, with higher density regions in Point Cloud 1 (left) compared to Point Cloud 2 (right). We can see that the optimized model exhibits well-defined edges and an accurate representation of the NTCR's structural features, including notches and curvilinear profile. This was corroborated by quantitative metrics such as surface roughness and curvature consistency, which showed marked improvements.

As shown in Fig. 7, the measurements for the notches are generally centered around the standard width of 1.5 mm. The variability is within acceptable limits, with most measurements not deviating significantly from the standard. In our reconstructed structure, out of 16 notches, 14 were observable and measurable in the final results. The result highlights individual variations in the widths of these notches, with some measurements notably above or below the standard value. The boxplot summarizes the data distribution, indicating that most measurements fall within a consistent range around the median. Additionally, the natural bending of the robot during measurement could influence the notch widths, making the overall results quite satisfactory. The statistical data for the 14 measured notches is shown

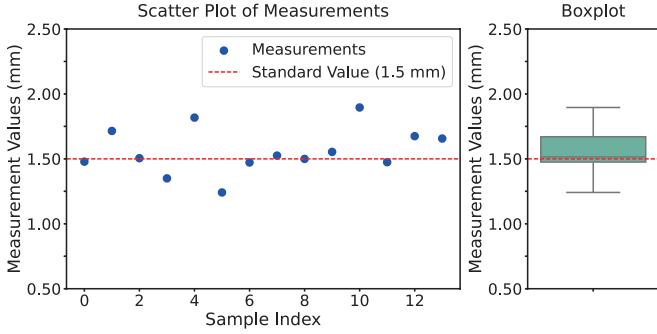


Fig. 7. Measurement results on the visible notches' width, highlighting the comparison with a standard reference width of 1.5 mm. (left) The scatter plot shows individual measurement variations, while the (right) boxplot provides a summary of the width distribution.

below, demonstrating the overall accuracy and precision of our reconstruction method.

The detailed analysis confirms that the KD-Tree-based optimization effectively enhances the 3D reconstruction process by improving accuracy and completeness and reducing noise. These improvements result in a high-fidelity model that accurately represents the intricate appearance of millimeter-scale soft robots, thereby advancing the capabilities for precise surgical interventions and other critical applications.

V. CONCLUSIONS

In this paper, we proposed a stereo vision-based method for the three-dimensional morphological reconstruction of millimeter-scale NTCRs. We developed a dual stereo vision-based point clouds data collection platform specifically for small NTCRs, and it might be applicable to other small objects. A series of filtering and optimization techniques were used to refine the sparsely located 3D point clouds to be more organized with the aim of reconstructing the robot's morphological appearance despite the hardware limitation.

Our work addresses the unique challenges posed by the small-scale and complex geometries of these robots, providing a useful foundation for 3D point cloud collection and post-processing techniques. Future work will focus on real-time reconstruction capabilities and further refinement of the models to enhance their applicability in dynamic and interactive environments, as well as vision-based self-modeling for soft robots.

REFERENCES

- [1] C. Lee, M. Kim, Y. J. Kim, N. Hong, S. Ryu, H. J. Kim, and S. Kim, "Soft robot review," *Int. J. Control Autom. Syst.*, vol. 15, pp. 3–15, 2017.
- [2] M. Hofer, C. Sferrazza, and R. D'Andrea, "A vision-based sensing approach for a spherical soft robotic arm," *Front. Robot. AI*, vol. 8, p. 630935, 2021.
- [3] O. Faris, R. Muthusamy, F. Renda, I. Hussain, D. Gan, L. Seneviratne, and Y. Zweiri, "Proprioception and exteroception of a soft robotic finger using neuromorphic vision-based sensing," *Soft Robot.*, vol. 10, no. 3, pp. 467–481, 2023.
- [4] Y. Rong and G. Gu, "Vision-based real-time shape estimation of self-occluding soft parallel robots using neural networks," *IEEE Robot. Autom. Lett.*, 2024.
- [5] C. Hegde, J. Su, J. M. R. Tan, K. He, X. Chen, and S. Magdassi, "Sensing in soft robotics," *ACS Nano*, vol. 17, no. 16, pp. 15 277–15 307, 2023.
- [6] H. Zheng, S. Pinzello, B. G. Cangan, T. J. Buchner, and R. K. Katzschmann, "Vision-based online key point estimation of deformable robots," *Adv. Intell. Syst.*, p. 2400105, 2024.
- [7] J. S. Lee, D. Kim, M. Roh, and J. B. Kim, "Vision based deformable wires recognition using point cloud in wire harness supply," in *Proc. 2022 13th Int. Conf. Inform. Commun. Techno. Converg. (ICTC)*, 2022, pp. 208–213.
- [8] J. Lai, K. Huang, B. Lu, Q. Zhao, and H. K. Chu, "Verticalized-tip trajectory tracking of a 3d-printable soft continuum robot: Enabling surgical blood suction automation," *IEEE/ASME Trans. Mechatron.*, vol. 27, no. 3, pp. 1545–1556, 2021.
- [9] A. Centurelli, L. Arleo, A. Rizzo, S. Tolu, C. Laschi, and E. Falotico, "Closed-loop dynamic control of a soft manipulator using deep reinforcement learning," *IEEE Robot. Autom. Lett.*, vol. 7, no. 2, pp. 4741–4748, 2022.
- [10] J. Lai, K. Huang, B. Lu, and H. K. Chu, "Toward vision-based adaptive configuring of a bidirectional two-segment soft continuum manipulator," in *Proc. 2020 IEEE/ASME Int. Conf. Adv. Intell. Mechatron. (AIM)*, 2020, pp. 934–939.
- [11] F. Xu, H. Wang, W. Chen, and Y. Miao, "Visual servoing of a cable-driven soft robot manipulator with shape feature," *IEEE Robot. Autom. Lett.*, vol. 6, no. 3, pp. 4281–4288, 2021.
- [12] M. M. Dalvand, S. Nahavandi, and R. D. Howe, "High speed vision-based 3d reconstruction of continuum robots," in *Proc. 2016 IEEE Int. Conf. Syst. Man Cyber. (SMC)*, 2016, pp. 618–623.
- [13] Z. Cui, W. Ma, J. Lai, H. K. Chu, and Y. Guo, "Coupled multiple dynamic movement primitives generalization for deformable object manipulation," *IEEE Robot. Autom. Lett.*, vol. 7, no. 2, pp. 5381–5388, 2022.
- [14] T. Wu, Y. Dong, X. Liu, X. Han, Y. Xiao, J. Wei, F. Wan, and C. Song, "Vision-based tactile intelligence with soft robotic metamaterial," *Mater. Des.*, vol. 238, p. 112629, 2024.
- [15] Z. Zhang, J. Dequidt, and C. Duriez, "Vision-based sensing of external forces acting on soft robots using finite element method," *IEEE Robot. Autom. Lett.*, vol. 3, no. 3, pp. 1529–1536, 2018.
- [16] B. Chen, R. Kwiatkowski, C. Vondrick, and H. Lipson, "Fully body visual self-modeling of robot morphologies," *Sci. Robot.*, vol. 7, no. 68, p. eabn1944, 2022.
- [17] G. Zhang and X. Wang, "Design and 3d reconstruction of a tendon-sheath-driven searching and rescuing robot," in *2013 IEEE International Symposium on Safety, Security, and Rescue Robotics (SSRR)*. IEEE, 2013, pp. 1–6.
- [18] J. Lai, T.-A. Ren, W. Yue, S. Su, J. Y. Chan, and H. Ren, "Sim-to-real transfer of soft robotic navigation strategies that learns from the virtual eye-in-hand vision," *IEEE Trans. Ind. Inform.*, vol. 20, no. 2, pp. 2365–2377, 2024.
- [19] Z. Zhang, "A flexible new technique for camera calibration," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 22, no. 11, pp. 1330–1334, 2000.
- [20] R. B. Rusu and S. Cousins, "3d is here: Point cloud library (pcl)," in *Proc. 2011 IEEE Int. Conf. Robot. Autom.*, 2011, pp. 1–4.
- [21] A. P. Bustos and T.-J. Chin, "Guaranteed outlier removal for point cloud registration with correspondences," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 40, no. 12, pp. 2868–2882, 2017.
- [22] B. Curless and M. Levoy, "A volumetric method for building complex models from range images," in *Proc. 23rd Annu. Conf. on Comput. Graph. Interact. Techniq.*, 1996, pp. 303–312.
- [23] M. Kazhdan, M. Bolitho, and H. Hoppe, "Poisson surface reconstruction," in *Proc. 4th Eurograph. Symp. Geom. Process.*, vol. 7, no. 4, 2006.
- [24] S.-J. Hong, J. Kim, and A. Lee, "Real-time morphological measurement of oriental melon fruit through multi-depth camera three-dimensional reconstruction," *Food Bioprocess Technol.*, pp. 1–15, 2024.
- [25] R. B. Rusu, *Semantic 3D object maps for everyday robot manipulation*. Springer, 2013.
- [26] J. L. Bentley, "Multidimensional binary search trees used for associative searching," *Commun. ACM*, vol. 18, no. 9, pp. 509–517, 1975.
- [27] M. Muja and D. G. Lowe, "Scalable nearest neighbor algorithms for high dimensional data," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 36, no. 11, pp. 2227–2240, 2014.
- [28] S. Rusinkiewicz and M. Levoy, "Efficient variants of the icp algorithm," in *Proc. 3rd Int. Conf. 3D Digit. Imaging Model.*, 2001, pp. 145–152.
- [29] P. J. Best, "A method for registration of 3-d shapes," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 14, pp. 239–256, 1992.