2022 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)
October 23-27, 2022, Kyoto, Japan

Ultrasound Tracking and Closed-Loop Control of a Magnetically-Actuated Biomimetic Soft Robot

Artur João Anjos de Oliveira, Jorge Batista, Sarthak Misra and Venkatasubramanian Kalpathy Venkiteswaran

***Abstract*— Small untethered soft robots have potential for diverse applications, particularly in constrained spaces where the use of a tethered device would be infeasible. Examples include biomedical applications such as brachytherapy, fine-needle biospy and micro-needle drug delivery. To advance soft robots towards these applications, there is a need to establish methods for tracking and control using clinically-relevant methods. This study demonstrates motion planning and magnetic control of a soft untethered robot, using ultrasound images as feedback. The closed-loop control of the Millipede soft robot is first validated using a camera-based tracker, where the deviation between the planned path and the trajectory of the robot is 1.71 mm. Afterwards, two methods for ultrasound-based tracking capable of estimating the pose of the robot are proposed, a geometric approach and a convolutional neural network (CNN), and their performance is compared using a video camera as ground truth. Following this, the CNN method replaces the camera tracker to estimate the position and orientation of the robot. The closed-loop system using ultrasound images guides the robot through the workspace while avoiding virtual obstacles, and achieves an average tracking error of 1.59 mm and an angle error of 2.24°.**

I. Introduction

Untethered magnetically-actuated soft robots can provide potential solutions for applications in constricted environments. The soft, untethered structure of magnetic soft robots allows navigation to difficult-to-reach targets without damaging the surroundings. Moreover, magnetic actuation removes the need for an on-board source of energy in the robot. Medical applications are a particular area of interest for these robots because magnetic actuation at low fields are not harmful for humans [1]. Recent studies have developed untethered soft robots capable of performing tasks such as locomotion, grasping and carrying loads [2]–[4]. Despite developments in the design and actuation of soft robots, localization and closed-loop control remain challenges to be solved.

Recent studies have demonstrated closed-loop control of untethered agents for pick-and-place tasks, using optical camera images [5], [6]. However, in many enclosed environments, such as minimally invasive procedures, localization using optical cameras may not be a viable option. Other works have utilized Magnetic Resonance Imaging (MRI) to track and control magnetic untethered robots [7], [8]. Nevertheless, the use of an MR system for the control of soft robots introduces some drawbacks, such as a high image acquisition time, time-delay between actuation and control systems, and a limited choice of materials for the fabrication of the robot. By contrast, ultrasound (US) is an inexpensive and real-time medical imaging modality. Despite the high signal-to-noise ratio, the high frame acquisition rate of US scanning allows real-time control of robots in clinical environments [9], [10].

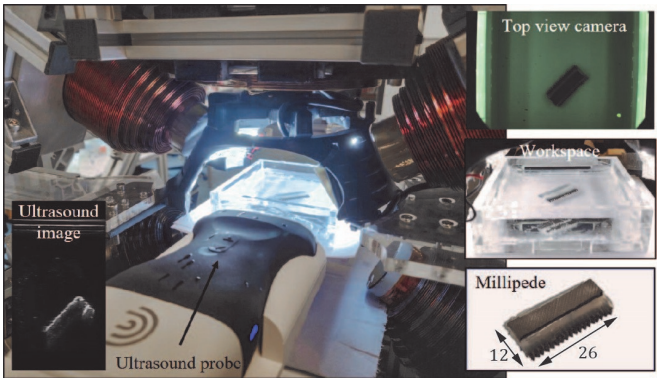

Fig. 1. Closed-loop control of the Millipede, an untethered magnetically-actuated soft robot, using ultrasound images as feedback and a magnetic actuation system. Dimensions are marked in mm.

Previous studies have demonstrated closed-loop control of magnetic robots on the micro-scale, using ultrasound imaging feedback [11], [12]. The tracking of microrobots relies on finding features, such as blobs and contours, to estimate the position of the robots. The geometric tracking algorithms employed in these studies are sensitive to noise and to variations in the intensity of the image pixels. Furthermore, the geometric approaches are susceptible to occlusions because the methods rely on tracking the same shape of the robot, which may not be feasible throughout the entire duration of the experiment.

To increase the robustness of trackers to noise and occlusions, deep learning techniques have been used in computer vision areas for object detection, and regression and classification problems. Recent works have demonstrated the use of Convolutional Neural Networks (CNN) to estimate the pose of objects in different environments [13]. However, none of the approaches have been used for the estimation of the pose

A. Oliveira, S. Misra and V. K. Venkiteswaran are affiliated with Surgical Robotics Laboratory, Department of Biomechanical Engineering, University of Twente, 7500 AE Enschede, The Netherlands. Contact: v.kalpathyvenkiteswaran@utwente.nl

S. Misra is also affiliated with the Department of Biomedical Engineering, University of Groningen and University Medical Centre Groningen, 9713 GZ Groningen, The Netherlands.

A. Oliveira and J. Batista are affiliated with the University of Coimbra, Department of Electrical Engineering and Computers and the Institute of Systems and Robotics, Polo II, Coimbra 3030-194, Portugal.

This research has received funding from the European Research Council (ERC) under the European Union's Horizon2020 Research and Innovation programme (Grant Agreement #866494 - project MAESTRO) and from the Faculty of Engineering Technology at the University of Twente under the project 'Symbiobot'.

2022 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS) | 978-1-6654-7927-1/22/$31.00 ©2022 IEEE | DOI: 10.1109/IROS47612.2022.9981635

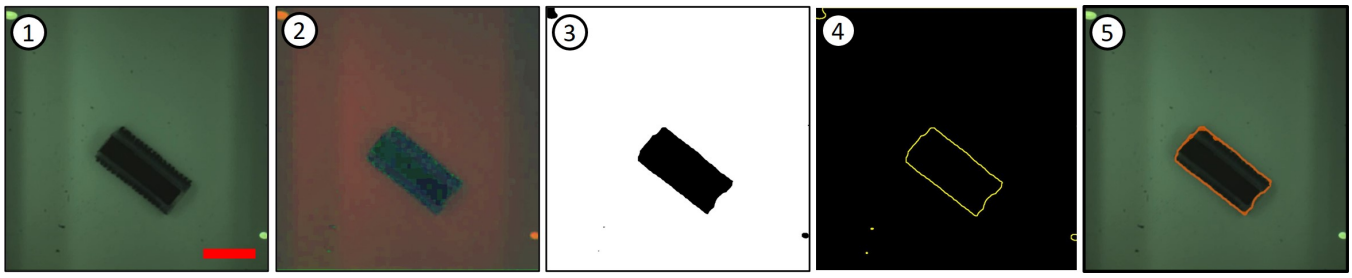

Fig. 2. Image analysis for estimating the position of the robot from camera images. The original Red-Green-Blue (RGB) image ① is converted to a Hue-Saturation-Value (HSV) image ②. ③ Adaptative threshold for obtaining a binary image (black and white), followed by morphological opening, i.e. dilatation of the erosion, is applied to remove small blobs. ④ Edge detection to find the contours in the image. ⑤ The contour of the robot is selected according to the area and shape of the contour. Then the centroid and orientation is estimated using the moments of the contour. The scale bar is 15 mm.

of untethered soft robots during closed-loop control.

In order to facilitate the use of magnetically-actuated soft robots in medical applications, there is a need to achieve robust localization and closed-loop control using clinically-relevant imaging methods. Additionally, compared to previous works on magnetic microrobots, new designs of magnetic soft robots also require sensing of robot orientation to determine the direction of the actuating magnetic field for functions such as cargo delivery, needle biopsy or maneuvering through the workspace [3], [4], [14].

In this study, we demonstrate closed-loop motion control and path planning for a biomimetic magnetic soft robot using 2D ultrasound images. A framework is set up for tracking the soft robot and applying the necessary magnetic field for actuation. Two methods for ultrasound-based tracking capable of estimating the position and orientation of the robot are investigated, and their performance compared using a video camera as ground truth. Both methods are improvements on trackers in the literature that only estimate the position of microrobots. The framework suggests the possibility of using soft robots to autonomously perform tasks in clinically-relevant scenarios.

The rest of the paper is organized as follows. Section II introduces the magnetic actuation system and the soft robot. The methodologies used to detect the robot are described in Sections III and IV, using camera and ultrasound feedback, respectively. Section V presents the path planner and the closed-loop control strategy. Section VI contains the experimental validation. Section VII summarizes the contributions of this work and proposes possible steps for future research.

II. Materials

A. Magnetic Actuation System

BigMag [15] is a magnetic actuation system with a moving array of six electromagnetic coils on two symmetric mobile frames. The two frames rotate around a spherical workspace with a diameter of 10 cm. It is equipped with 2 Dalsa Genie Nano-C1940 (Waterloo, Ontario, Canada) cameras, which provide a top and a side view of the entire workspace. Moreover, BigMag generates any desired magnetic field up to 60 mT with a bandwidth of 40 Hz within the workspace. The magnetic field is used to steer the soft robot through the workspace.

B. Soft Robot

The soft robot used in this study is a biomimetic Millipede robot developed in our previous work [14]. Briefly, the Millipede has two sets of magnetized legs connected by a silicone rubber body. The Millipede achieves locomotion in the presence of a rotating magnetic field. The tilting of the plane of rotation of the magnetic field by an angle results in a different displacement on the two sets of legs. Thus, the turning motion is achieved by controlling the displacement of each set of legs using the tilt angle. An increase of the magnitude of the magnetic field leads to greater displacement, increasing the speed of the Millipede. Details of the design and fabrication process of the soft robot can be found in [14]. For this study, we added a component of silicone rubber mixed with aluminium powder to increase the weight of the robot to achieve underwater motion.

C. Ultrasound Imaging

The ultrasound images are acquired using a wireless Clarius Scanner L15 HD. The Clarius probe acquires the ultrasound images of the workspace using a Clarius App on a smartphone and sends the images to a cast API on a laptop at a resolution of 960×600 pixels (50×70 mm). The detection algorithms are implemented on the cast API and the pose of the soft robot is sent to the BigMag system via an Ethernet cable. The cast API was implemented and developed on a computer running Linux Ubuntu 20.04, equipped with Intel i7-8750H CPU, GeForce GTX 1050 GPU and 16 GB of RAM.

III. Camera-based detection

In this section, the camera-based detection algorithm for the estimation of the pose of the soft robot is described, given by $P = [x, y, \alpha]$. The frames from the top camera of the BigMag system are processed in two stages which consists of a robot detection within the frame, and a position and orientation estimation, as shown in Fig. 2.

The first stage performs the detection of the contours of the soft robot from the binary image (black and white). The input image is a Red-Green-Blue (RGB) colour image acquired by BigMag at a 50Hz rate, and it is converted to the Hue-Saturation-Value (HSV) colour space. Then, channel values are bounded using a user-friendly interface to detect the colour of the robot. Morphological opening, i.e., erosion followed by dilatation, is applied to the binary image to

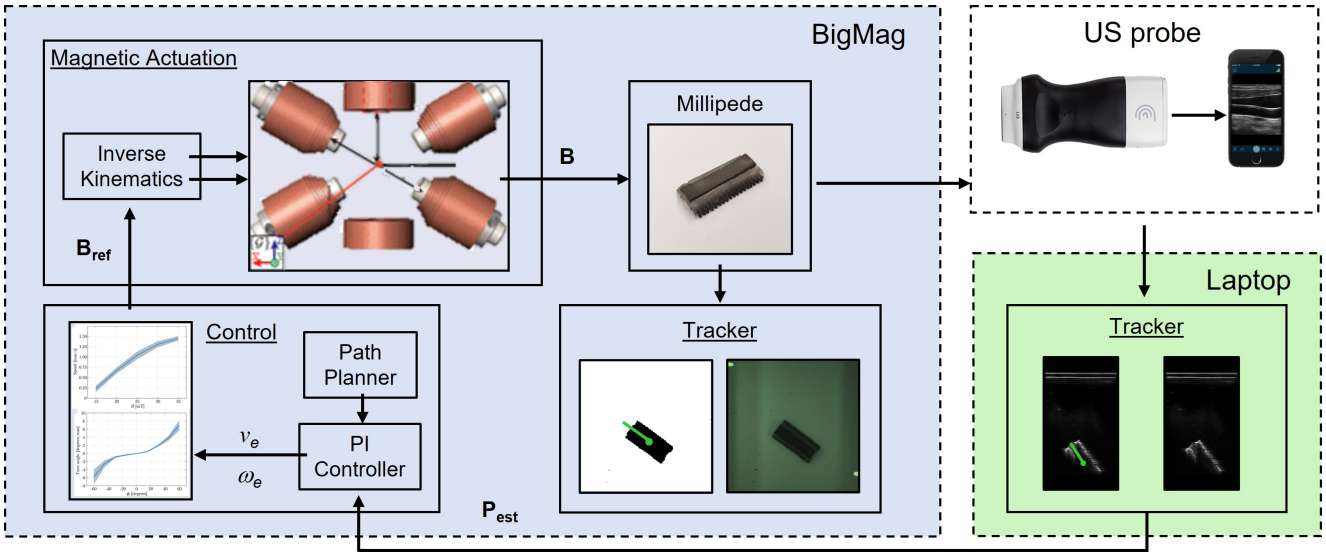

Fig. 3. Closed-loop system for controlling magnetic soft robots using ultrasound imaging. The ultrasound probe images the Millipede and a tracking algorithm estimates the pose of the soft robot ($\mathbf{P}_{est}$). The pose of the robot is sent to the Proportional Integral (PI) controller. Based on the linear and angular velocity errors(v_e and ω_e), the control block outputs the estimated magnetic field ($\mathbf{B}_{ref}$) to drive the soft robot along the path generated. Using inverse kinematics, the magnetic actuation system of BigMag generates the actual magnetic field ($\mathbf{B}$) on the Millipede. The camera tracker defines the ground truth of the pose of the Millipede, and it is not connected to the control block.

remove noise and small blobs, while preserving the shape of the robot. Afterwards, using a contour finding algorithm, the contour of the robot is selected among the others, accordingly to the shape, area, and size.

The second stage consists of processing the soft robot contour points to estimate the position and the orientation. The centroid (x, y) is calculated through the moments of the contours using the Green's Theorem by integrating over the perimeter of the contour. For the orientation estimation, the contours points are fitted to an ellipse and the angle (α) is obtained from the major axis of the ellipse.

IV. Ultrasound-based detection

In this section, we explain the integration of an ultrasound machine into the BigMag system to localize and control the magnetically actuated soft robot to a target, as shown in Fig. 3. Two different approaches are described for the real-time localization of the robot using an ultrasound image. The first approach utilizes a geometric algorithm to retrieve the pose of the robot, while the second approach uses a custom-made convolutional neural network (CNN).

A. Geometric algorithm

The first approach is a two stage algorithm to first perform image segmentation and then detect and estimate the position and orientation of the soft robot, as described in Fig. 4. The pose estimation relies on finding the perpendicular edges which define the rectangular shape of the robot. The first stage is similar to the camera-based algorithm described previously, but the ultrasound image is in the greyscale colour space. Also, the ultrasound image is noisier which requires additional filtering, such as a median filter to remove salt and pepper noise and a Gaussian filter to remove speckle noise and smoothen the image.

The edge points of the image are candidates to fit two perpendicular lines of the rectangular shape of the Millipede. A general GPU implementation of the RANSAC (RANdom SAmple Consensus) is modified to find two perpendicular edges of the Millipede soft robot [16]. The slope of the principal (longer) edge sets the orientation of the robot. Then, the longer and smaller edges define the corner of the rectangle detected, which is used to calculate the position of the robot.

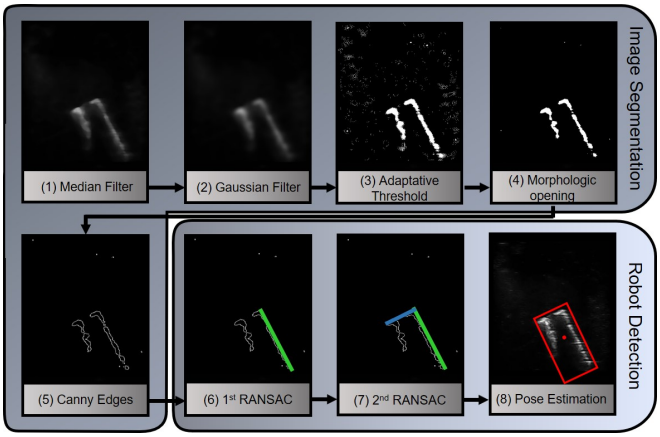

Fig. 4. Image processing for estimating the position of the robot using ultrasound images for the geometric algorithm. The first stage consists of image segmentation (1) - (5) followed by the detection of the robot (6) - (8). (1) (2) Image preprocessing of the ultrasound (US) image for speckle noise reduction and for image smoothing. (3) Adaptative threshold to obtain a binary image (black and white). (4) Morphological opening, i.e. dilatation of the erosion, is applied to remove small blobs. (5) Using Canny Edges, detection of the candidate points for estimating the pose of the robot. (6) First RANSAC iteration to find the principal edge of the soft robot. (7) Second iteration of RANSAC to find the perpendicular edge of the robot. (8) Pose estimation by calculating the centroid and the orientation of the robot.

B. Convolutional neural network

The second approach uses a convolutional neural network to estimate the position and orientation of the robot in the ultrasound image. The process of acquiring the ultrasound

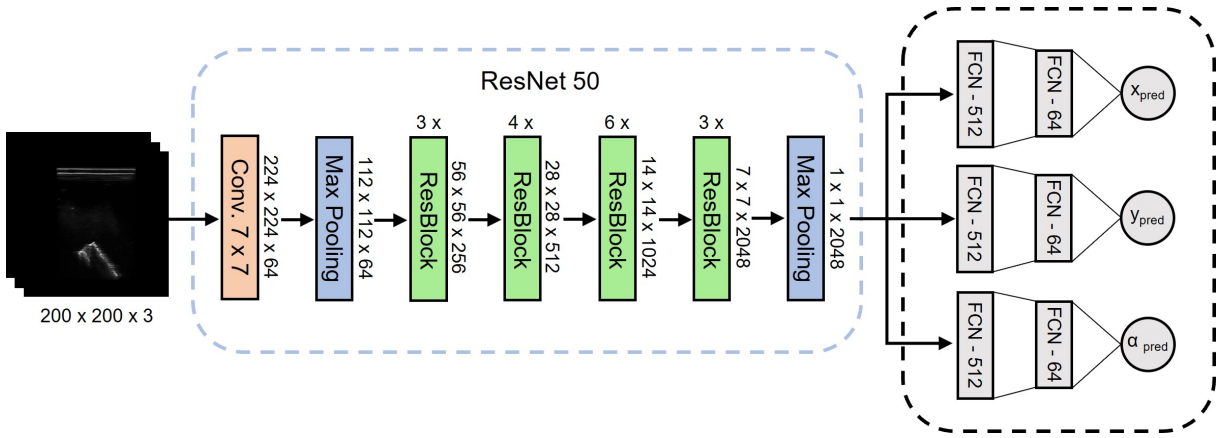

Fig. 5. Architecture of the adapted ResNet 50 convolution neural network (CNN) for pose regression. The proposed CNN uses a ResNet-50 as backbone for feature extraction. The features image vector is used in three separated branches for regression of the pose (x_{pred} and y_{pred}) and the orientation (α_{pred}) of the robot. Each branch is composed by three consecutive fully-connected layers with 512, 64 and 1 outputs.

images and the label for the dataset is described. The architecture of the proposed CNN and the training of the network is also presented.

1) Architecture: The proposed convolutional neural network (CNN) uses a ResNet-50 [17] as backbone, where the input is a greyscale image with a fixed size of 200x200 pixels, as described in Fig. 5. The output of the network is a 3x1 vector with the predicted pose of the robot $(x_{pred}, y_{pred}, \alpha_{pred})$. The ResNet-50 is a 50-layer deep residual network which has five convolutional blocks stacked on top of each other. The output of the last convolutional block is a 7x7x2048 dimensional array which is flattened to a 2048-dimensional vector using a max pooling layer. The output layer to classify the 1000 classes of the ImageNet dataset is replaced by three regression branches. Each branch is composed by a set of 3 fully connected layers, with 512, 64 and 1 output neurons to predict each pose value. The branches for the regression of x and y component utilize a root mean squared error loss function (RMSE) to compute the minimum distance between the predicted and labelled value, v and $\tilde{v}$, respectively. On the other hand, the regression of the angle uses a mean squared error between the angles by representing the angle in the complex plane using $z(a) = (\cos(a), \sin(a))$.

$$L_{RMSE} = \sqrt{\frac{1}{N}\sum_{n=1}^{N}(v_i - \tilde{v}_i)^2} \tag{1}$$

$$\begin{aligned} L_{MSE} &= \frac{1}{N}\sum_{n=1}^{N}\left|z(a_i) - z(\tilde{a}_i)\right|^2 \\ &= \frac{1}{N}\sum_{n=1}^{N}(\cos a_i - \cos\tilde{a}_i)^2 + (\sin a_i - \sin\tilde{a}_i)^2 \end{aligned} \tag{2}$$

2) DataSet: The dataset is constructed with ultrasound images and the pose of the soft robot. To avoid annotating all the video frames manually, we used the pose estimation from the camera-based algorithm to label each ultrasound image. Also, the dataset is augmented by simulating two different types of noise to the ultrasound image data. Gaussian noise with different standard deviations and salt and pepper noise are added to the ultrasound images, while keeping the label of the pose of the soft robot. The addition of noisy images increases the size of the dataset and the generalization of the CNN to ultrasound images with noise. The dataset is constituted of approximately 95000 ultrasound images and the respective labels. The dataset is divided into train(80%), validation(10%) and test(10%).

3) Training Parameters: In the training process, we followed the training of the original ResNet-50 and our data is zero centred by subtracting the mean value. The weights of the backbone of the network are initialized with the converged ResNet-50 on the ImageNet dataset. The fully connected layers of the regression branches are trained from scratch after random initialization. The Adaptative Moment Estimation (Adam) [18] is applied with the default values and a learning rate of 0.01.

V. Motion Planning and Control

The estimated pose P is provided to a path planner to compute a collision free trajectory. From the state-of-the-art motion planners, the A*-WAPP-waypoints path planner is used to feed the control algorithm with the points along the trajectory.

A. Path Planning

The A*-WAPP-waypoints algorithm [19] is a modified version of the A* path planner which reduces the probability of collision with obstacles, while improving the driving time, distance and stability. The path generator utilizes the information of a fitness cost function ($f(n) = g(n) + h(n) + d(n)$) associated with each node n. The goal cost $g(n)$ is the cost required to move from the start node to the n-node. The heuristic cost $h(n)$ of the node is the Euclidean distance from the n-node to the goal node. The distance cost $h(n)$ is inversely proportional to the distance between the node and the obstacle, i.e. the algorithm attributes a higher cost to nodes closer to obstacles. The path search begins in the starting position of the robot and encompasses finding the node with the lowest fitness cost among the eight adjacent nodes of the current searching node. The algorithm creates

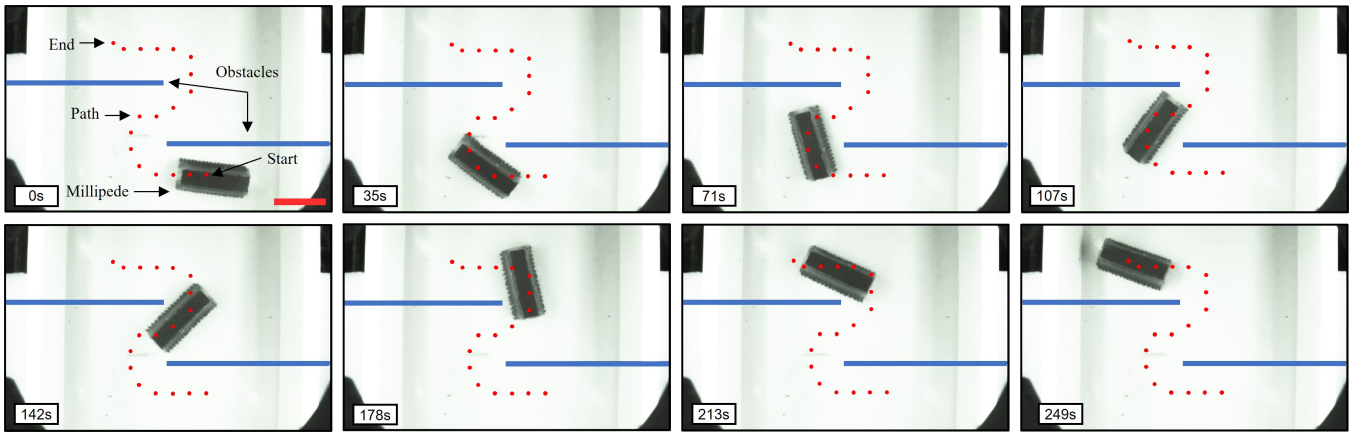

Fig. 6. Experiment demonstrating the closed-loop navigation of the Millipede using camera feedback. The robot follows a path that avoids the obstacles of the workspace to reach the target site. The scale bar is 20 mm. *Please refer to the accompanying video.*

the path when the searching node corresponds to the goal node.

B. Motion Control

The path planner provides the set of waypoints for the implementation of the closed-loop position controller. The controller also receives the pose of the soft robot (x_r, y_r, α_r) from the imaging tracking system. The pose of the robot and each waypoint coordinate (x_p, y_p) are used to estimate the magnetic field needed to steer the soft robot based on previously recorded motion data for the Millipede robot [14].

We control the linear velocity $v(k)$ of the robot using a Proportional Integral (PI) controller:

$$v(k) = k_v e(k) + k_i \int e(k)dt \tag{3}$$

$$e(k) = \sqrt{(x_p - x_r)^2 + (y_p - y_r)^2} - d(k) \tag{4}$$

where $e(k)$ is the Euclidean distance between the current position of the robot and the waypoint of the trajectory, subtracted by a distance $d(k)$. The velocity is controlled by the constant gains k_v and k_i. $d(k)$ is the distance that the robot keeps from the waypoint, when following the waypoints of the path. A second controller uses a proportional controller to turn the robot towards the waypoint:

$$\omega(k) = k_s \ (\beta(k) \perp \alpha(k)), \ k_s > 0 \tag{5}$$

$$\beta(k) = \tan^{-1} \frac{y_p - y_r}{x_p - x_r} \tag{6}$$

where $\beta(k)$ is the angle of the waypoint relatively to the robot. The operator $\perp$ represents the angular difference between the angles. Thus, the angular velocity $\omega(k)$ is the angle to turn the robot to the waypoint multiplied by a constant gain k_s. The PI controller is reset when the value of the error is below 1 mm for the waypoints of the path. The distance $d(k)$ is set to half the length of the robot to allow the turning of the robot and increase the steering stability. However, when the robot is moving towards the final waypoint (target node), $d(k)$ is set to 1 mm, so that the robot accurately reaches the target.

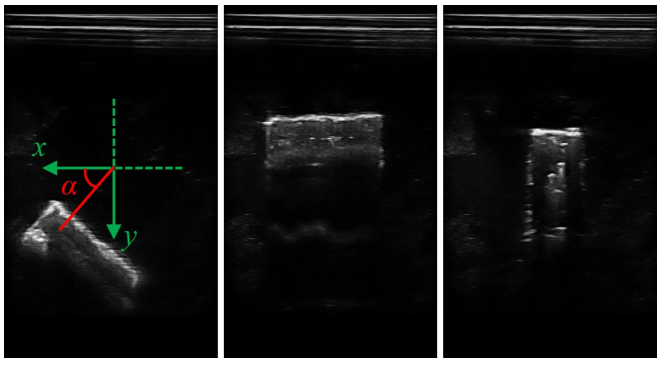

Fig. 7. Representation of the coordinate frame used for the experiments, showing x, y and α. Ultrasound images of the Millipede soft robot for orientations of 145°, 0°, and 90°, respectively, where the silhouette of the soft robot changes with respect to the orientation α. The ultrasound probe is located at the top edge of the images.

VI. Results

In this section, the motion planning and magnetic control of the Millipede soft robot are demonstrated using camera and ultrasound images, including evaluation of the performance of the CNN and geometric trackers methods.

A. Camera tracking experiments

The closed-loop control using camera feedback is demonstrated through an experiment where the soft robot avoids two obstacles to traverse a zig-zag path, as shown in Fig. 6. The path planning algorithm generates a set of waypoints from the start to end node, while the camera images are used to estimate the pose of the robot in real time. The average speed of the robot is 1.61 mm s^{-1}. The mean error of the deviation of the soft robot from the path planner is 1.71 ± 1.07 mm. This experiment validates the motion control and path planning for the Millipede soft robot.

B. Ultrasound tracking experiments

1) Evaluation of the US methods: The proposed ultrasound tracking methods are evaluated with the test dataset containing more than 9500 ultrasound images, which is 10% of the entire dataset. The dataset corresponds to a wide range of positions and orientations of the robot during its motion

TABLE I

AVERAGE TRACKING ERROR OF THE RANSAC AND CNN METHODS IN FOUR QUADRANTS OF THE WORKSPACE.

Coordinates	No. of data points	Axis	RANSAC (mm)	CNN (mm)
x>0 , y>0	2314	x	0.7 ± 1.2	0.2 ± 0.1
		y	0.9 ± 0.9	1.0 ± 2.0
		Norm	1.14 ± 1.5	1.0 ± 2.0
x>0 , y<0	3155	x	1.0 ± 1.6	0.2 ± 0.1
		y	1.4 ± 1.8	0.4 ± 0.3
		Norm	1.7 ± 2.4	0.5 ± 0.3
x<0 , y>0	2550	x	0.9 ± 0.9	0.2 ± 0.2
		y	1.1 ± 0.8	0.7 ± 1.0
		Norm	1.4 ± 1.2	0.7 ± 1
x<0 , y<0	1483	x	1.3 ± 2.1	0.2 ± 0.2
		y	1.9 ± 4.1	0.4 ± 0.2
		Norm	2.3 ± 4.6	0.5 ± 0.3

through the workspace. We aim to investigate the accuracy of the US tracking algorithms in detecting the pose of the Millipede, since its silhouette in the US image changes with respect to the orientation of the soft robot, as shown in Fig. 7.

The CNN shows a lower error and variance in tracking the soft robot compared to the geometric (RANSAC) method. For the RANSAC-based method, the tracking errors for the x-coordinate, y-coordinate and angle are 0.9 ± 1.4 mm, 1.2 ± 2.1 mm and 3.9 ± 6.7°, respectively. On the same dataset, the proposed CNN obtains errors of 0.28 ± 0.14 mm, 0.62 ± 0.48 mm and 2.3 ± 3.4° for the x-coordinate, y-coordinate and angle, respectively.

Table I shows the position tracking error in different areas of the the workspace, using both methods. The error using the CNN-based method is better in every quadrant of the workspace when compared with the RANSAC-based tracker. Furthermore, for the CNN tracker, the error in the y-coordinate is 2-3 times bigger when the robot is away from the US probe ($y > 0$), compared to the other half of the workspace.

To compare the accuracy of tracking the orientation of the robot, the test dataset was divided into intervals of 20° according to the angle labelled in each ultrasound image. The absolute error between the ground-truth and the angle estimated by both methodologies was calculated. Fig. 8 shows the dispersion of the mean angle for each angle interval.

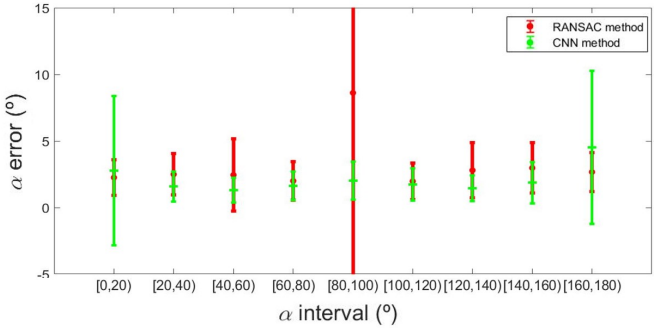

Fig. 8. Comparison of the absolute error between the angle predicted by the proposed tracking strategies and the ground-truth angle for the two US tracking methods. The mean error and the respective standard deviation are calculated for each 20° interval containing the ultrasound images, with the ground truth in the correspondent interval.

Overall, the dispersion of the error is lower in the CNN method, except for the intervals near the angle discontinuity (α = 0°, 180°). The RANSAC-based method presents an absolute angle error of 8.6 ± 21.0°, when 80°< α <100°, because only the small edge of the robot is imaged.

Overall, the estimation of the pose of the robot with the CNN method is better and more reliable, and therefore, the CNN method is used in the closed-loop experiments with ultrasound imaging.

2) Closed-loop control using ultrasound: The CNN architecture and the weights of the layers are stored and converted to be used in the C++ Cast API on the laptop. The pose estimation by the CNN runs at 15 Hz on the laptop computer. The CNN estimates the position and orientation of the Millipede for each frame acquired by the Cast API. Then, the information is sent to the control system implemented in BigMag and the required magnetic field is applied.

The closed-loop control is performed in two situations, with an obstacle on one side of the ultrasound image, as shown in Fig. 9. For the path avoiding the obstacle on the right, the absolute tracking error is 1.71 ± 0.85 mm, and the angle error is 2.29 ± 1.64 mm. For the other path, the absolute tracking error is 1.49 ± 0.83 mm and the angle error is 2.18 ± 1.33 mm. The mean deviation of the soft robot from the planned path is 1.87 ± 0.92 mm. In both cases, the robot is able to navigate the obstacles.

VII. CONCLUSIONS

In this paper, we achieve tracking, magnetic motion control and path planning of untethered biomimetic soft robot, using camera and ultrasound images. The motion planning and magnetic control are validated using camera feedback, and a deviation error of 1.71 ± 1.07 mm between the trajectory of the soft robot and the planned path is obtained. Also, the use of an A*-Wall Avoidance Path Planner to compute the obstacle free paths to control the robot is shown. The performance of the two proposed ultrasound tracking methods is compared using a dataset of images. The RANSAC and the CNN methods achieve an absolute tracking error of 1.5 ± 2.5 mm and 0.7 ± 0.5 mm, respectively, and an absolute angle error of 3.9 ± 6.7 ° and 2.3 ± 3.4 °, respectively. Despite the high signal-to-noise ratio of the ultrasound images, the soft robot autonomously reaches a target following a path at an average position tracking error of 1.59 ± 0.84 mm and an angle error of 2.24 ± 1.49 °, using the CNN method.

The current study takes a step forward in the localization and control of untethered soft robots using ultrasound feedback. Moreover, the CNN tracker achieves a lower error than the geometric method, since the CNN is more robust in dealing with occlusions of the silhouette of the soft robot. However, it is worth noting that in these experiments, the US probe images a stationary 2D workspace. Furthermore, the size of the workspace restricts the experiments to avoiding only one virtual obstacle on a horizontal surface. For applications such as the use of soft robots in medical procedures, the ultrasound imaging will have to consider

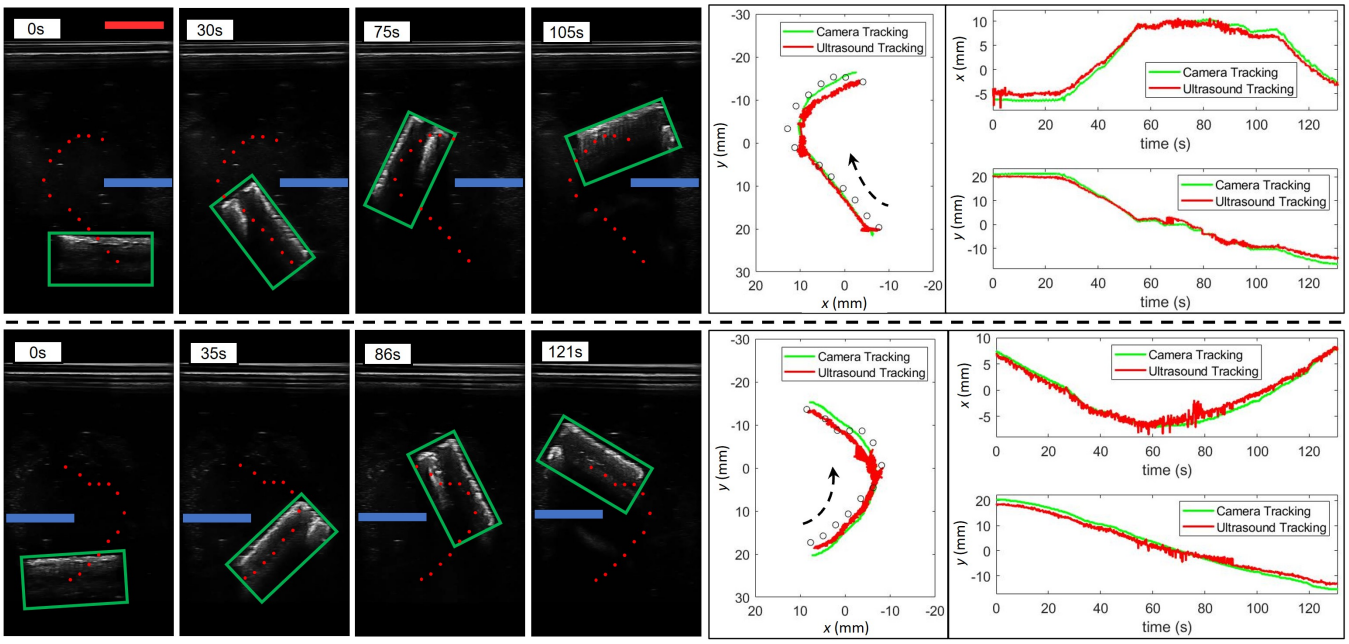

Fig. 9. Validation of the closed-loop control of a soft robot using a convolutional neural network (CNN) to detect the pose of the robot in ultrasound images. The experiment shows the steering of the Millipede robot to a target while avoiding an obstacle on the right (top column) and on the left side (bottom column). The scale bar is 20 mm. Plots on the right show the comparison of the trajectory obtained using the CNN tracker and the camera tracker (ground truth). *Please refer to the accompanying video.*

motion over greater distances and a dynamic 3D environment with moving obstacles and occlusions. Thus, for future work, we propose to test the robustness of the CNN tracker in a more application-relevant environment.

References

[1] M. Sitti, H. Ceylan, W. Hu, J. Giltinan, M. Turan, S. Yim, and E. Diller, "Biomedical applications of untethered mobile milli/microrobots," *Proceedings of the IEEE*, vol. 103, no. 2, pp. 205–224, 2015.

[2] V. K. Venkiteswaran, L. F. P. Samaniego, J. Sikorski, and S. Misra, "Bio-Inspired Terrestrial Motion of Magnetic Soft Millirobots," *IEEE Robotics and Automation Letters*, vol. 4, no. 2, pp. 1753–1759, Apr. 2019.

[3] J. Zhang, Z. Ren, W. Hu, R. H. Soon, I. C. Yasa, Z. Liu, and M. Sitti, "Voxelated three-dimensional miniature magnetic soft machines via multimaterial heterogeneous assembly," *Science Robotics*, vol. 6, no. 53, Apr. 2021.

[4] T. Xu, J. Zhang, M. Salehizadeh, O. Onaizah, and E. Diller, "Millimeter-scale flexible robots with programmable three-dimensional magnetization and motions," *Science Robotics*, vol. 4, no. 29, p. eaav4494, Apr. 2019.

[5] F. Ongaro, S. Scheggi, C. Yoon, F. v. den Brink, S. H. Oh, D. H. Gracias, and S. Misra, "Autonomous planning and control of soft untethered grippers in unstructured environments," *Journal of micro-bio robotics*, vol. 12, no. 1, pp. 45–52, 2017.

[6] J. Zhang, O. Onaizah, K. Middleton, L. You, and E. Diller, "Reliable grasping of three-dimensional untethered mobile magnetic microgripper for autonomous pick-and-place," *IEEE Robotics and Automation Letters*, vol. 2, no. 2, pp. 835–840, 2017.

[7] O. Erin, D. Antonelli, M. E. Tiryaki, and M. Sitti, "Towards 5-dof control of an untethered magnetic millirobot via mri gradient coils," in *Proceedings of the 2020 IEEE International Conference on Robotics and Automation (ICRA)*, 2020, pp. 6551–6557.

[8] T. Kensicher, J. Leclerc, D. Biediger, D. J. Shah, I. Seimenis, A. T. Becker, and N. V. Tsekos, "Towards mri-guided and actuated tetherless milli-robots: Preoperative planning and modeling of control," in *2017 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 2017, pp. 6440–6447.

[9] B. J. Nelson, I. K. Kaliakatsos, and J. J. Abbott, "Microrobots for minimally invasive medicine," *Annual review of biomedical engineering*, vol. 12, pp. 55–85, 2010.

[10] H. Ceylan, I. C. Yasa, U. Kilic, W. Hu, and M. Sitti, "Translational prospects of untethered medical microrobots," *Progress in Biomedical Engineering*, vol. 1, no. 1, p. 012002, 2019.

[11] S. Scheggi, K. K. T. Chandrasekar, C. Yoon, B. Sawaryn, G. van de Steeg, D. H. Gracias, and S. Misra, "Magnetic motion control and planning of untethered soft grippers using ultrasound image feedback," in *Proceedings of the 2017 IEEE International Conference on Robotics and Automation (ICRA)*, 2017, pp. 6156–6161.

[12] I. S. M. Khalil, P. Ferreira, R. Eleutério, C. L. de Korte, and S. Misra, "Magnetic-based closed-loop control of paramagnetic microparticles using ultrasound feedback," in *Proceedings of the 2014 IEEE International Conference on Robotics and Automation (ICRA)*, 2014, pp. 3807–3812.

[13] M. Turan, Y. Almalioglu, H. Araujo, E. Konukoglu, and M. Sitti, "Deep EndoVO: A recurrent convolutional neural network (RCNN) based visual odometry approach for endoscopic capsule robots," *Neurocomputing*, vol. 275, pp. 1861–1870, Jan. 2018.

[14] V. K. Venkiteswaran, D. K. Tan, and S. Misra, "Tandem actuation of legged locomotion and grasping manipulation in soft robots using magnetic fields," *Extreme Mechanics Letters*, vol. 41, p. 101023, 2020.

[15] J. Sikorski, I. Dawson, A. Denasi, E. E. Hekman, and S. Misra, "Introducing bigmag — a novel system for 3d magnetic actuation of flexible surgical manipulators," in *Proceedings of the 2017 IEEE International Conference on Robotics and Automation (ICRA)*, 2017, pp. 3594–3599.

[16] M. A. Fischler and R. C. Bolles, "Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography," *Communications of the Association for Computing Machinery (ACM)*, vol. 24, no. 6, pp. 381–395, 1981.

[17] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016, pp. 770–778.

[18] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," *arXiv preprint arXiv:1412.6980*, 2014.

[19] S.-h. Lim, S. W. Sohn, H. Lee, D. Choi, E. Jang, M. Kim, J. Lee, and S. Park, "Analysis and evaluation of path planning algorithms for autonomous driving of electromagnetically actuated microrobot," *International Journal of Control, Automation and Systems*, vol. 18, no. 11, pp. 2943–2954, 2020.