2023 IEEE International Conference on Robotics and Automation (ICRA 2023)
May 29 - June 2, 2023. London, UK

Image Segmentation for Continuum Robots from a Kinematic Prior

Connor M. Watson[1], Anna B. Nguyen[1], and Tania K. Morimoto[2]

***Abstract*— In this work, we address the problem of robust segmentation of a continuum robot from images without the need for training data or markers. We present a method that leverages information about the kinematics of these robots to produce an estimate of the robot shape, which is refined through optimization over global image statistics. Our approach can be straightforwardly applied to any continuum robot design and is able to handle partial occlusions of the robot body, as well as challenging background conditions. We validate our method experimentally for a concentric tube robot in a simulated surgical environment and show that our method significantly outperforms a naive projection of the robot shape and color thresholding, which is commonly used in current vision-based estimation algorithms for these robots. Overall, this work has the potential to improve the viability of vision-based state estimation for continuum robots in real-world settings.**

I. INTRODUCTION

Unlike traditional robotic manipulators, which change shape at discrete joint locations, continuum robots have shapes that vary continuously along their lengths [1]–[5]. Their flexibility, small form factor, and ability to navigate tortuous paths make continuum robots well suited for minimally invasive surgery (see [6] for a review) and other applications that require high dexterity in tight spaces.

These attractive, mechanical features, however, can also lead to challenges with sensing and estimation. For instance, because continuum robots are flexible, it is difficult to accurately estimate their shape, especially when the body of the robot is subjected to unknown forces. Phenomena like friction similarly impact the shape and dynamics of the robot in ways that are hard to predict. These issues are compounded by the small size of the robot body—whose diameter is typically on the millimeter scale— which imposes significant limitations on the integration of sensors.

Reliable state estimation is a critical challenge that must be solved before continuum robots can be safely deployed in environments like the surgical theatre. One approach to address these challenges, has been to use vision to estimate important aspects of the robot's state. Examples of this work include shape estimation [7]–[10], the identification of inertial parameters used in dynamics models [11], [12], and the estimation of forces the robot applies to unknown surroundings [13].

*This work was supported in part by the National Science Foundation under Grant 1935329.

[1]Connor Watson and Anna B. Nguyen are with the Department of Mechanical and Aerospace Engineering, University of California, San Diego, La Jolla, CA 92093 USA. cmwatson@eng.ucsd.edu

[2]Tania. K. Morimoto is with the Department of Mechanical and Aerospace Engineering and the Department of Surgery, University of California, San Diego, La Jolla, CA 92093 USA. tkmorimoto@eng.ucsd.edu

Fig. 1. An example of robust image segmentation for a concentric tube continuum robot. The robot is successfully identified, as illustrated by the white overlay, despite the presence of multiple, partial occlusions and a background of similar color.

Importantly, a common feature of these works on vision-based estimation is the requirement to have an accurate segmentation of the robot, that distinguishes the pixels belonging to the continuum robot body from the rest of the image. In order to obtain this segmentation reliably however, the images used in these previous works are taken under ideal conditions, which do not exist outside the laboratory, severely limiting their real-world applicability [14], [15]. To address this issue, we propose a new method for accurate continuum robot segmentation (Fig. 1) that is robust to more realistic imaging conditions and can be used to help further the growing body of research on vision-based estimation for continuum robots.

A. Continuum Robot Image Segmentation

In the last decade, most image segmentation research has been dominated by deep learning-based methods—largely spurred by the success of ImageNet [16]. It has become clear that, given sufficient computational power and training data, the results achieved by deep networks significantly outperform all other existing methods, especially on general image segmentation tasks.

However, amassing the requisite amount of labeled training data for the very specialized task of segmenting a continuum robot is a challenging undertaking. Even transfer learning approaches—where a successful deep network architecture that has first been trained on a very general set of labels is then retrained on a more specific set of labels using fewer images—require on the order of $\sim 10e4$ images. An

2023 IEEE International Conference on Robotics and Automation (ICRA) | 979-8-3503-2365-8/23/$31.00 ©2023 IEEE | DOI: 10.1109/ICRA48891.2023.10161229

individual collecting and labeling such a large amount of data is a highly impractical undertaking, especially given the large variety of continuum robot designs for which the process would need to be repeated. Online datasets, which have catalyzed advancements in other sub-disciplines of robotics, such as self-driving vehicles [17] and manipulation [18], do not yet exist for continuum robots. Finally, perhaps the most significant impediment to gathering useful training data for continuum robots is that many of the intended applications are surgical, which adds regulatory and privacy concerns for how the data is collected [19].

As an alternative approach to using deep networks for vision-based estimation for continuum robots, researchers often rely on composing an image scene to contain the entire robot body against a uniform, high-contrast background [7], [8], [10]–[13]. This control over the composition of the scene, enables straightforward, background-subtraction techniques based on color or another photometric quantity to effectively segment the robot without the need for large amounts of training data. These simple strategies are not robust, however, in that as the complexity of the scene increases (e.g. the robot is only partially visible, foreign objects are present, the background color is non-uniform and similar to the robot body, etc.) the segmentation becomes less reliable [14], [15]. Although it is possible to improve color-based segmentation methods through proper coloring of the instrument or the addition of reflective markers, such modifications suffer from sterilization and integration challenges [20] [21].

In the decade preceding the explosion of deep learning approaches, researchers studied segmentation methods based on models of the 3D shape of a target object. By leveraging this prior knowledge, these methods [22], [23] can effectively segment an object in a more complicated scene where a simple color-based method would fail. In this work, we propose to augment these methods with a kinematic model for a continuum robot that can be used for segmentation, without the need for a large set of training data.

B. Contributions

Specifically, the contribution of this work is an algorithm for segmentation of continuum robots that: requires minimal training data, is general across continuum robot designs, and segments the robot in realistic image conditions that include partial views and challenging background conditions (where color-based segmentation of the robot alone performs poorly). In the following section we detail the components of our algorithm that leverage information about the robot's shape derived from a kinematic model to construct an optimization problem which can be solved to segment the robot. Next, we test our algorithm empirically using images of the robot collected from a simulated surgical environment and demonstrate its improved performance relative to current techniques. Finally, we conclude with a discussion of the current limitations of the work, along with future directions.

TABLE I
NOMENCLATURE

$\hat{(\cdot)}$	conversion of an element of $\mathbb{R}^3$ to an element of $\mathfrak{se}(3)$
s	arclength variable of the continuum robot
L	total arclength of the robot
$\mathbf{u}$	curvature of the robot centerline
$\mathbf{p}$	position vector of the robot centerline
$\mathbf{R}$	orientation matrix of the robot centerline
$\mathbf{e}_3$	the third element of the standard basis
$\mathbf{q}$	vector of the robot joint positions
$f(\cdot)$	mapping from robot joint positions to centerline curvature
$\Gamma(\cdot)$	mapping from robot centerline pose to 3D shape
S	3D shape of the robot
$\mathbf{x}_0$	a 3D point on the surface of the robot in the robot frame
$\mathbf{\Omega}$	rotation between camera frame and robot frame
$\mathbf{t}$	translation between camera frame and robot frame
$\mathbf{x}$	a 3D point on the surface of the robot in the camera frame
$\pi(\cdot)$	a perspective projection mapping
$\mathbf{y}$	a 2D pixel coordinate in the domain of the image
I	raw image
M	binary mask of background
I'	background subtracted image
Φ	domain of image
B	subset of image domain belonging to background
R	subset of image domain belonging to projection of shape model
R^c	complement of R
$I(\mathbf{y})$	color value of the image of a pixel coordinate $\mathbf{y}$

II. Methods

Our algorithm for continuum robot segmentation consists of three main parts (Fig. 2). First, the image to be processed undergoes an initial background subtraction. Second, joint measurements of the robot's configuration are fed into a kinematic model to produce an initial parameterization of the shape of the robot. Finally, this parameterization is refined through optimization to minimize an energy defined for the background-subtracted image, which produces a segmentation for the robot. An example of this workflow is shown in Fig. 4. The nomenclature used throughout this article is summarized in Table I.

A. Background Subtraction

This first step is not expected to achieve a perfect segmentation—only to remove elements of the image that are grossly inconsistent with the expected appearance of the robot. Consequently, our algorithm *does not* require a specific background subtraction method. Instead, this choice is left to the user as a design decision, which may depend on the information available for the particular problem.

Consider an initial image I, a binary mask $M(I, \boldsymbol{\theta})$, and background-subtracted image $I' = I \circ M$, where $\circ$ is the Hadamard product. The mask, M, which delineates the foreground from the background, is a function of the image, I, and a vector of parameters, $\boldsymbol{\theta}$, used to encode any information about the background available to the user. For instance, if the user has access to images of the background *a priori*, $\boldsymbol{\theta}$ may be the weights of a Gaussian Mixture Model [24], neural network [25], or other learned function [26],

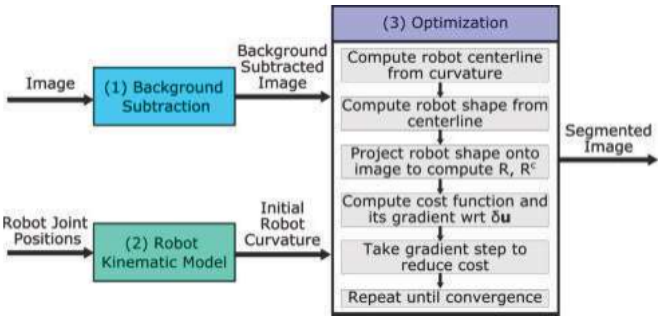


Fig. 2. Workflow diagram of the proposed algorithm.

meant to capture statistics about the background pixels. If, however, the user does not have data to train such a model, $\boldsymbol{\theta}$ may be the parameters used to specify a simple color mask that removes unlikely robot colors. In our experiments, we adopt this later approach, but incorporating more complicated models is straightforward.

B. Continuum Robot Shape Model

The key element of our proposed approach is to leverage a model of the expected shape of the continuum robot to improve segmentation. Specifically, we consider a mapping $\mathbf{u}(s) = f(\mathbf{q})$ from the robot's joint configuration, $\mathbf{q} \in \mathbb{R}^m$, to the *curvature* of its centerline, $\mathbf{u}(s)$, which is parameterized by the arclength of the centerline, $s \in [0, L]$. There is extensive literature on forward kinematic models specific for different continuum robot designs that define such mappings (for a review see [27] and [28]). These models range in complexity from simple constant curvature models, which do not vary with respect to s, to more complicated mechanics-based models that define $f(\cdot)$ as the solution to a set of differential equations.

Given such a mapping, it is then straightforward to determine the position, $\mathbf{p}(s)$, and orientation, $\mathbf{R}(s)$, of the robot centerline as a function of arclength by integrating the system of equations,

$$\begin{aligned} \dot{\mathbf{p}} &= \mathbf{R}\mathbf{e}_3 \\ \dot{\mathbf{R}} &= \mathbf{R}\hat{\mathbf{u}}, \end{aligned} \tag{1}$$

with an appropriate set of initial conditions, $\mathbf{p}(0), \mathbf{R}(0)$, where $\hat{(\cdot)}$ maps the robot curvature to $\mathfrak{so}(3)$. It is important to note that the 1D centerline of the robot is distinct from its expected *shape*, which is a surface embedded in 3D space. Therefore it is necessary to define another mapping, $S = \Gamma(\mathbf{p}, \mathbf{R}, s)$, which maps information about the robot centerline computed in Eq. 1 to the shape of the robot, S, effectively "drawing" the robot shape from its centerline. The exact definition of $\Gamma(\cdot)$ depends on the specific design of the robot to be segmented and is straightforward to determine given that continuum robot designs to date have cross sections that are radially symmetric about their centerline.

C. Segmentation by Shape Optimization

Consider an image I with domain Φ and area element $d\Phi$. Our approach to segmentation is inspired by methods such as [22] [29], which pose an optimization problem to minimize energies of the form

$$E = \int_R E_{in}(I)d\Phi + \int_{R^c} E_{out}(I)d\Phi, \tag{2}$$

where E_{in} and E_{out} are functions that are integrated to capture statistics about pixels within a region of the image, $R \subset \Phi$, and within its complement, $R^c = \Phi \backslash R$, respectively. The region R is determined from a model of the surface of the object to be segmented by first transforming the object into the reference frame of the camera and then projecting the object onto the image domain, Φ, using a projection model. Specifically, consider a model of an object with a shape defined by surface S. Every point on the surface, $\mathbf{x}_0 \in S$, is transformed into the camera frame by

$$\mathbf{x} = \boldsymbol{\Omega}\mathbf{x}_0 + \mathbf{t}, \tag{3}$$

where $\boldsymbol{\Omega}$ and $\mathbf{t}$ are the orientation and translation of the camera frame relative to the coordinate frame in which S is defined. This transformation is followed by a projection,

$$\mathbf{y} = \pi(\mathbf{x}), \tag{4}$$

which maps 3D points in the reference frame of the camera, $\mathbf{x} = [x, y, z]^T$, to 2D pixel locations in the image domain, $\mathbf{y}$. The region R is then defined to be the set of all points $\mathbf{y}$ in the image domain (Fig. 3). The choice of $\pi(\cdot)$ is another design decision afforded to the user. Here, we assume an ideal perspective projection model so that $\mathbf{y} = [\frac{x}{z}, \frac{y}{z}]^T$.

Prior works typically assume that the pose of the target object in the camera frame, $\{\boldsymbol{\Omega}, \mathbf{t}\}$, is unknown and solve for these variables by minimizing Eq. 2. In our case, however, the target object is a continuum robot for which the relative pose of the camera is usually known. In many robotic applications, the camera is either mounted in a fixed position relative to the robot [30] or is under direct robotic control—as is the case for many surgical robots [31]. Consequently, we instead choose to segment a continuum robot from an image by minimizing an energy functional like Eq. 2 over a parameterization of the robot shape.

As described in Section II-B, a model of the shape of a continuum robot is completely defined by its curvature, $\mathbf{u}$,

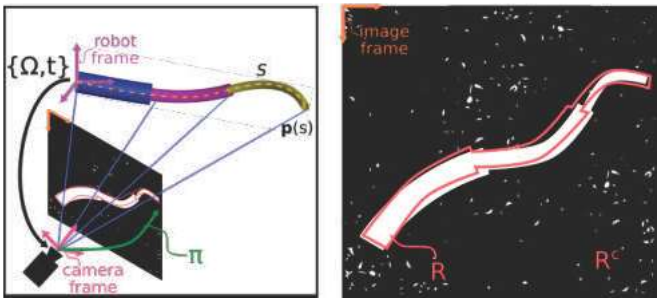


Fig. 3. A schematic of a continuum robot shape model and its projection onto an image. The centerline of the model, $\mathbf{p}(s)$, is determined from the robot's joint configuration and a forward kinematic model. Points on the surface of the robot, S, are transformed from the robot reference frame to the camera reference frame, given the relative pose of the camera as defined by $\{\boldsymbol{\Omega}, \mathbf{t}\}$, and are then projected into the image frame with a mapping defined by $\pi(\cdot)$.

which can be determined from the robot's joint configuration, $\mathbf{q}$. We pose a modified optimization problem to find a new robot curvature, $\mathbf{u}' = \mathbf{u} + \delta\mathbf{u}$, that solves the following:

$$\min_{\delta\mathbf{u}} \gamma_1 \int_{R-B} E_{in}(I'(\mathbf{y}))d\Phi + \gamma_2 \int_{R^c-B} E_{out}(I'(\mathbf{y}))d\Phi + \underbrace{\gamma_3 \int_L ||\delta\mathbf{u}||^2 ds}_{E_{reg}} + \underbrace{\gamma_4 \int_{R\cap B} d\Phi}_{E_{pen}}. \quad (5)$$

Here, E_{in} and E_{out} are the same as in Eq. 2, but we use a background subtracted version of the image (Section II-A), and the domains of integration are restricted to not include any pixels in the domain of the background, $B \subset \Phi$. Compared to Eq. 2, we also add two new terms to our modified objective function. The first term, E_{reg}, is a regularization term that penalizes large changes to the curvature of the robot and is integrated over the arclength of the robot centerline. The second additional term, E_{pen}, is the integral over all pixels that are common to both the background and the projection of the robot shape onto the image. This term penalizes solutions where large regions of the robot projection lie in the domain of the background. The weights, γ_{1-4}, are used to adjust the relative importance of each of the terms. We find that setting the weights such that all terms in the cost function have the same order of magnitude tends to produce effective results. There are many possible choices for the functions E_{in} and E_{out} [22]. Here we adopt the method from [29]:

$$E_{in,out}(I(\mathbf{y})) = (I(\mathbf{y}) - \boldsymbol{\mu}_{in,out})^2, \quad \boldsymbol{\mu}_{in,out} = \frac{\int_{R,R^c} I(\mathbf{y})d\Phi}{\int_{R,R^c} d\Phi}, \quad (6)$$

where $I(\mathbf{y}) \in \mathbb{R}^3$ is the color at pixel location $\mathbf{y}$, and $\boldsymbol{\mu}_{in,out}$ is the average color over the region R, R^c.

III. Experimental Evaluation

In this section, we empirically evaluate the performance of our algorithm under different conditions. We do so for the segmentation of a concentric tube robot (CTR), which is a specific class of continuum robot. We outline the details of the experiment, discuss metrics for characterizing the performance, and finally analyze the results.

A. Experimental Setup

CTRs consist of precurved, flexible, telescoping tubes which can be translated and rotated relative to one another by an actuation unit. As the tubes are actuated, they interact in bending and torsion to reach mechanical equilibrium and produce different robot configurations [4], [5]. The specific CTR used in this experiment has the same tubes and actuation unit as described in [32]. The kinematic model we use to determine the robot shape from its joint angles is based on the mechanics of Cosserat rods and is detailed in [33]. To draw the shape of the robot from its centerline, we define the function

$$S(s,\phi) = \Gamma(\mathbf{p},\mathbf{R},s,\phi) = \mathbf{p}(s) + \mathbf{R}(s)\mathbf{R}_z(\phi)\begin{pmatrix}\rho(s)\\0\\0\end{pmatrix} \quad (7)$$

where $\mathbf{R}(s)\mathbf{R}_z(\phi)$ is a rotation of ϕ radians about the tangent to the robot centerline and $\rho(s)$ is the diameter of the outermost robot tube at arclength location s.

To evaluate our proposed method, we gather 4 sets of 20 images of this robot from an external camera (Logitech C920) as the tip of the CTR follows a trajectory that spans a large region of its workspace (Fig. 5). At the time when each image is taken, the robot joint values are also recorded for use in the algorithm. Half of the image data collected (i.e. 40 images) comes from an "orthogonal" viewpoint—where the camera is oriented orthogonal to the insertion axis—while the other half comes from an "endoscopic" viewpoint—where the camera is oriented nearly parallel to the insertion axis of the robot. The first viewpoint is more common in lab settings while the second, more challenging viewpoint is similar to how a camera might be mounted for use with a minimally invasive surgical robot. The endoscopic viewpoint is more challenging since differences in the robot's configuration are difficult to discern and because the robot is often partially occluded by its own body and/or its actuation system.

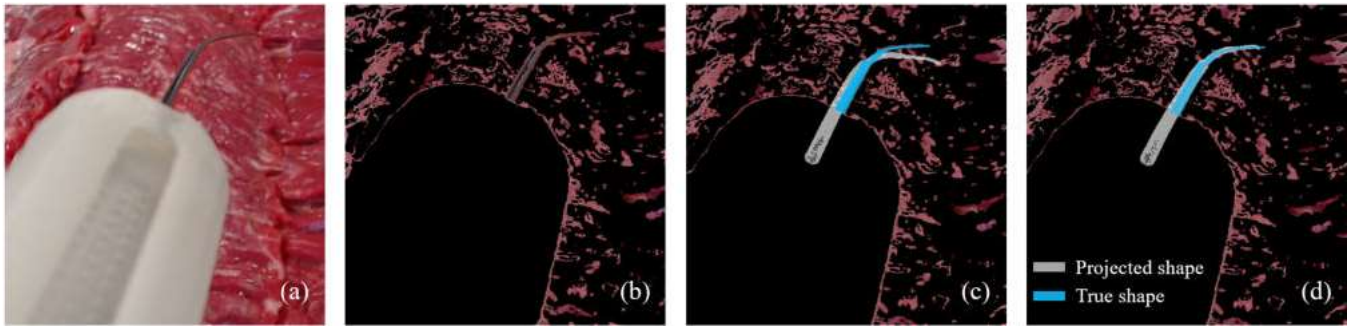


Fig. 4. (a) Example of a raw, color image prior to segmentation by our proposed algorithm. (b) The image after background subtraction has removed colors inconsistent with the robot body. (c) Projection of the initial robot shape (white) onto the background subtracted image clearly deviates from the true robot shape (blue). (d) Projection of the optimized robot shape (white) onto the background subtracted image. The optimized shape aligns much better with the true robot shape (blue).

TABLE II

COMPARISON OF IOUS FOR THE DIFFERENT SEGMENTATION METHODS ACROSS VARIOUS IMAGE SETS. ADJUSTED IOU SCORES ARE REPORTED TO THE RIGHT OF THE | (SEE SECTION III-B FOR DETAILS)

	Orthogonal	Orthogonal Occluded	Endoscopic	Endoscopic Occluded
Background Subtraction	0.016 $\pm$ 0.001	0.015 $\pm$ 0.001	0.022 $\pm$ 0.003	0.013 $\pm$ 0.003
Unoptimized Projection	0.284 $\pm$ 0.087	0.292 $\pm$ 0.101	0.284 $\pm$ 0.092 | 0.415 $\pm$ 0.143	0.254 $\pm$ 0.088 | 0.427 $\pm$ 0.157
Optimized Projection	**0.665 $\perp$ 0.063**	**0.603 $\perp$ 0.134**	0.414 $\perp$ 0.104 | **0.636 $\perp$ 0.153**	0.3 $\perp$ 0.109 | **0.52 $\perp$ 0.171**

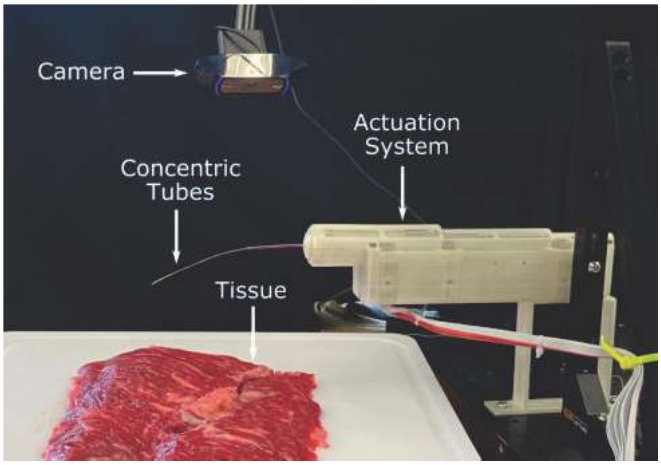


Fig. 5. The experimental setup consists of a concentric tube robot mounted above a piece of tissue. An external camera is attached to an articulating locking arm to enable images to be taken from both an "orthogonal" and an "endoscopic" viewpoint.

From each viewpoint, we collect 2 sets of data, one in which the robot is partially occluded by a laparoscopic instrument (Endopath monopolar cautery scissors, Ethicon) and one in which it is not. The camera pose relative to the robot is determined from markers affixed to the robot at known locations. All images are taken against a background of animal tissue which is similar in color to the robot tubes and meant to simulate a surgical environment (Fig. 5). After collection, the images are hand-segmented to produce ground truth labels that can be used to evaluate the performance of the algorithm. We use the Intersection Over Union (IoU) of the algorithm segmentation against the ground truth as the performance metric, which is computed as:

$$IoU = \frac{|A \cap B|}{|A \cup B|}. \tag{8}$$

Higher IoUs indicate better performance and an IoU of 1 indicates a perfect match between the algorithm segmentation and ground truth. We also report IoUs for the segmentation produced by the unoptimized projection of the robot (i.e. $\delta \mathbf{u} = 0$) and for the background subtraction for comparison with our algorithm. The color mask used for background subtraction (Section II-A) and the weights in the objective function used for optimization are the same for all images.

B. Results

We record the mean IoU plus or minus one standard deviation for each method on each data set in Table II. From these numbers, it is clear that background subtraction by color thresholding alone is a poor way to segment the robot when it is similar in appearance to the background (top row of Table III), regardless of the viewing angle, which is consistent with previous works [14], [15]. This result motivates the need for a more advanced approach for segmentation, even when the entire robot body is visible in the image.

Projecting the unoptimized shape of the robot gives better results than background subtraction alone but still fails to accurately capture all of the pixels of the robot (middle row of Table III). Although we use a sophisticated mechanical model to predict the shape of the robot, there are many sources of error that contribute to the misalignment between the projection of the robot shape model and the actual image of the robot. Examples of possible sources of error include, but are not limited to, the effects of unmodeled phenomena like friction, backlash, and fabrication tolerances [34]; uncertainty in model parameters such as bending stiffness or tube curvatures; and small calibration errors in camera pose or joint offsets.

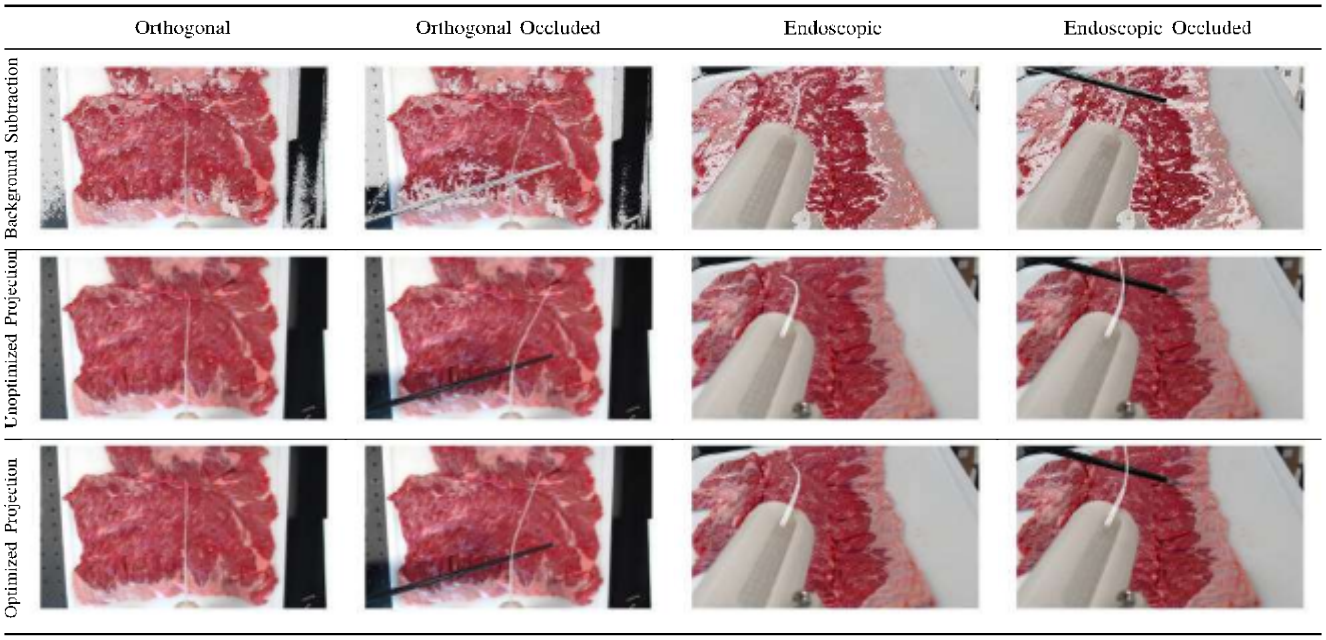
Our proposed algorithm, however, results in significantly higher IoUs across all test conditions. This improved result is due to the fact that our method uses data from the image to refine the model of the robot's shape, consequently producing a projection that more closely agrees with the ground truth labels (bottom row of Table III). Note that for the endoscopic viewpoint datasets, we report two different IoUs: a raw IoU, calculated as previously described, and an adjusted IoU, which excludes segmentation pixels that are in occluded regions. The rationale for this adjusted IoU is that the ground truth labels only capture regions of the image where the robot is visible. The projection of the robot shape model, however—whether optimized or not—does not account for any occlusions in the environment. This feature is beneficial in that it provides the user with a segmentation of the entire robot—even portions that are not visible—but makes the IoUs of the optimized and unoptimized projections more difficult to distinguish numerically because a large percentage of both projections is occluded by the actuation unit of the robot. By removing the pixels projected into the occluded regions from the IoU calculation, the improvement due to optimization of the robot shape model becomes more apparent.

IV. CONCLUSION

In conclusion, we have presented a new algorithm for robust continuum robot segmentation that does not require labeled training data. Our approach poses an optimization over changes to a parameterization of the shape of the robot. It produces a segmentation of the robot even for challenging cases, such as when the background is similar in color to

TABLE III

DIFFERENT IMAGE SEGMENTATION METHODS APPLIED TO EACH OF THE FOUR DATASETS DESCRIBED IN SECTION III-A. THE RESULTING SEGMENTATION IS SHOWN AS A TRANSPARENT, WHITE OVERLAY ON TOP OF EACH OF THE ORIGINAL IMAGES.



the robot or the robot is only partially visible. We validated our proposed approach empirically and demonstrated that our method shows significant, quantifiable improvement over color-based segmentation approaches, and is robust to errors in the model used to describe the shape of the robot and its projection.

There are several possible extensions of this method and directions for future work. First, our current method assumes that the robot operates in free space, unaffected by external forces arising from, for example, tissue interaction that may be present in a real-world scenario. To account for this, our approach can be adapted to utilize more sophisticated kinematic models that account for the effect of external loads on the robot shape [13]. Second, the computational efficiency of the algorithm can be improved by investigating lower-dimensional parameterizations of the robot shape, warm-starting the optimization, and exploiting parallel computations that could enable use in real-time applications. Finally, this method could be combined with learning-based segmentation techniques for improved performance, similar to the automated labeling method described in [35].

REFERENCES

[1] S. Hirose and M. Mori, "Biologically inspired snake-like robots," in *2004 IEEE International Conference on Robotics and Biomimetics*. IEEE, 2004, pp. 1–7.

[2] W. McMahan, V. Chitrakaran, M. Csencsits, D. Dawson, I. D. Walker, B. A. Jones, M. Pritts, D. Dienno, M. Grissom, and C. D. Rahn, "Field trials and testing of the octarm continuum manipulator," in *Proceedings 2006 IEEE International Conference on Robotics and Automation, 2006. ICRA 2006.* IEEE, 2006, pp. 2336–2341.

[3] E. W. Hawkes, L. H. Blumenschein, J. D. Greer, and A. M. Okamura, "A soft robot that navigates its environment through growth," *Science Robotics*, vol. 2, no. 8, p. eaan3028, 2017.

[4] R. J. Webster, A. M. Okamura, and N. J. Cowan, "Toward active cannulas: Miniature snake-like surgical robots," in *2006 IEEE/RSJ international conference on intelligent robots and systems*. IEEE, 2006, pp. 2857–2863.

[5] P. Sears and P. Dupont, "A steerable needle technology using curved concentric tubes," in *2006 IEEE/RSJ international conference on intelligent robots and systems*. IEEE, 2006, pp. 2850–2856.

[6] P. E. Dupont, N. Simaan, H. Choset, and C. Rucker, "Continuum robots for medical interventions," *Proceedings of the IEEE*, vol. 110, no. 7, pp. 847–870, 2022.

[7] J. M. Croom and J. M. Romano, "Visual sensing of continuum robot shape using self-organizing maps," *IEEE Robotics and Automation Letters*, vol. 2, no. 3, pp. 4591–4596, May 2010.

[8] R. A. Manakov, D. Y. Kolpashchikov, V. V. Danilov, N. V. Laptev, I. Skirnevskiy, and O. M. Gerget, "Visual shape and position sensing algorithm for a continuum robot," in *IOP Conference Series: Materials Science and Engineering*, vol. 1019, no. 1. IOP Publishing, 2021, p. 012066.

[9] B. Weber, P. Zeller, and K. Kühnlenz, "Multi-camera based real-time configuration estimation of continuum robots," in *2012 IEEE/RSJ International Conference on Intelligent Robots and Systems*. IEEE, 2012, pp. 3350–3355.

[10] M. M. Dalvand, S. Nahavandi, and R. D. Howe, "Fast vision-based catheter 3d reconstruction," *Physics in Medicine & Biology*, vol. 61, no. 14, p. 5128, 2016.

[11] J. Till, V. Aloi, and C. Rucker, "Real-time dynamics of soft and continuum robots based on cosserat rod models," *The International Journal of Robotics Research*, vol. 38, no. 6, pp. 723–746, 2019.

[12] J. Till, V. Aloi, K. E. Riojas, P. L. Anderson, R. J. Webster III, and C. Rucker, "A dynamic model for concentric tube robots," *IEEE Transactions on Robotics*, vol. 36, no. 6, pp. 1704–1718, 2020.

[13] V. A. Aloi and D. C. Rucker, "Estimating loads along elastic rods," in *2019 International Conference on Robotics and Automation (ICRA)*. IEEE, 2019, pp. 2867–2873.

[14] S. Dambreville, Y. Rathi, and A. Tannen, "Shape-based approach to robust image segmentation using kernel pca," in *2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'06)*, vol. 1, 2006, pp. 977–984.

[15] X. Fu, C.-Y. Wang, C. Chen, C. Wang, and C.-C. J. Kuo, "Robust image segmentation using contour-guided color palettes," in *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*, December 2015.

[16] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification

with deep convolutional neural networks," *Communications of the ACM*, vol. 60, no. 6, pp. 84–90, 2017.

[17] Y. Kang, H. Yin, and C. Berger, "Test your self-driving algorithm: An overview of publicly available driving datasets and virtual testing environments," *IEEE Transactions on Intelligent Vehicles*, vol. 4, no. 2, pp. 171–185, 2019.

[18] J. Mahler, F. T. Pokorny, B. Hou, M. Roderick, M. Laskey, M. Aubry, K. Kohlhoff, T. Kröger, J. Kuffner, and K. Goldberg, "Dex-net 1.0: A cloud-based network of 3d objects for robust grasp planning using a multi-armed bandit model with correlated rewards," in *2016 IEEE international conference on robotics and automation (ICRA)*. IEEE, 2016, pp. 1957–1964.

[19] M. Bhandari, T. Zeffiro, and M. Reddiboina, "Artificial intelligence and robotic surgery: current perspective and future directions," *Current opinion in urology*, vol. 30, no. 1, pp. 48–54, 2020.

[20] D. Stoyanov, "Surgical vision," *Annals of biomedical engineering*, vol. 40, pp. 332–345, 2012.

[21] D. Bouget, M. Allan, D. Stoyanov, and P. Jannin, "Vision-based and marker-less surgical tool detection and tracking: a review of the literature," *Medical image analysis*, vol. 35, pp. 633–654, 2017.

[22] S. Dambreville, R. Sandhu, A. Yezzi, and A. Tannenbaum, "A geometric approach to joint 2d region-based segmentation and 3d pose estimation using a 3d shape prior," *SIAM journal on imaging sciences*, vol. 3, no. 1, pp. 110–132, 2010.

[23] H. Chang, Q. Yang, and B. Parvin, "A bayesian approach for image segmentation with shape priors," in *2008 IEEE Conference on Computer Vision and Pattern Recognition*. IEEE, 2008, pp. 1–8.

[24] Z. Zivkovic, "Improved adaptive gaussian mixture model for background subtraction," in *Proceedings of the 17th International Conference on Pattern Recognition, 2004. ICPR 2004.*, vol. 2. IEEE, 2004, pp. 28–31.

[25] M. Braham and M. Van Droogenbroeck, "Deep background subtraction with scene-specific convolutional neural networks," in *2016 international conference on systems, signals and image processing (IWSSIP)*. IEEE, 2016, pp. 1–4.

[26] N. M. Oliver, B. Rosario, and A. P. Pentland, "A bayesian computer vision system for modeling human interactions," *IEEE transactions on pattern analysis and machine intelligence*, vol. 22, no. 8, pp. 831–843, 2000.

[27] R. J. Webster III and B. A. Jones, "Design and kinematic modeling of constant curvature continuum robots: A review," *The International Journal of Robotics Research*, vol. 29, no. 13, pp. 1661–1683, 2010.

[28] H. B. Gilbert, "On the mathematical modeling of slender biomedical continuum robots," *Frontiers in Robotics and AI*, vol. 8, p. 732643, 2021.

[29] T. F. Chan and L. A. Vese, "Active contours without edges," *IEEE Transactions on image processing*, vol. 10, no. 2, pp. 266–277, 2001.

[30] M. Remacle, V. MN Prasad, G. Lawson, L. Plisson, V. Bachy, and S. Van der Vorst, "Transoral robotic surgery (tors) with the medrobotics flex™ system: first surgical application on humans," *European Archives of Oto-Rhino-Laryngology*, vol. 272, no. 6, pp. 1451–1455, 2015.

[31] A. R. Lanfranco, A. E. Castellanos, J. P. Desai, and W. C. Meyers, "Robotic surgery: a current perspective," *Annals of surgery*, vol. 239, no. 1, p. 14, 2004.

[32] C. Girerd and T. K. Morimoto, "Design and control of a hand-held concentric tube robot for minimally invasive surgery," *IEEE Transactions on Robotics*, vol. 37, no. 4, pp. 1022–1038, 2021.

[33] D. C. Rucker, B. A. Jones, and R. J. Webster III, "A geometrically exact model for externally loaded concentric-tube continuum robots," *IEEE Trans. Robot.*, vol. 26, no. 5, pp. 769–780, 2010.

[34] J. Ha, G. Fagogenis, and P. E. Dupont, "Modeling tube clearance and bounding the effect of friction in concentric tube robot kinematics," *IEEE Trans. Robot.*, vol. 35, no. 2, pp. 353–370, 2018.

[35] C. da Costa Rocha, N. Padoy, and B. Rosa, "Self-supervised surgical tool segmentation using kinematic information," in *2019 International Conference on Robotics and Automation (ICRA)*. IEEE, 2019, pp. 8720–8726.