Continuous Shape Estimation of Continuum Robots Using X-ray Images

Edgar J. Lobaton[1], Jinghua Fu[2], Luis G. Torres[2], and Ron Alterovitz[2]

***Abstract*— We present a new method for continuously and accurately estimating the shape of a continuum robot during a medical procedure using a small number of X-ray projection images (e.g., radiographs or fluoroscopy images). Continuum robots have curvilinear structure, enabling them to maneuver through constrained spaces by bending around obstacles. Accurately estimating the robot's shape continuously over time is crucial for the success of procedures that require avoidance of anatomical obstacles and sensitive tissues. Online shape estimation of a continuum robot is complicated by uncertainty in its kinematic model, movement of the robot during the procedure, noise in X-ray images, and the clinical need to minimize the number of X-ray images acquired. Our new method integrates kinematics models of the robot with data extracted from an optimally selected set of X-ray projection images. Our method represents the shape of the continuum robot over time as a deformable surface which can be described as a linear combination of time and space basis functions. We take advantage of probabilistic priors and numeric optimization to select optimal camera configurations, thus minimizing the expected shape estimation error. We evaluate our method using simulated concentric tube robot procedures and demonstrate that obtaining between 3 and 10 images from viewpoints selected by our method enables online shape estimation with errors significantly lower than using the kinematic model alone or using randomly spaced viewpoints.**

I. Introduction

Continuum robots have a continuously bending, curvilinear structure and have the potential to enable new medical procedures by maneuvering through constrained anatomical spaces in a snake-like manner. Examples of continuum robots with the ability to reach difficult-to-access sites in the human body include bevel-tip steerable needles [1], concentric tube robots [2], [3], superelastic backbone robots [4], and highly articulated robotic manipulators [5]. To fully harness the potential of these devices, physicians must know the entire curvilinear shape of the continuum robot inside the body. This shape information is required to guide the robot's end effector to the clinical site while ensuring that the robot's curvilinear shape avoids anatomical obstacles and sensitive tissues such as bones, arteries, and nerves.

Accurately and continuously estimating the shape of a continuum robot over the duration of a procedure is challenging using currently available tools. Kinematics models of continuum robots are imprecise. The curvilinear shape of a continuum robot inside the human body is difficult to predict due to the robot's compliance, noisy actuation resulting from miniaturization, and the uncertainties resulting from device/tissue interaction. The shape of the robot could be precisely reconstructed at discrete time points from a CT scan (or from multiple X-ray images), but these imaging modalities should ideally not be used continuously since they rely on ionizing radiation; an estimated 2% of cancers are attributable to excessive use of radiation-based medical imaging [6]. Other available imaging modalities are either prohibitively expensive for routine procedures (e.g. MRI) or do not offer sufficient resolution to accurately track a continuum robot (e.g. 3D ultrasound).

In this paper, we propose a new method to continuously and accurately estimate the shape of a continuum robot over the duration of a procedure using only a small number of X-ray projection images (e.g., radiographs or fluoroscopy images). Our approach begins by estimating the continuum

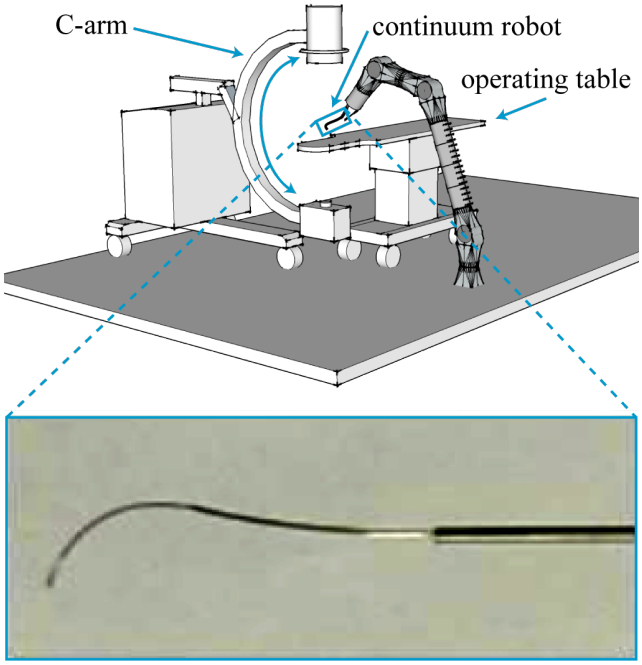


Fig. 1. Our objective is to accurately estimate the shape of a continuum robot over time during a procedure using a small number of optimally selected 2D X-ray projection images. We assume the X-ray sensor is mounted on a C-arm (top), a commonly used medical device that rotates the X-ray sensor about the patient lying on an operating table. Our method computes optimal viewpoints for the X-ray sensor to maximize the quality of the online shape estimation of the continuum robot. We apply our method to concentric tube robots, a type of continuum robot (bottom).

*This work was supported in part by the National Science Foundation (NSF) through awards #IIS-0905344 and #IIS-1149965 and by the National Institutes of Health (NIH) under award #R21EB011628.

[1]E. J. Lobaton is with the Department of Electrical and Computer Engineering, North Carolina State University, USA, edgar.lobaton@ncsu.edu

[2]J. Fu, L. G. Torres, and R. Alterovitz are with the Department of Computer Science, University of North Carolina at Chapel Hill, USA, ron@cs.unc.edu

robot's shape using a kinematics model, and then refines the estimate using data extracted from X-ray projection images. Although the X-ray images are acquired one at a time at discrete time points, our method is capable of accurately estimating the robot's shape continuously over the duration of the procedure by integrating the kinematics model with data from previously acquired images.

Our method is directly applicable to procedures involving continuum robots in which an X-ray imager is mounted on a C-arm, which is capable of rotating the imager in a circle around the patient, as shown in Fig. 1. This type of imaging is often used in interventional radiology procedures. We assume the angle of the C-arm can be measured and controlled. We note that our approach can be applied to continuum robots with or without embedded fiducial markers. To further improve shape estimation quality, we also optimize the viewpoint of the X-ray imager for each image in order to maximize the improvement in the quality of the shape estimation. This approach could provide physicians with a low-cost, high quality estimate of the robot's shape with minimal radiation exposure for the patient.

To enable continuous shape estimation, we represent the shape of the continuum robot over time as a deformable surface which can be described as a linear combination of time and space bases. With the bases specified or learned a priori, our online problem becomes the optimization of the coefficients which correspond to the bases. We optimize the coefficients by using information from a sequentially optimal set of viewpoints. Our method takes advantage of Bayes' rule and numeric optimization to estimate the uncertainty of the continuum robot shape reconstruction and minimize reconstruction errors by minimizing the uncertainty. We apply our method in simulation to concentric tube robots, a class of thin, dexterous continuum robots. We can accurately, continuously estimate robot shape in simulation using only 3 to 10 X-ray images over the duration of a procedure.

II. Related Work

Our approach for online shape estimation of continuum robots integrates information from an uncertain kinematics model with computer vision methods for 3D reconstruction. In vision-based 3D reconstruction, a 3D shape model of an object is computed from projection images taken from multiple viewpoints, where the images can be optical or X-ray projection images. Such 3D reconstruction methods typically require images from at least two viewpoints as well as identification of corresponding points that are visible in each image [7].

Methods using epipolar geometry have been introduced to compute 3D reconstructions of curvilinear, tubular objects from optical or X-ray projection images. To address the challenge of point correspondences, methods have used brute-force search [8], fiducial markers embedded in the tubes [9], self-organizing maps [10], and minimization of reprojection error subject to device-specific plausibility constraints [11]. Another approach used learning algorithms in conjunction with training data to map visual features identified in the images to a shape estimate of the robot [12]. The above methods focus on creating a single 3D reconstruction, and new images need to be acquired to re-estimate the robot shape every time the robot moves. Consequently, for medical procedures relying on X-ray images these 3D reconstruction methods will require excessive radiation exposure to the patient over the course of a dynamic procedure.

An alternative to optical and X-ray images is 3D ultrasound imaging. Although the resolution of 3D ultrasound is low, recent algorithmic advancements based on geodesic active contours are enabling shape estimation of portions of continuum robots in 3D ultrasound [13], although errors are still substantially greater than when using projection images for 3D reconstruction. Another option is integrating into the continuum robot a Fiber Bragg Grating (FBG) sensor, which uses an optical technique to estimate the shape of a flexible device. To overcome the large errors that arise in practice when using FBG sensors, a filtering method was developed for articulated snake robots [14], but the method is not easily transferrable to other continuum robots. Another sensing modality is to use a magnetic tracker at the tip of the continuum robot [15], although this approach does not scale well to reconstructing the full shape of a continuum robot.

When computing a 3D reconstruction of an object using projection images, the placement, and hence the view, of the camera or X-ray sensor for each image has a large impact on shape reconstruction accuracy. Methods have been developed to optimize sensor placement and camera parameter settings in order to minimize ambiguity in object recognition [16], minimize uncertainty in object tracking [17], [18], minimize pose estimation uncertainty of mobile robots [19], and estimate an object's configuration [20]. We focus on maximizing the quality of the 3D reconstruction of a continuum robot.

Recent research has begun investigating 3D reconstructions of moving objects. Paladini et al. and Xiao et al. [21], [22] used linear shape models to represent nonrigid 3D structures. This representation has been successfully applied to reconstructing facial and body motion, but it requires an orthographic camera and cannot handle missing information. Park et al. [23] reconstructed the 3D trajectory of a moving point by describing the motion as a linear combination of trajectory bases. With the bases known a priori, the problem is transformed into obtaining the coefficients of the bases, which significantly reduces the complexity and uncertainty. This method can also handle missing data. In our method, we represent the motion of the continuum robot in a similar manner, as a linear combination of bases.

III. Problem Statement

Our objective is to accurately and continuously estimate the shape of a continuum robot during a procedure by integrating information from kinematics modeling and a small finite number of X-ray images. We assume that we are given a kinematic model of the robot, but due to modeling errors and unexpected forces from the environment, the actual robot shape will differ from the model's prediction. We also assume that we have an X-ray imaging sensor that can

acquire perspective projection images from viewpoints that we select, as shown in Fig. 1. To limit radiation exposure to the patient to clinically acceptable doses, we cap the number of images taken to 10, with fewer images being better.

We define the shape of the continuum robot as a curve over time:

$$\gamma_t(s) = [\gamma_{t,1}(s)\ ,\ \gamma_{t,2}(s)\ ,\ \gamma_{t,3}(s)]^\top \tag{1}$$

where $\gamma_t : [0, S(t)] \to \mathbb{R}^3$, $s \in [0, S(t)]$ parameterizes the curve spatially at a given time $t \in [0, T]$, $S(t)$ is the length of the curve, T is the duration of the procedure, and $\gamma_{t,j} : [0, S(t)] \to \mathbb{R}$ for $j \in \{1, 2, 3\}$ are the corresponding x, y, z components of the curve.

Due to uncertainty, the actual shape of the robot may not match the kinematic model. We represent the real-world shape as a perturbation from the kinematic model, namely,

$$\gamma_t(s) = \gamma_t^K(s) + \gamma_t^P(s), \tag{2}$$

where γ_t^K is the kinematic model and γ_t^P is the perturbation of the actual shape relative to the kinematic model.

We assume that the imaging sensor can make observations (i.e., acquire images) of the continuum robot from various configurations. The configuration of the imaging sensor at time t is defined by the variable $C_t := (R_t, d_t)$, where $R_t = [R_{t,1}\ ,\ R_{t,2}\ ,\ R_{t,3}]^\top$ and $d_t = [d_{t,1}\ ,\ d_{t,2}\ ,\ d_{t,3}]^\top$ are the rotation and translation that define the position of the sensor. For a point on the curvilinear device, its 3D coordinates X and 2D projection $h_t(X)$ satisfy

$$h_t(X) = f \cdot \left[\frac{R_{t,1}^\top X + d_{t,1}}{R_{t,3}^\top X + d_{t,3}}\ ,\ \frac{R_{t,2}^\top X + d_{t,2}}{R_{t,3}^\top X + d_{t,3}} \right]^\top, \tag{3}$$

where f is the focal length for the perspective projection. Over the duration of the procedure, we assume that N_I images are taken: $\{C_{t_n} \mid t_n = (n - 0.5)T/N_I, n = 1, \cdots, N_I\}$.

X-ray images are inherently noisy. To quantify the uncertainty in our observations, we define our observation at time t_n, $\rho_{t_n} : [0, 1] \to \mathbb{R}^2$, as:

$$\rho_{t_n}(s) = h_{t_n}(\gamma_{t_n}(s)) + \eta_{t_n}(s), \tag{4}$$

where $\eta_{t_n}(s)$ models the noise in the image as a 2-dimensional vector composed of i.i.d. random variables with distribution $N(0, \sigma_o^2)$.

Our objective is to estimate the shape γ_t of the deforming continuum robot. To maximize the quality of this estimate, our variables are the sensor configurations, C_{t_n}, i.e. the locations from which we will acquire X-ray images. We define the following objective function at any given time t:

$$F_t(\tilde{\gamma}_t) = \mathbf{E}\left[||\gamma_t - \tilde{\gamma}_t||^2 \mid (\rho_{t_n}, C_{t_n}) \text{ for } 0 < t_n \leq t \right], \tag{5}$$

where $\tilde{\gamma}_t$ is our estimate of the robot's shape at time t. This objective function quantifies the expected error at time t given the current and prior observations. The output of our method is a sequence of sensor viewpoints C_{t_n} and the resulting shape estimate $\tilde{\gamma}_t$ for the robot as a function of time.

IV. Method

We represent the trajectory of a curvilinear device over time as a surface γ in 4D (space and time), as shown in Fig. 2. Under the assumption that perturbations to this surface can be captured by a linear combination of known basis functions (see Sec. IV-A), then estimating the shape of a perturbed surface is equivalent to finding the coefficients of the corresponding bases (see Sec. IV-B). Furthermore, reducing the uncertainty on these coefficients by selecting appropriate sensor configurations translates into higher precision in shape estimation (see Sec. IV-C).

First, we focus on the case where markers are present in the continuum robot. We assume $N_m(t)$ markers are present at discrete values $\{s_k\}_{k=1}^{N_m(t)}$ in Eqn. 1. Then, we turn to the markerless case by introducing a simple method to automatically find correspondences between points on a 3D model and its 2D projections (see Sec. IV-D).

A. Surface Representation

Park et al. [23] present a linear representation for the 3D trajectory of a moving point. Based on this representation, they provide a linear solution to reconstruct the 3D trajectory of a moving point from a series of 2D projections.

Inspired by their work, we represent γ_t^P as:

$$\gamma_t^P(s) = \sum_i^{N_b} b_i \psi_i(s, t), \tag{6}$$

where $\{\psi_i\}$ are the space and time basis functions, N_b is the number of bases, and b is the vector of coefficients for each basis function. Each b_i is assumed to come from a normal distribution $N(0, \sigma_{b,i}^2)$.

The first step of our method is to identify appropriate basis functions $\{\psi_i\}$ and corresponding variance $\left\{\sigma_{b,i}^2\right\}$. One approach is to use a standard set of basis functions, such as two-dimensional sine and cosine basis functions, and tune the variance parameters. Alternatively, the basis functions can be learned through the use of functional data analysis [24]. In particular, if we are given several realizations of the shape of a curvilinear robot over time, we can estimate the basis functions by computing the covariance of the 3D coordinates of sample points on the curve. The eigenvectors of this matrix provide an estimate to the basis functions, and the eigenvalues are estimates to $\sigma_{b,i}^2$. Hereafter, we assume that the bases are known, which simplifies the complexity

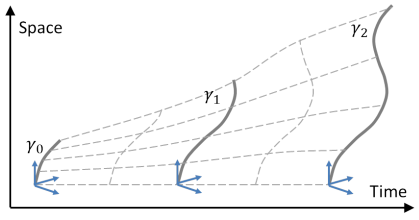


Fig. 2. Shape of the curvilinear device over time can be represented as a surface. The horizontal axis denotes time and the vertical axis denotes shape of the device in space.

of the shape recovery problem to the estimation of the coefficients b.

Before discussing our algorithm in detail, we summarize the inputs and outputs at a given time $t \in [0, T]$:

Input

- Duration of task: T
- Number of images: N_I
- Times of image acquisitions: $t_n = (n - 0.5)T/N_I$ for $n = 1, \ldots, N_I$
- Number of markers visible at time t: $N_m(t)$
- Kinematic model: γ_t^K for $t \in [0, T]$
- Prior sensor configurations C_{t_n} for $0 < t_n \leq t$
- Prior observations: ρ_{t_n} for $0 < t_n \leq t$
- Variance of observation noise: σ_o^2
- Number of perturbation bases: N_b
- Perturbation basis: $\{\psi_i\}_1^{N_b}$
- Variance of perturbation bases: $\{\sigma_{b,i}^2\}_{i=1}^{N_b}$

Output

- Estimated shape: $\hat{\gamma}_t$
- Next sensor configuration: $C_{t_{n'}}$, $n' = \min\{n \mid t_n > t\}$

B. Coefficient Estimation

Based on Eqns. 2 and 6, $||\gamma_t - \tilde{\gamma}_t||^2 = (b - \tilde{b})^\top M_t (b - \tilde{b})$ where $M_{t,ij} := \int_0^1 \psi_i(s,t)^\top \psi_j(s,t)\, ds$. Hence, we can rewrite Eqn. 5 as:

$$F_t(\tilde{b}) = \int (b - \tilde{b})^\top M_t (b - \tilde{b})\; f_t(b \mid \rho_{t_n}, C_{t_n})\; db, \tag{7}$$

where $f_t(b \mid \rho_{t_n}, C_{t_n})$ is the conditional probability distribution of having b as the correct coefficients given all the observations and sensor configurations up to time t.

In this section, we find an estimator that minimizes the objective function above. That is, we are after

$$\hat{b}_t = \arg\min_{\tilde{b}} F_t(\tilde{b}). \tag{8}$$

Our derivation resembles in spirit to that of the Extended Kalman Filter [25].

The optimal value $\hat{b}_t$ must satisfy $\nabla F(\hat{b}_t) = 0$, which implies that

$$\hat{b}_t = \frac{\int b\; f_t(b \mid \rho_{t_n}, C_{t_n})\; db}{\int f_t(b \mid \rho_{t_n}, C_{t_n})\; db}. \tag{9}$$

By Bayes' rule, $f_t(b \mid \rho_{t_n}, C_{t_n}) \cdot f_t(\rho_{t_n} \mid C_{t_n}) = f_t(\rho_{t_n} \mid b, C_{t_n}) \cdot f_t(b \mid C_{t_n})$. Furthermore, since b is independent of the sensor configurations then $f_t(b \mid C_{t'}) = f(b)$. These observations imply that

$$\hat{b}_t = \frac{\int b\; f_t(\rho_{t_n} \mid b, C_{t_n}) \cdot f(b)\; db}{\int f_t(\rho_{t_n} \mid b, C_{t_n}) \cdot f(b)\; db}. \tag{10}$$

Note that by definition of our model,

$$f_t(\rho_{t_n} \mid b, C_{t_n}) \cdot f(b) \propto \exp\left(-\tfrac{1}{2}\left(\textstyle\sum_{i=1}^{N_b} \frac{b_i^2}{\sigma_{b,i}^2} + \cdots \right.\right. \\ \left.\left. \textstyle\sum_{n=1}^{n'-1} \sum_{k=1}^{N_m(t_n)} \frac{||\rho_{t_n}(s_k) - h_{t_n}(\gamma_{t_n}(s_k))||^2}{\sigma_o^2}\right)\right), \tag{11}$$

where $n' = \min\{n \mid t_n > t\}$. Since this is a complicated function to evaluate, we assume that $\sigma_{b,i}^2$ is small and linearize $h_{t_n}(\gamma_{t_n}(s))$ around $b = \mathbf{0}$. That is, $h_{t_n}(\gamma_{t_n}(s)) \approx h_{t_n}(\gamma_{t_n}^K(s)) + \nabla h_{t_n}\left(\gamma_{t_n}^K(s)\right) \sum_i \psi_i(s, t_n) b_i$, where

$$\nabla h_t(X) = \begin{bmatrix} \frac{f}{(R_{t,3}^\top X + d_{t,3})} R_{t,1}^\top - \frac{f(R_{t,1}^\top X + d_{t,1})}{(R_{t,3}^\top X + d_{t,3})^2} R_{t,3}^\top \\ \frac{f}{(R_{t,3}^\top X + d_{t,3})} R_{t,2}^\top - \frac{f(R_{t,2}^\top X + d_{t,2})}{(R_{t,3}^\top X + d_{t,3})^2} R_{t,3}^\top \end{bmatrix}.$$

Hence, we have that

$$\rho_{t_n}(s) - h_{t_n}(\gamma_{t_n}(s)) \approx c_{t_n}(s) - A_{t_n}(s) b \tag{12}$$

where $A_{t_n}(s) := \nabla h_{t_n}\left(\gamma_{t_n}^K(s)\right) [\psi_1(s, t_n), \cdots, \psi_{N_b}(s, t_n)]$ and $c_{t_n}(s) := \rho_{t_n}(s) - h_{t_n}(\gamma_{t_n}^K(s))$. Furthermore, if we let $c_{t_n} := [c_{t_n}(s_1)^\top, \cdots, c_{t_n}(s_{N_m(t_n)})^\top]^\top$ and $A_{t_n} := [A_{t_n}(s_1)^\top, \cdots, A_{t_n}(s_{N_m(t_n)})^\top]^\top$, then

$$\begin{aligned} &\textstyle\sum_n \sum_k ||\rho_{t_n}(s_k) - h_{t_n}(\gamma_{t_n}(s_k))||^2 \\ &\quad \approx \textstyle\sum_n (c_{t_n} - A_{t_n} b)^\top (c_{t_n} - A_{t_n} b). \end{aligned} \tag{13}$$

If we let $D := \operatorname{diag}(\sigma_{b,1}^{-2}, \cdots, \sigma_{b,N_b}^{-2})$ then

$$\begin{aligned} &\tfrac{1}{\sigma_o^2} \textstyle\sum_n \sum_k ||\rho_{t_n}(s_k) - h_{t_n}(\gamma_{t_n}(s_k))||^2 + \sum_i \frac{b_i^2}{\sigma_{b_i}^2} \\ &\quad \approx \tfrac{1}{\sigma_o^2} \textstyle\sum_n (c_{t_n} - A_{t_n} b)^\top (c_{t_n} - A_{t_n} b) + b^\top D b \\ &\quad = (b - v)^\top \Lambda_t (b - v) + \tfrac{1}{\sigma_o^2} \textstyle\sum_n c_{t_n}^\top c_{t_n} - v^\top \Lambda_t v \end{aligned}$$

where $v := \Lambda_t^{-1} \left(\frac{1}{\sigma_o^2} \sum_{n=1}^{n'-1} A_{t_n}^\top c_{t_n}\right)$ and $\Lambda_t := D + \frac{1}{\sigma_o^2} \sum_{n=1}^{n'-1} A_{t_n}^\top A_{t_n}$.

Therefore, by replacing the previous approximations into Equation 10, we come up with the following estimate:

$$\hat{b}_t \approx \frac{\int b\; \exp\left(-\frac{1}{2}(b - v)^\top \Lambda_t (b - v)\right)\; db}{\int \exp\left(-\frac{1}{2}(b - v)^\top \Lambda_t (b - v)\right)\; db}, \tag{14}$$

which yield

$$\hat{b}_t \approx \left(D + \frac{1}{\sigma_o^2} \sum_{n=1}^{n'-1} A_{t_n}^\top A_{t_n}\right)^{-1} \left(\frac{1}{\sigma_o^2} \sum_{n=1}^{n'-1} A_{t_n}^\top c_{t_n}\right). \tag{15}$$

C. Optimal View Selection

In order to determine an optimal sensor configuration, we define the following cost function:

$$G(C_t) = \mathbf{E}\left[||\gamma_t - \hat{\gamma}_t||^2 \mid (\rho_{t_n}, C_{t_n}) \text{ for } 0 < t_n \leq t, C_t\right]. \tag{16}$$

This cost measures the expected estimation error before the image is acquired given that a configuration C_t is chosen.

Note that:

$$G(C_t) = \int \int ||\gamma_t - \hat{\gamma}_t||^2 f_{t_{n'-1}}(\rho_t, b | \rho_{t_n}, C_{t_n}, C_t)\; d\rho_t\; db,$$

and

$$\begin{aligned} &f_{t_{n'-1}}(\rho_t, b | \rho_{t_n}, C_{t_n}, C_t) \\ &\quad = f_{t_{n'-1}}(\rho_t | b, \rho_{t_n}, C_{t_n}, C_t) f_{t_{n'-1}}(b | \rho_{t_n}, C_{t_n}, C_t) \\ &\quad = f(\rho_t | b, C_t) f_{t_{n'-1}}(b | \rho_{t_n}, C_{t_n}). \end{aligned}$$

The last equality follows from the fact: (1) the observation ρ_t is independent of previous observations ρ_{t_n} given the coefficients b, and (2) b is independent of the configuration C_t if no image ρ_t is provided.

Hence, we have

$$G(C_t) = \int\int ||\gamma_t - \hat{\gamma}_t||^2 f(\rho_t|b, C_t)\, d\rho_t \cdots f_{t_{n'-1}}(b|\rho_{t_n}, C_{t_n})\, db. \quad (17)$$

Let us define

$$G_1(C_t, b) = \int ||\bar{\gamma}_t - \hat{\gamma}_t||^2 f(\rho_t|b, C_t)\, d\rho_t = \int (\alpha_0 - \Gamma\eta_t)^\top M_t (\alpha_0 - \Gamma\eta_t) f(\eta_t)\, d\eta_t \quad (18)$$

where η_t is the stacked vector of noise, $\Gamma := \frac{\Lambda_t^{-1}}{\sigma_o^2} A_t^\top$, and $\alpha_0 := b - \frac{\Lambda_t^{-1}}{\sigma_o^2}\left(\sum_{n=1}^{n'-2} A_{t_n}^\top c_{t_n} + A_t^\top \left(h_t(\gamma_t) - h_t(\gamma_t^K)\right)\right)$.

Hence,

$$\begin{aligned} G_1(C_t, b) &= \alpha_0^\top M_t \alpha_0 + \int \eta_t^\top \Gamma^\top M_t \Gamma \eta_t f(\eta_t)\, d\eta_t \\ &= \alpha_0^\top M_t \alpha_0 + \text{trace}(\int \eta_t \eta_t^\top f(\eta_t)\, d\eta_t\ \Gamma^\top M_t \Gamma) \\ &= \alpha_0^\top M_t \alpha_0 + \sigma_o^2\ \text{trace}(\Gamma^\top M_t \Gamma). \end{aligned}$$

We are left with the task to compute

$$\begin{aligned} G(C_t) &= \int G_1(C_t, b) f_{t_{n'-1}}(b|\rho_{t_n}, C_{t_n})\, db \\ &= \sigma_o^2\ \text{trace}(\Gamma^\top M_t \Gamma) + \cdots \\ &\quad \int \alpha_0^\top M_t \alpha_0 f_{t_{n'-1}}(b|\rho_{t_n}, C_{t_n})\, db. \end{aligned} \quad (19)$$

We proceed, as before, by performing some approximations. Then, $\alpha_0 \approx b - \frac{\Lambda_t^{-1}}{\sigma_o^2}\left(\sum_{n=1}^{n'-2} A_{t_n}^\top c_{t_n} + A_t^\top A_t b\right) = H(b - \hat{b}_{t_{n'-2}}) + h$, where $h := \left(I - \frac{\Lambda_t^{-1}}{\sigma_o^2} A_t^\top A_t\right)\hat{b}_{t_{n'-2}} - \frac{\Lambda_t^{-1}}{\sigma_o^2}\left(\sum_{n=1}^{n'-2} A_{t_n}^\top c_{t_n}\right)$ and $H := \left(I - \frac{\Lambda_t^{-1}}{\sigma_o^2} A_t^\top A_t\right)$. Hence,

$$\begin{aligned} &\int \alpha_0^\top M_t \alpha_0 f_{t_{n'-1}}(b|\rho_{t_n}, C_{t_n})\, db \cdots \\ &\approx h^\top M_t h + \int (b - \hat{b}_{t_{n'-1}})^\top H^\top M_t \cdots \\ &\quad H(b - \hat{b}_{t_{n'-1}}) f_{t_{n'-1}}(b|\rho_{t_n}, C_{t_n})\, db \\ &= h^\top M_t h + \text{trace}(\int (b - \hat{b}_{t_{n'-1}}) \cdots \\ &\quad (b - \hat{b}_{t_{n'-1}})^\top f_{t_{n'-1}}(b|\sigma_{t_n}, C_{t_n})\, db\ H^\top M_t H) \\ &\approx h^\top M_t h + \text{trace}\left(\Lambda_{t_{n'-1}}^{-1} H^\top M_t H\right). \end{aligned} \quad (20)$$

Finally, we have that

$$G(C_t) \approx \sigma_o^2\ \text{trace}(\Gamma^\top M_t \Gamma) + h^\top M_t h + \cdots \text{trace}(\Lambda_{t_{n'-1}}^{-1} H^\top M_t H). \quad (21)$$

We select optimal sensor placement by minimizing $G(C_t)$ via a sampling based search over the configuration space. The shape of the device is then estimated by using the methodology presented in Sec. IV-B.

D. Markerless Scenarios

In the previous subsections, we assumed that there are $N_m(t)$ markers along the 3D model at time t and we are provided with observations of these markers. That is, correspondence between observations and the 3D model is known, which is required for Eqns. 15 and 21. For the more general case, we place $N_m(t)$ virtual markers on our 3D kinematic model and find the correspondence to each observed image.

As Eqn 2 suggests, the actual model is a perturbation around the kinematic model. Since the basis functions ψ_i and the distribution of b are known a priori, we try to generate an estimation of the 3D model $\tilde{\gamma}_t$ based on Eqn 2. $\tilde{\gamma}_t$ is then projected into 2D space via Eqn 4. Let us call the projected curve $\tilde{\rho}_t$.

By defining the distance between a point x and ρ_t as $d(x, \tilde{\rho}_{t'}) = \min_{x' \in \tilde{\rho}_{t'}} ||x - x'||$. Then, we quantify the reprojection error as $\int_s d(\tilde{\rho}_{t'}(s), \rho_{t'}) ds$. We generate multiple ($\sim$ 100) models of $\tilde{\gamma}_t$ by generating random coefficients b. The one that minimizes the reprojection errors will be selected, and its projection in 2D space will be considered as the projections of virtual markers.

V. Results

We apply our online shape estimation method to continuum robots in simulation. We focus on concentric tube robots [2], [3], a class of continuum robots composed of several nested, pre-curved tubes that can each be axially translated and rotated at the robot's base. As each tube is translated and rotated, the tubes elastically interact, enabling the robot to achieve a wide variety of curvilinear shapes. Concentric tube robots have the potential to enable new minimally invasive surgical procedures for cases in which straight instruments cannot reach the clinical target, including procedures in the brain [26], [27], lung [28], and heart [29].

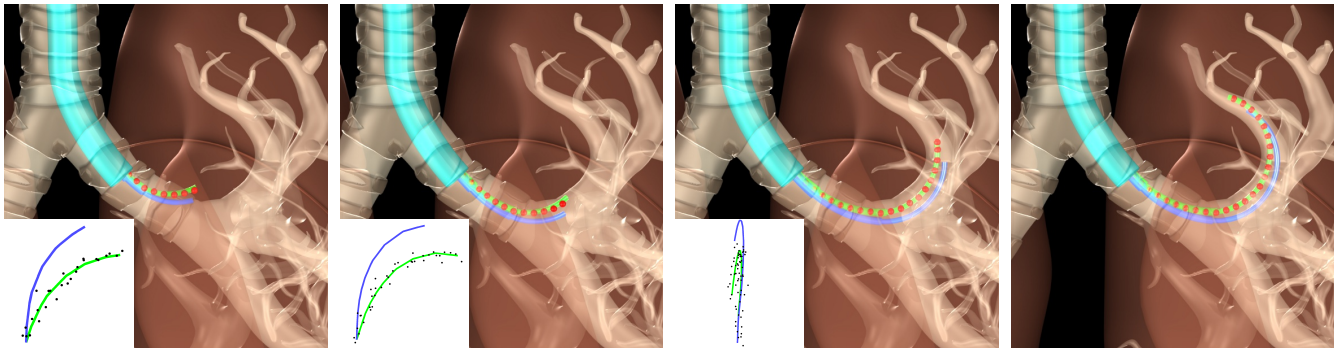
A. Evaluation Scenarios

We assume a C-arm rotates a projection image sensor (e.g. X-ray imager or optical camera) around the continuum robot in a circle (see Fig. 1). The sensor's configuration space is parameterized by an angle $C_t \in [0°, 360°]$ at time t. As is typical with clinical C-arms, the radius of the circle is 40 cm. We set the focal length of the sensor to 1 cm so projections of the model into the image domain are in the range of $[-M_p, M_p]$ where $M_p = 0.0375$ cm is the *projection magnitude*.

For our kinematic model of concentric tube robots we use a mechanically accurate model that considers both the elastic and torsional interactions of the component tubes [30]. To model uncertainty in the actual robot shape relative to the kinematic model, we note that much of the error in kinematic modeling is likely due to the modeling assumption that inner tubes and outer tubes share the same tangent vector at the point where the inner tube protrudes from the outer tube [31]. Hence, we represent modeling noise as a random deviation in the tangent vector of each tube's protrusion from its enclosing tube. We used a Gaussian distribution with a standard deviation of 0.08 radians, which resulted in average tip errors between the kinematic model and actual shape of between 0.25 cm and 0.5 cm, which is consistent with physical experiments [31].

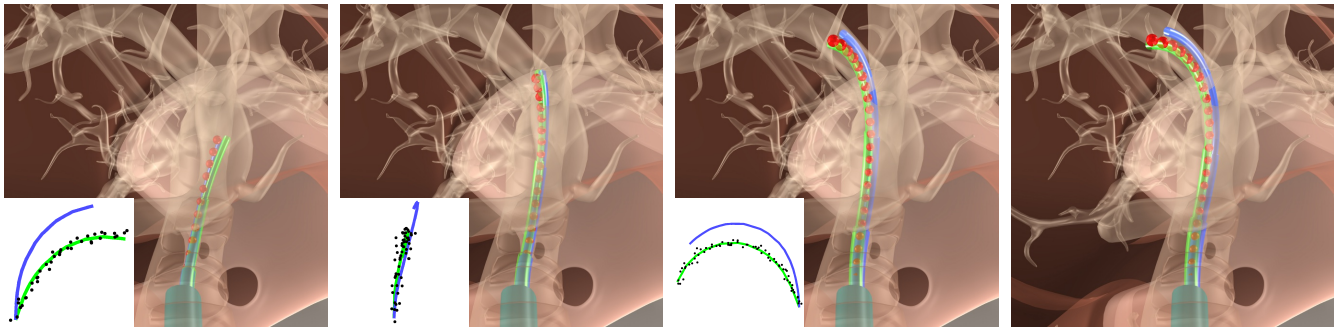
We consider two scenarios shown in Fig. 3 in which a concentric tube robot is used to reach a specific location within a lung. The final configuration in each scenario is approximately 7.5 cm in length. The model had $N_m(T) = 138$ for scenario 1 and $N_m(T) = 108$ for scenario 2 and present results for the markerless approach.

We compare three methods for estimating robot shape over the course of a procedure: (1) the *kinematic* model,



$t =$ 1st image, $C_t = 157°$ $t =$ 2nd image, $C_t = 175°$ $t =$ 3rd image, $C_t = 258°$ Scenario 1, $t = T$

(a) Scenario 1



$t =$ 1st image, $C_t = 166°$ $t =$ 2nd image, $C_t = 249°$ $t =$ 3rd image, $C_t = 314°$ Scenario 2, $t = T$

(b) Scenario 2

Fig. 3. Scenarios 1 and 2 involve maneuvering a concentric tube robot, which is deployed via a bronchoscope (cyan), through bronchial tubes in a human lung to clinical targets in simulation. We show the actual concentric tube robot shape (green), the kinematic model (blue), and the 3D reconstructed shape using our optimal method (red dots) after each image acquisition and at the final time T. Inlayed are the simulated camera views from the viewpoint selected by our approach. The camera views include the actual concentric tube robot shape (green), simulated noisy segmented points along the image of the actual concentric tube robot that are used for the reconstruction (black dots), and the kinematic model for reference purposes (blue).

(2) the shape estimate obtained via our method using *random* sensor placements in which images are captured by selecting random sensor configurations C_t from a uniform distribution between $0°$ and $360°$, and (3) the shape estimate obtained using our method with *optimal* sensor placements. We also vary N_I (the number of images taken during the procedure) between 1 and 10, and we assume images are taken at equal time intervals. We compute the *error* between a shape estimate $M \in \{\text{kinematic}, \text{random}, \text{optimal}\}$ and the continuum robot's actual shape at time t as

$$E^M(t) = \sqrt{\sum_{k=1}^{N_m(t)} \sum_{j=1}^{3} \frac{(\gamma_{t,j}(s_k) - \gamma^M_{t,j}(s_k))^2}{N_m(t)}}. \qquad (22)$$

B. Learned Basis Functions

We used the learning approach in Sec. IV-A to define the basis functions. This approach used 200 simulated runs with the noise models described above. Fig. 4 illustrates a set of basis functions learned for our first scenario. As a representative set, we show the first 3 learned basis functions for the y-coordinate of the continuum robot, which explain over 99% of the total variance in the data. We note that due to the variable length of the device over time, the bases are not defined over a square domain. Furthermore, these bases are not smooth. Ridges are present at locations associated with the junction between concentric tubes.

C. Shape Estimation Results

For each scenario and online shape estimation method, we ran 200 trials with the actual robot shape being determined by the kinematic model with random noise as discussed above. For each trial we computed the error as a function of time using equation 22. We illustrate the error for the case of $N_I = 3$ for scenarios 1 and 2 in Fig. 5. When using the kinematic model (top), the error grows as the procedure progresses. For shape estimates obtained using our method, at every time point where an image is obtained (shown by the red bar) the error in reconstruction declines. As the amount of time increases since the last acquired image, the error slowly rises as imaging data gradually becomes obsolete and the estimation becomes more dependent on the kinematic model. After 2 images, the error drops as more images are acquired until sub-millimeter error is achieved.

We also evaluated the error of the tip location of the device at the end of the procedure. Using the kinematic model, the tip error averaged 2.9 mm for scenario 1 and 3.7 mm for scenario 2. For scenario 1, using our approach the tip error was reduced to 1.1 mm with $N_I = 5$ images and less than 1 mm with $N_I = 8$ images. For scenario 2, using our approach

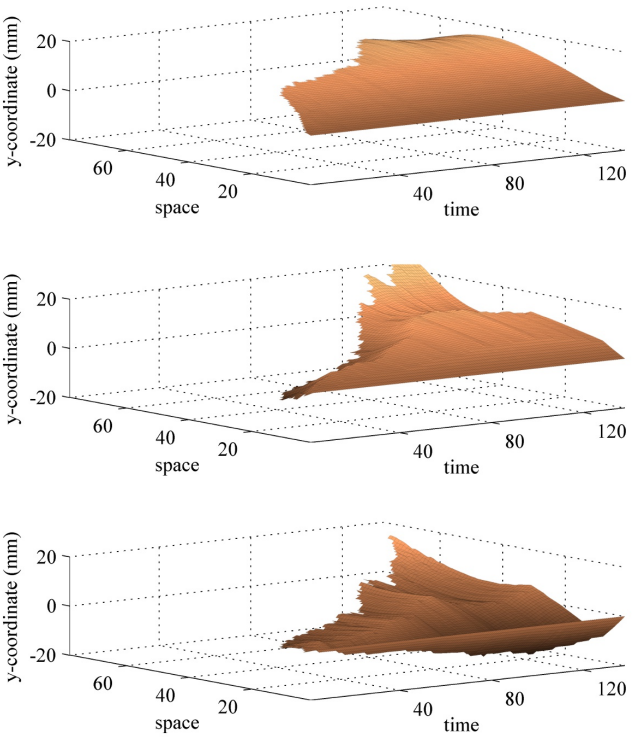


Fig. 4. Basis functions for the y-coordinate of the continuum robot.

the tip error was reduced to 1 mm with $N_I = 5$ images and less than 0.8 mm with $N_I = 8$ images. Our method accurately estimates the entire shape of the continuum robot, including the tip location.

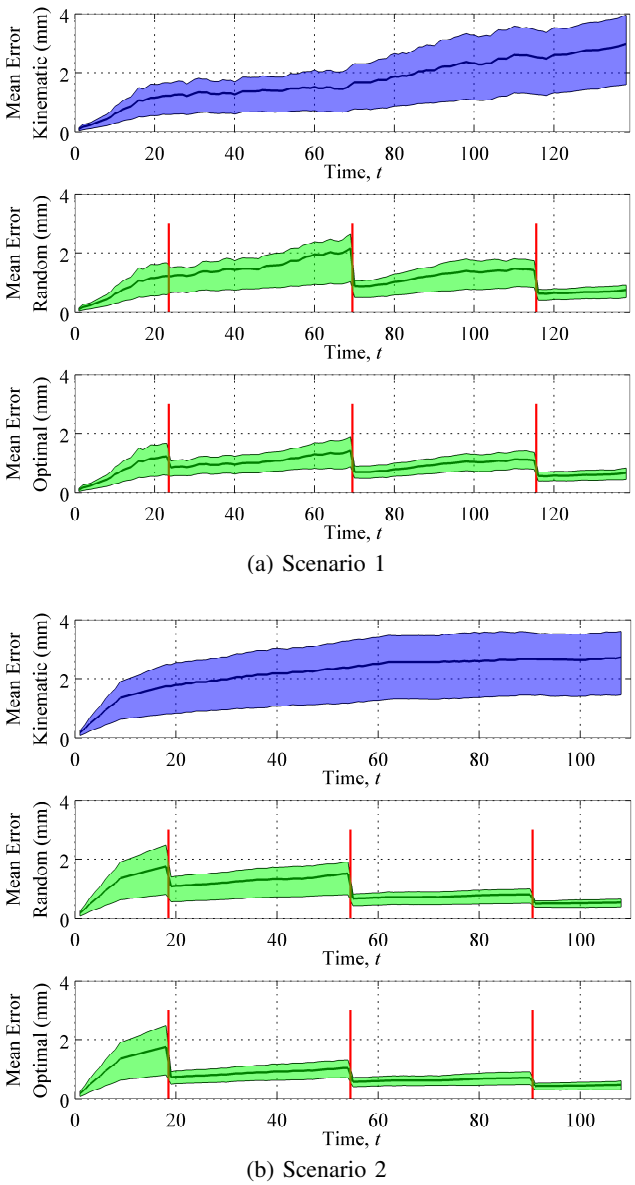
We also evaluated 3D shape estimation accuracy as a function of the number of X-ray images N_I taken during the procedure. For each trial, we compute the *average error*:

$$\bar{E}^M = \int_{t=0}^{T} \frac{E^M(t)}{T} dt. \tag{23}$$

In Fig. 6, we display the mean of the average error over the 200 trials for each scenario for the three methods for estimating robot shape. In both scenarios, our estimation approach with optimal viewpoint selection performs better than our approach with random sampling of viewpoints and substantially better than the kinematic model alone for any given positive number of images chosen. As expected, accuracy improves as the number of images increases, but the improvement levels off quickly. The results indicate that, when using our approach with optimal viewpoint selection, only a small number of X-ray images is required to accurately and continuously estimate continuum robot shape during a task.

VI. Conclusions and Future Work

We presented a new method for continuously estimating the shape of a continuum robot over the duration of a medical procedure while using a small number of X-ray projection images (e.g., radiographs or fluoroscopy images). Our new



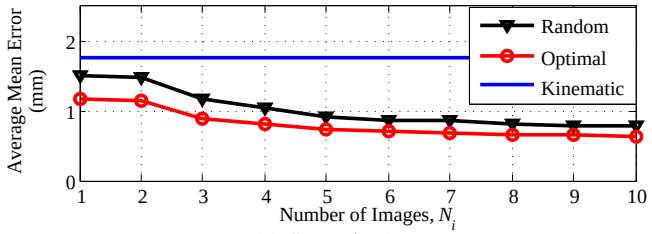
(a) Scenario 1

(b) Scenario 2

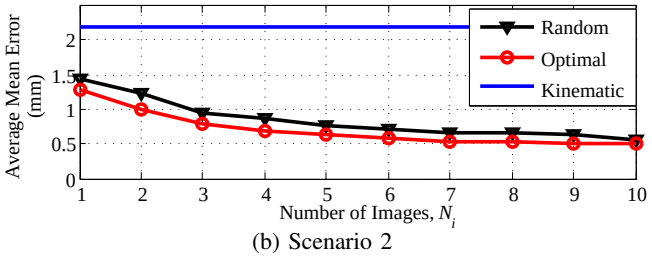
Fig. 5. Mean error $E^M(t)$ as a function of time for scenarios 1 and 2 for three shape estimation approaches: the kinematic model, our estimation approach with random sampling of viewpoints, and our estimation approach with optimal viewpoint selection. We set $N_I = 3$ and acquire images at the times of the red bars. The shaded envelope shows the range of the middle 50% of the data.

method represents the shape of the continuum robot over time as a deformable surface which can be described as a linear combination of time and space bases. We estimate the bases by effectively combining the robot's shape estimate from its kinematics model with data extracted from X-ray projection images as they are taken. We optimize the viewpoint of the X-ray sensor for each image to maximize the quality of the shape estimate.

We evaluated our method using simulated concentric tube robot procedures. We demonstrated that, using only 3 to 10 projection images, we can estimate continuum robot shape continuously over time with significantly higher accuracy



(a) Scenario 1



(b) Scenario 2

Fig. 6. Mean average error $\bar{E}^M$ as a function of the number N_I of images acquired for scenarios 1 and 2 for three shape estimation approaches: the kinematic model, our estimation approach with random sampling of viewpoints, and our estimation approach with optimal viewpoint selection. With only a small number of X-ray images, our method can accurately estimate continuum robot shape.

than kinematics modeling alone. In future work, we plan to investigate methods to further improve accuracy and to evaluate our approach in new scenarios and for other continuum robots.

VII. Acknowledgments

The authors thank Nathan Dierk for helping create visualizations and Robert J. Webster III for insightful discussions and creating the robot depicted in Fig. 1(bottom).

References

[1] K. B. Reed, A. Majewicz, V. Kallem, R. Alterovitz, K. Goldberg, N. J. Cowan, and A. M. Okamura, "Robot-assisted needle steering," *IEEE Robotics and Automation Magazine*, vol. 18, no. 4, pp. 35–46, Dec. 2011.

[2] R. J. Webster III, A. M. Okamura, and N. J. Cowan, "Toward active cannulas: Miniature snake-like surgical robots," in *Proc. IEEE/RSJ Int. Conf. Intelligent Robots and Systems (IROS)*, Oct. 2006, pp. 2857–2863.

[3] P. Sears and P. E. Dupont, "A steerable needle technology using curved concentric tubes," in *Proc. IEEE/RSJ Int. Conf. Intelligent Robots and Systems (IROS)*, Oct. 2006, pp. 2850–2856.

[4] K. Xu and N. Simaan, "An investigation of the intrinsic force sensing capabilities of continuum robots," *IEEE Trans. Robotics*, vol. 24, no. 3, pp. 576–587, June 2008.

[5] A. Degani, H. Choset, A. Wolf, and M. A. Zenati, "Highly articulated robotic probe for minimally invasive surgery," in *Proc. IEEE Int. Conf. Robotics and Automation (ICRA)*, May 2006, pp. 4167–4172.

[6] D. J. Brenner and E. J. Hall, "Computed tomography–an increasing source of radiation exposure." *The New England Journal of Medicine*, vol. 357, no. 22, pp. 2277–2284, Nov. 2007.

[7] M. Yi, S. Soatto, J. Kosecka, and S. S. Sastry, *An Invitation to 3-D Vision: From Images to Geometric Models*. New York, NY: Springer-Verlag, 2004.

[8] W.-S. Lee and T. Poston, "Rapid 3d tube reconstruction from nearby views," in *Int. Conf. in Central Europe on Computer Graphics and Visualization*, Feb. 1997, pp. 262–271.

[9] R. J. Webster III, J. M. Romano, and N. J. Cowan, "Mechanics of precurved-tube continuum robots," *IEEE Trans. Robotics*, vol. 25, no. 1, pp. 67–78, Feb. 2009.

[10] J. M. Croom, D. C. Rucker, J. M. Romano, and R. J. Webster III, "Visual sensing of continuum robot shape using self-organizing maps," in *Proc. IEEE Int. Conf. Robotics and Automation (ICRA)*, May 2010, pp. 4591–4596.

[11] J. Burgner, S. D. Herrell, and R. J. Webster III, "Toward fluoroscopic shape reconstruction for control of steerable medical devices," in *Proc. ASME 2011 Dynamic Systems and Control Conference*, Oct. 2011, pp. 791–794.

[12] A. Reiter, R. E. Goldman, A. Bajo, K. Iliopoulos, N. Simaan, and P. K. Allen, "A learning algorithm for visual pose estimation of continuum robots," *Proc. IEEE/RSJ Int. Conf. Intelligent Robots and Systems (IROS)*, pp. 2390–2396, Sept. 2011.

[13] H. Ren and P. E. Dupont, "Tubular enhanced geodesic active contours for continuum robot detection using 3d ultrasound," *Proc. IEEE Int. Conf. Robotics and Automation (ICRA)*, pp. 2907–2912, May 2012.

[14] S. Tully, G. Kantor, M. A. Zenati, and H. Choset, "Shape estimation for image-guided surgery with a highly articulated snake robot," in *Proc. IEEE/RSJ Int. Conf. Intelligent Robots and Systems (IROS)*, Sept. 2011, pp. 1353–1358.

[15] N. Abolhassani, R. V. Patel, and M. Moallem, "Needle insertion into soft tissue: a survey," *Medical Engineering & Physics*, vol. 29, no. 4, pp. 413–431, May 2007.

[16] T. Arbel and F. P. Ferrie, "Entropy-based gaze planning," *Image and Vision Computing*, vol. 19, no. 11, pp. 779–786, Sept. 2001.

[17] J. Denzler and C. Brown, "An information theoretic approach to optimal sensor data selection for state estimation," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 24, no. 2, pp. 145–157, Feb. 2002.

[18] J. Denzler, M. Zobel, and H. Niemann, "Information theoretic focal length selection for real-time active 3d object tracking," in *Proc. IEEE Int. Conf. Computer Vision (ICC)*, Oct. 2003, pp. 400–407.

[19] M. P. Vitus and C. J. Tomlin, "Sensor placement for improved robotic navigation," in *Proc. Robotics: Science and Systems*, June 2010.

[20] S. Wenhardt, B. Deutsch, J. Hornegger, H. Niemann, and J. Denzler, "An information theoretic approach for next best view planning in 3-d reconstruction," in *Proc. Int. Conf. Pattern Recognition (ICPR)*, Aug. 2006, pp. 103–106.

[21] M. Paladini, A. D. Bue, M. Stosic, M. Dodig, J. Xavier, and L. Agapito, "Factorization for non-rigid and articulated structure using metric projections," in *Intl. Conf. on Computer Vision and Pattern Recognition (CVPR)*, June 2009, pp. 2898–2905.

[22] J. Xiao and T. Kanade, "Non-rigid shape and motion recovery: Degenerate deformations," in *Intl. Conf. on Computer Vision and Pattern Recognition*, June 2004, pp. 668–675.

[23] H. Park, T. Shiratori, L. Matthews, and Y. Sheikh, "3d reconstruction of a moving point from a series of 2d projections," in *European Conf. on Computer Vision (ECCV)*, 2010, pp. 158–171.

[24] J. Ramsay and B. Silverman, *Functional Data Analysis*. Springer, 2005.

[25] G. Welch and G. Bishop, "An introduction to the Kalman filter," University of North Carolina at Chapel Hill," Tech Report TR 95-041, 2006.

[26] J. Burgner, P. J. Swaney, D. C. Rucker, H. B. Gilbert, S. T. Nill, P. T. Russell III, K. D. Weaver, and R. J. Webster III, "A bimanual teleoperated system for endonasal skull base surgery," in *Proc. IEEE/RSJ Int. Conf. Intelligent Robots and Systems (IROS)*, Sept. 2011, pp. 2517–2523.

[27] L. G. Torres and R. Alterovitz, "Motion planning for concentric tube robots using mechanics-based models," in *Proc. IEEE/RSJ Int. Conf. Intelligent Robots and Systems (IROS)*, Sept. 2011, pp. 5153–5159.

[28] L. G. Torres, R. J. Webster III, and R. Alterovitz, "Task-oriented design of concentric tube robots using mechanics-based models," in *Proc. IEEE/RSJ Int. Conf. Intelligent Robots and Systems (IROS)*, Oct. 2012, pp. 4449–4455.

[29] N. V. Vasilyev and P. E. Dupont, "Robotics and imaging in congenital heart surgery," *Future Cardiology*, vol. 8, no. 2, pp. 285–296, 2012.

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

[31] D. C. Rucker, "The mechanics of continuum robots: Model-based sensing and control," Ph.D. dissertation, Vanderbilt University, 2011.