Unified Tracking and Shape Estimation for Concentric Tube Robots

Alessandro Vandini, *Student Member, IEEE*, Christos Bergeles, *Member, IEEE*, Ben Glocker, Petros Giataganas, *Student Member, IEEE*, and Guang-Zhong Yang*, *Fellow, IEEE*

***Abstract*—Tracking and shape estimation of flexible robots that navigate through the human anatomy are prerequisites to safe intracorporeal control. Despite extensive research in kinematic and dynamic modelling, inaccuracies and shape deformation of the robot due to unknown loads and collisions with the anatomy make shape sensing important for intra-operative navigation. To address this issue, vision-based solutions have been explored. The task of 2D tracking and 3D shape reconstruction of flexible robots as they reach deep-seated anatomical locations is challenging, since the image acquisition techniques usually suffer from low signal-to-noise ratio (SNR) or slow temporal responses. Moreover, tracking and shape estimation are thus far treated independently despite their coupled relationship. This paper aims to address tracking and shape estimation in a unified framework based on Markov Random Fields (MRF). By using concentric tube robots as an example, the proposed algorithm fuses information extracted from standard monoplane X-ray fluoroscopy with the kinematics model to achieve joint 2D tracking and 3D shape estimation in realistic clinical scenarios. Detailed performance analyses of the results demonstrate the accuracy of the method for both tracking and shape reconstruction.**

I. Introduction

Continuum robots are emerging surgical platforms that can navigate through the human anatomy to reach deep-seated pathological sites [2], [3]. A representative continuum robot is the concentric tube robot [4], [5]. With diameter similar to a catheter but with the capability to exercise forces at its distal end due to its increased stiffness, the concentric tube robot has been proposed for several applications that require intracorporeal navigation. For example, Gosline et al. [6] investigated cardiosurgical applications, Burgner et al. [7] neurosurgical applications, Hendrick et al. [8] applied it to transurethral prostate surgery, and Lin et al. [9] to eye surgery.

Safe teleoperation of concentric tube robots during minimally invasive surgery (MIS) requires careful monitoring of their trajectory within the patient. Their kinematic and dynamic modelling [4], [5], [10], however, may lead to inaccurate tip position and shape estimation due to uncertainties in the joint values, mechanical parameters, and forces affecting the robot's behaviour. For example, shape deformations caused by collisions with the anatomy cannot be fully modelled, despite the body of work investigating the effect of known distal and distributed forces on the robot's shape [10], [11]. In fact, the intraoperative sensing of these forces is not always possible. Therefore, accurate continuous exteroceptive shape sensing is required for safe intraoperative control. Finally, shape sensing is necessary for assistive control such as path planning [12], [13] and collision detection [14].

A. Related Work

In general, shape sensors can be categorised based on their operating principles into electromagnetic, optical, and vision-based. Electromagnetic shape sensors usually provide discrete measurements of the shape of the robot, such as position of the tip or a limited number of points along its length [15], [16]. These sensors suffer from environment-related inaccuracies due to the interference of ferromagnetic materials. Optical fibres fabricated with FBGs have gained interest for shape sensing in recent years. Although FBG shape sensors can provide accurate and fast shape measurements of continuum robots, their integration with small flexible robots remains a challenge and adds significant cost [17], [18], [19], [20], [21].

Vision-based sensing can provide accurate shape measurements of the robot without requiring hardware modifications. Thus far, this sensing technology has mainly been used for general continuum robots, or implemented using ad hoc imaging configurations. Work presented in [22] and [23] demonstrates that vision-based shape sensing of a continuum robot can have improved results over direct kinematics modelling. Shape estimation using shape-from-silhouette and three orthogonal cameras is proposed in [24]. In addition, self-organising maps and stereo vision are employed in [25] to sense the shape of a continuum robot.

More recently, techniques that rely on intraoperative imaging modalities, such as endoscopy and fluoroscopy, have been proposed. The pose of a flexible instrument in endoscopic images is estimated by either marker-based or marker-less methods using virtual visual servoing [26]. In [27], the shape of a continuum robot is estimated by tracking it in endoscopic images while considering its workspace constraints.

Shape estimation of a concentric tube robot is achieved in [28] by triangulating robot points that are segmented on stereo images acquired from biplanar X-ray fluoroscopy. Biplanar fluoroscopy increases, however, the delivered radiation, and

A. Vandini, P. Giataganas, and G.-Z. Yang are with the Hamlyn Centre for Robotic Surgery, Imperial College London, SW7 2AZ, London, United Kingdom. A. Vandini was supported by EPSRC via EP/L020688/1.

C. Bergeles is with the Translational Imaging Group, Centre for Medical Image Computing, Department of Medical Physics and Biomedical Engineering, UCL, WC1E 6BT, London, United Kingdom. He was supported by an EPSRC-funded UCL Future Leaders Award.

B. Glocker is with Biomedical Image Analysis Group, Imperial College London, SW7 2AZ, London, United Kingdom.

A preliminary version of this work was presented at the 2015 IEEE/RSJ Int. Conf. Intelligent Robots and Systems [1]. The current manuscript proposes and evaluates a radically improved algorithm.

*Correspondence: g.z.yang@imperial.ac.uk

is generally not clinically preferred. Shape reconstruction of a robotic catheter using optimal monoplane C-arm positioning and appearance priors is presented in [29]. A vision-based shape sensor of a continuum robot based on optimal viewpoints of a monoplane C-arm and deformable surface parametrisation is proposed in [30]. Shape estimation reaches good accuracy but the method is only evaluated with simulated data and requires offline learning of the basis functions that model the deformations. Moreover, both methods proposed in [29], [30] rely on adaptive positioning of the C-arm to achieve shape reconstruction. This is not always possible within the clinical workflow due to the limited workspace in the operating theatre. The work presented in [31] uses fluoroscopy for pose estimation of a snake-robot. The algorithm is based on a computationally intensive intensity-based 2D/3D registration and the validation is conducted on a benchtop setup without collisions or 3D ground truth shape of the robot.

B. Contributions

In our preliminary work in [1], accurate intraoperative tracking of a concentric tube robot in monoplane fluoroscopy is fused with the robot kinematics, acting as a shape-prior, via a fast 2D/3D non-rigid registration to achieve intraoperative shape sensing. Our current paper enhances this work in a number of ways.

In this paper we propose a continuum robot shape-estimation algorithm, which is automatic, accurate, and relies on standard intraoperative monoplane fluoroscopy. In addition, the algorithm does not require the repositioning of the C-arm, making it applicable to restricted clinical environments or, more importantly, those without a robotic C-arm. Furthermore, our algorithm is general and only requires the joint-to-shape mapping (*i.e.* kinematics) of the continuum robot. To illustrate key steps involved, as an application we use concentric tube robots. More specifically, we introduce several contributions beyond our previous work on intraoperative shape sensing [1]:

- The algorithm unifies the tasks of tracking and estimating the shape of the robot from intraoperative images. In [1], these tasks were executed separately and the tracking algorithm relied only on image information; noisy images led to erroneous shape estimation. The new formulation, however, solves the two tasks simultaneously as a single MRF-based energy minimisation problem considering image and kinematics information concurrently. Thus, tracking, i.e. image segmentation, and shape reconstruction are simultaneously improved.
- Regularisation terms based on the mechanical variables of the robot are introduced to incorporate robot-based priors into the shape estimation framework.
- Tube endpoints are detected in the images to guide the unified tracking and shape estimation algorithm.
- The robot is modelled using B-spline curves, which allow a smoother and more accurate description of its shape compared to [1].
- Detailed simulation with synthesised X-ray-like images and experimental evaluations of the algorithm using two concentric tube robot designs demonstrate the robustness and clinical value of our method.

The paper is structured as follows: the kinematics of the continuum robot and the proposed vision-based shape sensing algorithm are described in Sec. II. The evaluation of the algorithm in simulation and laboratory experiments is reported in Sec. III, IV and V. The relative strength of the proposed method and its potential pitfalls are discussed in Sec. VI and final conclusions are presented in Sec. VII. Table I lists the nomenclature.

II. Vision-based Shape Sensing for Continuum Robots

This section proposes shape sensing of concentric tube robots by unifying robot tracking and shape reconstruction into a single energy minimisation problem. It will define every term of the overall energy that will be minimised, i.e. its three components:

$$E_{track\&rec} = E_{tracking} + E_{shape} + E_{fusion} \tag{1}$$

where $E_{tracking}$ relates to tracking, i.e. segmentation of the robot in the fluoroscopic images, E_{shape} relates to a shape prior based on the mechanical variables of the robot, and E_{fusion} unifies the tracking and shape reconstruction. Each energy component of the framework is described in the following subsections. A 2D/3D non-rigid registration will be formulated using an MRF framework in Sec. II-D. This formulation will then be extended in Sec. II-E to derive $E_{track\&rec}$ as a joint energy function resulting from the tracking and shape reconstruction tasks.

A. Kinematics of Concentric Tube Robots

Concentric tube robots comprise multiple precurved telescoping tubes made from superelastic materials (usually NiTi). The relative translation and rotation of the tubes give rise to curvature interactions and control the robot shape and tip pose (see Fig. 1). Therefore, the joint variables of the robot can be considered as the relative translations, ϕ, of the tubes, and the relative rotations, α. The value of α for each tube varies along its length due to the exhibited tube twist. The shape based on the kinematics modelling of the robot is calculated by solving a Boundary Value Problem (BVP) for this relative twist angle, α, and for bending and torsional curvature $\kappa = \{\kappa_x, \kappa_y, \kappa_z\}$ as functions of arc length. Based on the twist angles and curvature, the robot's centreline is estimated using matrix exponentials [4] and is defined as S_{kin}. When the kinematic variables $\{(\phi, \alpha)\}$ are given, the shape is found employing root-finding similar to [32].

We use the terminology of [4] for variable and fixed robot sections. Variable curvature sections possess two tubes of similar bending stiffness, thus, the curvature of the section depends on their combined curvature as tuned by applying relative rotations to the tubes. Fixed curvature sections consist of a single curvature. Finally, outer sections are stiffer than inner sections, with a stiffness ratio of Γ. Fig. 1 depicts a robot that has a three fixed curvature sections.

TABLE I
NOMENCLATURE

Kinematics	
ϕ_i	Relative extension of the i^{th} section or tube
α_i	Relative rotation of the i^{th} section or tube
S_{kin}	Robot shape estimated by kinematics
$\tilde{S}_{kin}$	Projections of the B-spline points on the image plane
Γ	Stiffness ratio
Tracking & Tubes' Endpoint Detection	
I_t	Image observation at frame t
Hp_t	Tracking hypothesis at frame t
$\hat{Hp}_t$	Position of the robot at frame t
r_{roi}	Radius of the sphere for the ROI; set to 15 mm
$dist$	Distance along the normal; set to 4 pixels
$\mathbf{ec}_j$	Tube endpoint candidates of j^{th} tracking hypothesis
w	Neighbourhood cardinality; set to 19
$\mathbf{end}(j)$	Tube endpoint of j^{th} tracking hypothesis
3D Robot Model	
s	Abscissa of the B-spline curve
$C(s)$	B-spline curve at abscissa s
N_i	Basis function of the B-spline curve
P_i	Position of the control points
M	Number of control points; set to 8
Joint Tracking & Shape Reconstruction	
V	Set of nodes
E	Set of edges
x_i	Label i^{th}
θ^{ij}_{proj}	Pairwise potential for re-projection error
Ω_{ij}	s values between i^{th} and j^{th} control points
$\mathbf{D}$	Re-projection error function
$E_{track\&rec}$	Overall energy to be minimised
E_{MRF}	Energy first order MRF
$E_{registration}$	Energy non-rigid registration
$E_{tracking}$	Energy tracking
E_{shape}	Energy shape prior based on robot's mechanics
E_{fusion}	Energy of tracking and shape reconstruction
$\theta^{z}_{tracking}$	Unary: likelihood of the tracking hypotheses
θ^{ij}_{shape}	Pairwise: mechanical variables of the robot
θ^{ijz}_{fusion}	Triplewise: tracking and shape reconstruction
θ^{ij}_{curv}	Pairwise: curvature of the robot tubes
θ^{ij}_{smooth}	Pairwise: smoothness of control point displacement
ι_c, ι_s	Curvature and smoothness terms; set to 5 and 100
$\mathbf{C}$	Absolute value of the curvature at s
$t(z)$	Curvature of the z^{th} tube of the robot
d_i	Displacement of the i^{th} control points
s_{2D}	2D robot shape estimated by image tracking
o_{end}	Endpoints' distance weight; set to 5
Evaluation Methodology	
S_{gt}	Ground truth robot shape during evaluation
S_{est}	Estimated robot shape during evaluation
S_{recon}	Robot shape estimated by vision-based sensing
$S_{kin\&vision}$	Shape estimated by [1]
$S_{track\&rec}$	Shape estimated by joint tracking and reconstruction
$\sigma(h)$	Noise profile of the h^{th} Gaussian distribution
$\mathbf{dis}(h,j)$	Displacement of the j^{th} point of the robot shape
$\mathbf{v}(j)$	Displacement vector for unknown "force/load"
c	Displacement for unknown "force/load"; set to 12

B. Tracking of the Robot in Fluoroscopic Images

The algorithm that robustly tracks the robotic tool in X-ray fluoroscopic video sequences builds upon the method that we have developed previously for tracking endovascular tools [33]. We refer to this algorithm as GSEG and its in-depth description together with detailed performance evaluation can be found in [33]. In this section, GSEG is summarised while the unification of tracking and shape reconstruction is derived in Sec. II-E.

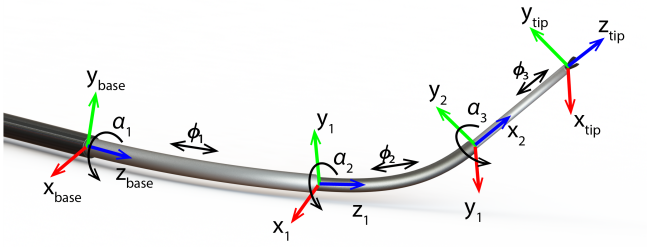


Fig. 1. The kinematic parameters of a concentric tube robot comprising a three fixed curvature sections, i.e. three tubes.

The visual appearance of a concentric tube robot in fluoroscopic frames is characterised by a continuous curve with thickness depending on the size of the robot's cross section, as can be seen in Fig. 2(a). Therefore, its appearance is similar to endovascular tools such as catheters and guidewires making GSEG a suitable algorithm for tracking the robot in fluoroscopic video sequences. Instead of being purely image driven, the upgraded GSEG presented in this work is informed by the robot kinematics to identify the most plausible robot segmentation in the image (see Sec. II-E).

In summary, the main components of the GSEG are:

- Features that represent straight segments in the image (SEGlet) and specifically designed for tracking tubular tools are used. SEGlets are detected by taking into account temporal information of the tool's segments and background structures. Thus, they represent a sparse sample of the robot in the image [see Fig. 2(a)];
- A tool-tracking formulation is defined based on generation of tracking hypotheses, i.e. the organisation of the SEGlets in plausible tool shapes [see Fig. 2(a)];
- A tool model is employed as a regularisation term during the evaluation of the tracking hypotheses through Bayesian inference. To incorporate temporal information during tracking, the tool model is recursively updated using a linear Kalman filter where constant velocity is used to describe the motion of the tool and predict its position in subsequent frames.

Given a set of tracking hypotheses, the best is selected as the position of the tool in the current frame. In the original formulation of GSEG [33], the hypotheses are evaluated in a Bayesian inference framework. Thus, the hypothesis that maximises the posterior probability $P(Hp_t|I_t)$ is the segmented position of the robot at frame t, i.e. $\hat{Hp}_t$, and is defined as:

$$\hat{Hp}_t = \arg\max_{Hp_t} P(Hp_t|I_t) \tag{2}$$

where the image observation at frame t is I_t, and Hp_t is the tracking hypothesis. Hypothesis Hp_t at time t is generated by organising and interpolating the detected SEGlets at time t in a plausible 2D tool shape (please refer to Fig. 2(a) of the manuscript for an example). Therefore, Hp_t is described by

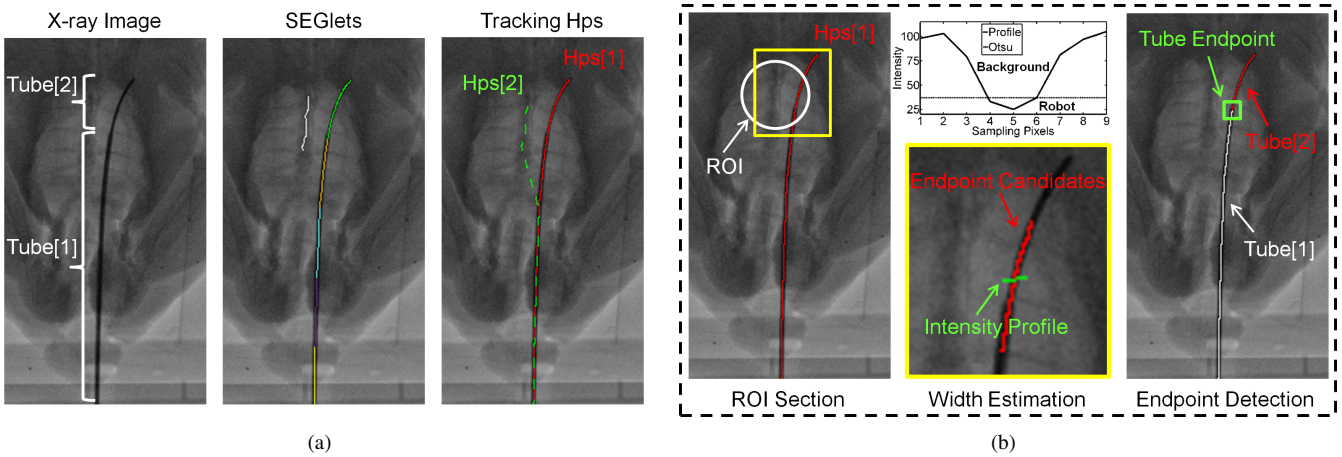


Fig. 2. (a) The steps of the tracking algorithm can be summarised as follows: First a cropped fluoroscopic frame of a concentric tube robot within a tissue model is shown, followed by the detection of the SEGlets and their organization and interpolation in tracking hypotheses. (b) The detection steps of the tube endpoint for one of the tracking hypotheses is illustrated. This detection allows subsequently the segmentation of each single tube of the robot in the image.

a set of 2D points within the image coordinate system. This posterior probability of the robot at frame t is defined as:

$$P(Hp_t|I_t) \propto P(Hp_t)P(I_t|Hp_t) \tag{3}$$

The prior $P(Hp_t)$ is the regularisation term, which measures the distance between the tracking hypothesis Hp_t and the position of the tool model as predicted from the Kalman filter.

The conditional probability $P(I_t|Hp_t)$ provides the likelihood of the tracking hypothesis Hp_t given I_t. This term is composed by two image measurements of the robot: the "lineness" model, and the Spline Local Binary Pattern (SLBP) measurement model [34]. The "lineness" is an image feature calculated by analysing the eigenvalues of the Hessian matrix and it exhibits high responses in presence of line-like objects. Its model is described by a Gaussian distribution of its response along the robot in the image. The SLBP model is an intensity measurement based on discrete samples of pixel intensities within a local area around the tool. More details on the tracking algorithm can be found in [33].

C. Detection of the Tube Endpoints

Tube endpoints are the tip of each robot tube as illustrated in Fig. 1 and Fig. 2(b). Depending on the design of the robot, several endpoints can be present. Detection of tube endpoints allows the segmentation of each single tube of the robot in the image (see Fig. 2(b)) and direct linking of robot kinematics to image information.

Each tube endpoint is detected along the curve of each tracking hypothesis by considering both kinematics and image information. The robot's tubes have different diameters (see Fig. 2(a)), and this change in width along the robot backbone is evident in the images and is sufficient to detect each endpoint.

First, a Region Of Interest (ROI) in the image is created to limit the search area for the endpoint. For this purpose, points belonging to a sphere of radius r_{roi} centred at the tube endpoint, as estimated by kinematic analysis, are projected on the image. The projection of these points generates the ROI. Only points within the ROI and belonging to a tracking hypothesis are considered as tube endpoint candidates (see Fig. 2(b)). The set of tube endpoint candidates for each tracking hypothesis is defined as $\mathbf{ec}$.

Second, for each candidate, the local size/width of the tube cross section in the image is calculated. The intensity of the image is sampled for each candidate within distance $dist$ along the normal to the robot backbone. This set of values represents the intensity profile of the cross section of the robot around the candidate points as shown in Fig. 2(b). It includes, however, background points since the size of the cross section of the robot is still unknown. Otsu's thresholding [35] is used to classify the set of intensity values into the two classes: a robot class, and a background class. The class with the lowest mean intensity corresponds to the robot, since the robot appears dark in X-ray images due to its metallic nature. In general, metallic objects appear darker than the anatomy. The variance between the contrast of these two class of objects depends on different factors, including the patient's anatomy, C-arm settings and markers used. Therefore, we do not use a fixed threshold to segment the tube from the anatomy but an adaptive algorithm, i.e. Otsu's thresholding, which handles this unknown variation in contrast. The cardinality of the set of pixels assigned to the robot class corresponds to the width measured in pixels belonging to the tube for that point on the tracking hypothesis.

As a result, for all the candidate points in the ROI, the width of the robot is now known in pixels. A tube endpoint is the point that exhibits maximal width difference from its neighbouring points along the tracking hypothesis. Formally, this is written as:

$$\mathbf{end}(j) = \arg\max_{k \in \mathbf{ec}_j} \left| \sum_{i=k-1}^{-w} \mathbf{ec}_j(i) - \sum_{i=k}^{w+1} \mathbf{ec}_j(i) \right| \tag{4}$$

where k is a tube endpoint candidate, w is a constant and represents the cardinality of the neighbourhood, and j is the j^{th} tracking hypothesis. An example of detection of the tube endpoint for a single tracking hypothesis is shown in Fig. 2(b).

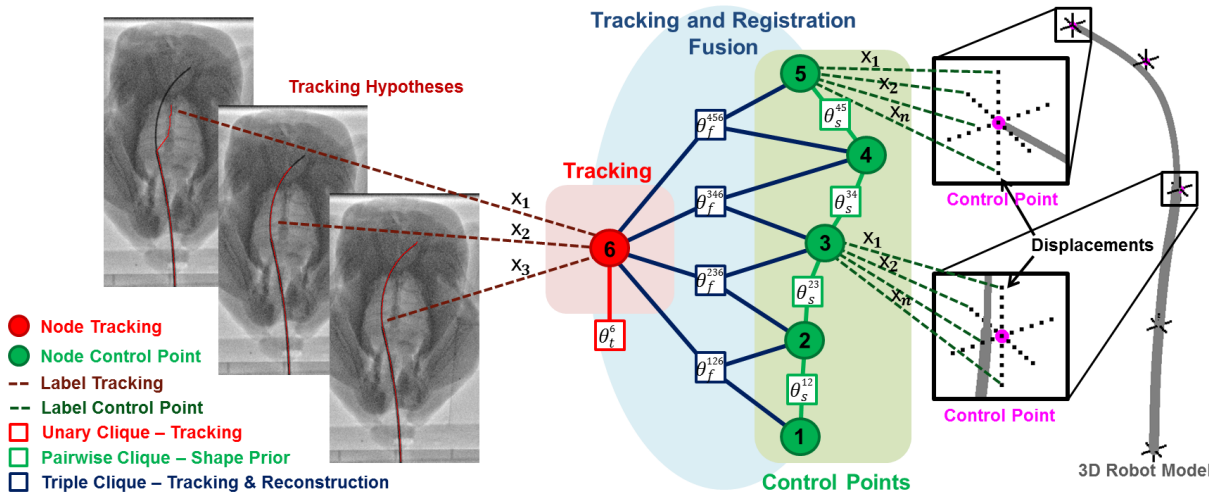


Fig. 3. A graphical representation of the framework that unifies both tracking and non-rigid registration is shown in the centre. On the left, tracking hypotheses are associated to the tracking node as labels. On the right, a label set composed of 3D displacements is assigned at each control point. In addition, the 3D robot model is shown together with a close-up of two control points and their possible spatial displacements.

D. Non-Rigid Registration for Shape Reconstruction

This section briefly introduces a new robot-shape modelling scheme based on B-splines. In addition, a 2D/3D non-rigid registration algorithm for shape reconstruction of continuum robots is formulated using an MRF framework. This formulation then leads to the unification of the tracking and registration tasks as a single energy minimisation problem, which will be presented in Sec. II-E.

1) 3D Robot Model: The robot shape is modelled by a 3D B-spline curve, which represents the robot centreline. B-splines are extensively used for modelling flexible interventional tools, such as catheters and guidewires, in 2D tracking and detection algorithms [36], [37]. This model allows a reduction of the dimensionality of the representation of the curve and inherently provides smoothness. B-splines describe the robot's shape accurately and robustly. In our case, a polynomial basis is used, similar to [38], [39]. A linear combination of control-point positions generates a B-spline, which is formally defined as:

$$C(s) = \sum_{i=1}^{M} N_i(s) P_i \text{ where } s \in [0,1] \tag{5}$$

where M is the number of control points, P_i their positions and N_i the i^{th} basis function of the B-spline curve. Following this notation, the points of S_{kin} are interpolated using the B-spline model. In addition, the control points are equally distributed along the robot and they do not need to lie on top of the parameterised curve. An illustration of the 3D robot model is shown in Fig. 3. It can be seen that control points 1 and 5 are on the curve, while points 2, 3, 4 are away from the curve.

2) Non-Rigid Registration using MRFs: In order to reconstruct the 3D shape of the continuum robot, a 2D/3D non-rigid registration is required to incorporate the information from the fluoroscopic/projection images into the reconstruction process. The 2D/3D registration problem can be effectively formulated as an MRF optimisation problem. First, the underlying graph structure is described, followed by the introduction of individual energy terms.

A graph G is defined by a set of discrete variables V, i.e. the graph nodes, and a set of edges E. Edges connecting two or more nodes represent dependencies between variables. Each variable is associated with a control point of the B-spline representing the 3D robot model, as shown in Fig. 3 by the green nodes. The variables take values from a finite set of labels, and, in our context, the set defines a discretisation of the space of control point displacements (see Fig. 3). Therefore, the assignment of a label to a particular graph node corresponds to displacing the control point of the B-spline by a certain (prescribed) magnitude.

The task of non-rigid registration then becomes a graph labelling problem wherein one seeks to assign the optimal displacements to control points such that a registration energy function is minimised. Here, the energy function corresponds to the re-projection error, which describes the distance between the tracked robot in the image plane, i.e. s_{2D}, and the projection of the points of the B-spline on the image. The energy of an MRF labelling is defined as a sum of so called clique potentials:

$$E_{MRF} = \sum_{c \in C} \theta^c(\mathbf{x}_c) \tag{6}$$

Each clique c is a subset of nodes that have inter-dependencies (including unary cliques with only one node), and $\mathbf{x}_c$ are the labels assigned to these nodes. The potential function θ^c evaluates the labelling on the subset of nodes and returns an energy, where lower energies correspond to better labelling according to some task-specific objective. The most commonly used MRF model is a first-order model that has only unary and pairwise cliques. It has been shown that for the task of interventional tool tracking, a pairwise model for which the tracking energy is solely encoded in pairwise terms yields very good performance [37]. As the assignment of a label x_i to the

i^{th} node corresponds to displacing the i^{th} control point of the B-spline, it is clear that unary potentials are not appropriate as they model the label assignment independently for each variable. The displacement of a control point has an effect on points on the spline, which are related to more than one control point, and, thus, using pairwise terms which model the simultaneous assignment of two labels, x_i and x_j, to the neighboring control points, i and j, is a good compromise between approximation and performance. The MRF energy comprising pairwise terms only is then defined as:

$$E_{registration} = \sum_{(i,j)\in E} \theta^{ij}_{proj}(x_i, x_j) \tag{7}$$

The pairwise term θ^{ij}_{proj} describes the differences between the projection of the robot in the image plane, s_{2D}, and the projection of the points of the B-spline, and is defined as:

$$\theta^{ij}_{proj}(x_i, x_j) = \sum_{s\in\Omega_{ij}} \mathbf{D}\left(C_{ij}(x_i, x_j, s)\right) \tag{8}$$

where $\mathbf{D}$ is the distance between s_{2D} and the projections of the B-spline points on the image plane, i.e. $\tilde{S}_{kin}$. Only the B-spline points with abscissas s being included between the two control points associated to the i^{th} and j^{th} nodes, namely Ω_{ij}, are considered. A similar notation is also used in [37]. The nodes i and j are displaced by the amount defined by their corresponding labels, i.e. x_i and x_j.

Such MRF energies can be efficiently optimised using discrete optimisation methods based on iterative graph-cuts. We employ the QPBO algorithm with a fusion move strategy for multi-labelling problems [40], which has been also used in [37]. Additionally, we use higher-order clique reduction as introduced in [41] enabling us to further extend the model with higher-order clique potentials leading to a unified tracking and shape reconstruction model, as discussed in the following.

E. Unified Tracking and Shape Reconstruction

It should be noted that the 2D/3D non-rigid registration as introduced in Sec. II-D2 is inherently ill-posed, as we are seeking 3D displacements via an energy function based on 1D measurements obtained from image intensities. In theory, there are multiple 3D configurations that lead to the same 2D projection, and the initial estimate of the robot configuration needs to be sufficiently close to the optimal solution. This limitation can be overcome by incorporating additional constraints on the reconstructed shape.

Here, the MRF registration framework is expanded to model the tracking and shape reconstruction of the robot as a single energy minimisation problem, which allows the integration of multiple such constraints. For this purpose, a second-order MRF framework, which includes unary, pairwise and triple clique potentials is proposed, enhancing what was presented in Sec. IIE. This formulation allows the estimation of the best tracking hypothesis together with the optimal displacements of the control points, which leads to robust, joint tracking and reconstruction of the robot shape.

A node that is associated to the tracking result is added to the set of graph nodes V. This special node will have its own label set with the number of labels equal to the number of tracking hypotheses as shown in Fig. 3 by the red node and red dotted segments. Furthermore, additional edges are introduced that connect this new node with all pairs of neighboring control point nodes to form a set of triple cliques (see blue edges in Fig. 3). These cliques represent the bridge between tracking and shape reconstruction enabling the joint formulation. The structure of the graph is summarised in Fig. 3. The corresponding joint energy is defined as:

$$\begin{aligned} E_{track\&rec} = & \sum_{z\in V} \theta^{z}_{tracking}(x_z) \\ & + \sum_{(i,j)\in E_2} \theta^{ij}_{shape}(x_i, x_j) \\ & + \sum_{(i,j,z)\in E_3} \theta^{ijz}_{fusion}(x_i, x_j, x_z) \end{aligned} \tag{9}$$

where E_2 and E_3 are the sets of edges defining the pairwise and triple cliques, and the individual potential functions of the energy are:

- $\theta^{z}_{tracking}$ describes the likelihood of the tracking hypotheses;
- θ^{ij}_{shape} is the regularisation term based on mechanical variables of the robot;
- θ^{ijz}_{fusion} is the term responsible for unifying the tracking and shape reconstruction.

The first term of (9) relates to $E_{tracking}$ in (1), the second term to E_{shape}, whereas the final one to E_{fusion}. Each term is described in more detail in the following.

1) Tracking Hypotheses: The likelihood of each tracking hypothesis is evaluated in (3) and is used as the unary potential of the energy:

$$\theta^{z}_{tracking}(x_z) = 1 - P(Hp_t|I_t) \tag{10}$$

This term introduces the tracking information into the overall energy, combining 2D image-based with 3D kinematics-based information.

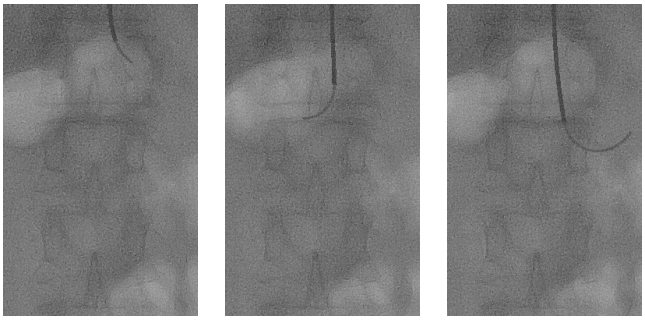
2) Shape Prior Based on Mechanical Variables: A shape-prior based on mechanical variables is introduced in the overall energy as a pairwise potential that constrains the solution space to plausible robot shapes. For this purpose, a regularisation term, i.e. θ^{ij}_{curv}, penalises configurations of the control points that generate curvature profiles along the robot shape different than the ones anticipated by the design specifications of the robot. In addition, a smoothness term, i.e. θ^{ij}_{smooth}, considers the fact that neighboring control points are characterised by coherent displacements. The shape prior based on mechanical variables is defined as:

$$\theta^{ij}_{shape}(x_i, x_j) = \iota_c \theta^{ij}_{curv}(x_i, x_j) + \iota_s \theta^{ij}_{smooth}(x_i, x_j) \tag{11}$$

where ι_c and ι_s are weights. The regularisation term, $\theta^{ij}_{curv}(x_i, x_j)$, based on the curvature of the robot, is zero or:

$$\theta^{ij}_{curv}(x_i, x_j) = \sum_{s\in\Omega_{ij}} \left|\mathbf{C}\left(C_{ij}(x_i, x_j, s)\right) - t(z)\right| \tag{12}$$

$$\begin{cases} \text{if, the tube:} \\ \textit{has fixed curvature}, \text{ or} \\ \textit{has variable curvature}, \text{ and } \mathbf{C}\left(C_{ij}(x_i, x_j, s)\right) > t(z) \end{cases}$$



(a) (b) (c)

Fig. 4. The first (a), middle (b) and last frame (c) of the synthetic sequence which is generated in order to recreate realistic fluoroscopic sequences of the robot within the patient. This is achieved by projecting the 3D robot points on in-vivo fluoroscopic images recorded during an endovascular procedure.

where $\mathbf{C}\left(C_{ij}(x_i, x_j, s)\right)$ describes the absolute value of the curvature calculated at the abscissa s between the i^{th} and j^{th} control points, $t(z)$ describes the curvature of the z^{th} tube of the robot from its design specifications. Finally, the smoothness term is defined as:

$$\theta^{ij}_{smooth}(x_i, x_j) = ||d_i - d_j||_2 \tag{13}$$

where d_i and d_j describe the displacements of neighbouring control points associated with labels x_i and x_j.

3) Tracking and Registration Fusion: The triple clique term is responsible for unifying tracking and shape reconstruction, and is defined as:

$$\theta^{ijz}_{fusion}(x_i, x_j, x_z) = \sum_{s \in \Omega_{ij}} \mathbf{D}_{x_z}\left(C_{ij}(x_i, x_j, s)\right) \tag{14}$$

where $\mathbf{D}_{x_z}$ is the distance between the tracking hypothesis associated to the label x_z, and $\tilde{S}_{kin}$. Depending on which tube s it is associated to, the corresponding image segmentation of the tube from the tracking hypothesis is used. In addition, since 2D to 3D correspondences between the endpoints of the tracking hypothesis and $\tilde{S}_{kin}$ are known, their distances are calculated via their Euclidean distances weighted by o_{end}. Therefore, the modelling of (14) is indeed different than (8), since a different tracking hypothesis provides the projection of the robot in the image depending on label x_z. This enables the unification of the tracking and shape reconstruction into a single formulation.

III. Evaluation Methodology

Extensive evaluation of our algorithm through performance metrics such as reconstruction error of robot shape, tip position error, tracking error, and tip tracking error, quantifies the accuracy of our method and demonstrates its value.

The shape reconstruction error, $Dist(S_{gt}, S_{est})$, between the estimated shape, S_{est}, and the ground truth shape, S_{gt}, is defined as:

$$Dist(S_{gt}, S_{est}) = \frac{1}{b+f}\left(\sum_{i=1}^{b} d_{min}(S_{gt}(i), S_{est}) + \sum_{j=1}^{f} d_{min}(S_{est}(j), S_{gt})\right) \tag{15}$$

TABLE II
Robot Designs

Robot 1 - Simulations			
Section stiffness ratio: $\Gamma = 10$			
		Length	
Section Type	**Curvature [1/mm]**	**Straight [mm]**	**Curved [mm]**
(1) variable	$1/101$	0	75
(2) fixed	$1/23$	75	55
Robot 2 - Experiments			
Section stiffness ratio: $\Gamma = 5.4$			
		Length	
Section Type	**Curvature [1/mm]**	**Straight [mm]**	**Curved [mm]**
(1) fixed	$1/293$	0	160
(2) fixed	$1/60$	160	100

where $d_{min}(S_{gt}(i), S_{est})$ is the distance between the i^{th} point of S_{gt} to the closest point on S_{est}, while b and f are the number of points of S_{gt}, and S_{est}, respectively [36]. In the error evaluation, the estimated shape S_{est} can be either:

1) The shape S_{kin}, calculated using forward kinematics;
2) The shape $S_{kin\&vision}$, estimated using the method presented in [1], i.e. by fusing kinematics and vision. In addition, the algorithm presented in [33] is used to track the robot;
3) The shape $S_{track\&rec}$, estimated using the proposed method, i.e. by unifying tracking and registration;

For S_{kin} and $S_{kin\&vision}$, the points resulting from the shape estimation are interpolated using a B-spline to generate a smooth and continuous 3D curve.

The tip position error is the Euclidean distance between the tip of the estimated shape and the tip of the ground truth shape. The tracking error is measured with (15), with the difference that the tracking results (in 2D) are used instead of S_{est} (in 3D), and the ground truth position of the robot in the image (in 2D) is used instead of S_{gt} (in 3D). Finally, the tip tracking error is the Euclidean distance between the tip of the robot given by the tracking and the tip of the ground truth position of the robot in the image.

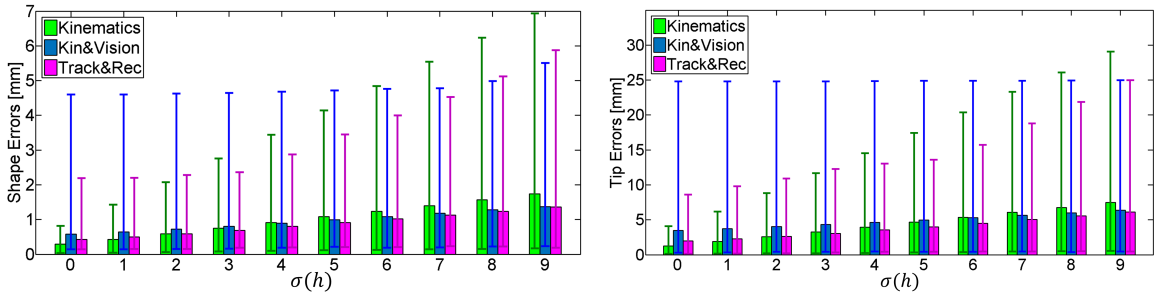
The proposed algorithm runs on a desktop PC with the following specifications: i7 $-$ 2600 at 3.40GHz and 16GB of RAM. The generation of all plausible hypothesis completes in less than 1s, while tracking and shape reconstruction in approximately 9s, with an unoptimised C++ implementation.

IV. Simulations

Robot 1, given in Table II, is used to evaluate the proposed algorithm via two simulation scenarios. The first simulation scenario evaluates shape reconstruction under noisy kinematics, and the second scenario evaluates shape reconstruction when the shape is deformed due to unknown, unmodelled, virtual forces acting on the robot from the patient's anatomy.

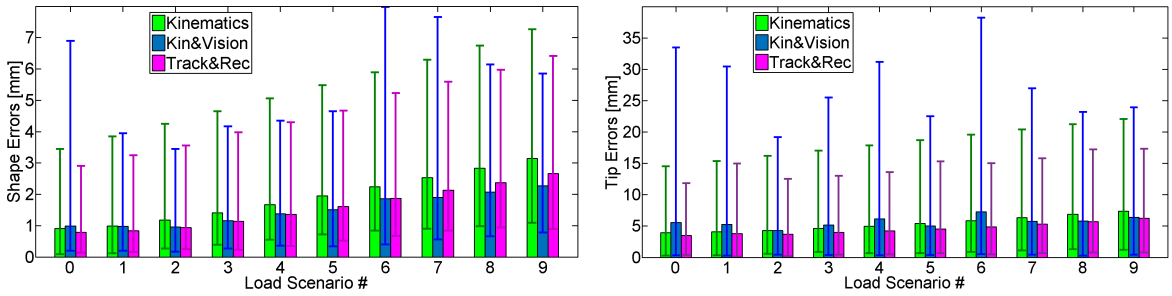
A. Robustness to Kinematic Uncertainties

In the first scenario, robot centrelines based on 145 discrete sets of joint variables, which cause the robot to perform a spiral-like trajectory, are generated. The trajectory is selected



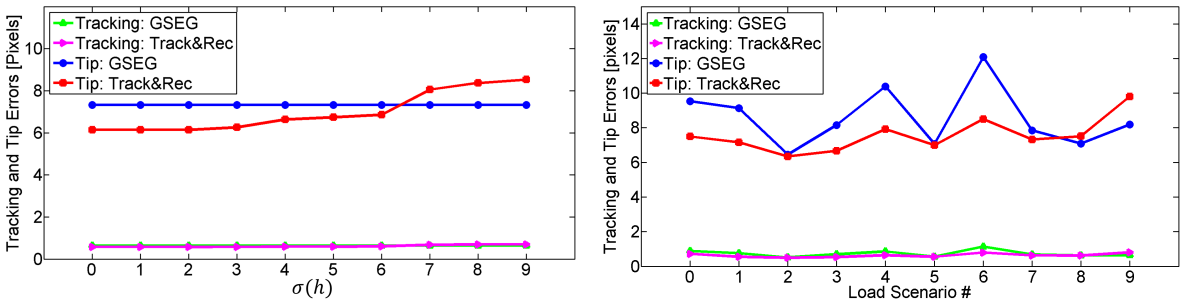
(a) (b)

Fig. 5. The mean, max and min reconstruction errors for the forward kinematics only, our previous method [1], and the proposed joint method for the first simulation scenario for 145 robot shapes per each noise case. (a) Shape reconstruction errors, and (b) tip position errors.



(a) (b)

Fig. 6. The mean, max and min reconstruction errors for the forward kinematics only, our previous method [1], and the proposed joint method for the second simulation scenario for 145 robot shapes per each load scenario. (a) Shape reconstruction errors, and (b) tip position errors.



(a) (b)

Fig. 7. The mean tracking and tip errors for GSEG and the proposed joint method for the (a) first simulation and (b) second simulation scenario for 145 robot shapes per each noise case and load scenario, respectively.

so that it includes short/long robots in both in-plane and out-of-plane configurations. These centrelines result in robot shapes that are used as the ground truth shapes. Each of the 145 centrelines is projected on the image plane using a realistic C-arm camera projection matrix. In order to recreate the fluoroscopic sequences of the robot within the patient, synthetic fluoroscopic sequences are generated. This is achieved by projecting the 3D robot points resulting from each discrete set of joint variables into images, which are part of a real in-vivo fluoroscopic sequence[1]. The projections do not consider any physical interactions between the robot and the imaged patients anatomy. The sequence is 20 s long at 7 fps, leading to 145 discrete frames, each of which is associated with a single robot shape. The trajectory of the robot and the intensity profile of the robot's tubes in the images were tuned in order to generate synthetic fluoroscopic sequences in which tracking errors were comparable with the ones measured during in vivo tracking experiments. Several frames of this sequence are shown in Fig. 4 as an example. Finally, the projection of the actual robot centreline into the image is used as ground truth position of the robot in the images.

The noise for each joint/kinematics variable type is modelled by two Gaussian distributions: $N_1(0, \sigma_1)$, $N_2(0, \sigma_2)$, where N_1 is applied to the relative extension of the i^{th} tube or section, ϕ_i, while N_2 is applied to the relative rotation of the i^{th} tube, α_i^s, and the rotation of the whole robot α_1^0. Ten noise profiles for each Gaussian distribution are generated by

[1]The sequence was recorded during an endovascular procedure, i.e. angioplasty of the iliac artery, and includes several anatomical areas such as the spine.

increasing the values of σ_1 and σ_2 as follows:

$$\sigma(h) = \sigma^{min} + h\left(\frac{\sigma^{max} - \sigma^{min}}{N}\right), h = 0, \cdots, N-1 \quad (16)$$

where $N = 10$, $\sigma_1^{min} = 1\,\text{mm}$, $\sigma_1^{max} = 3.7\,\text{mm}$, $\sigma_2^{min} = 1^\circ$ and $\sigma_2^{max} = 12^\circ$. These parameters were tuned to reach realistic kinematics inaccuracies for the simulations, comparable (and more challenging) to experimental results shown in the concentric tube robot kinematics modelling literature [4], [5], [10]. The mean tip error reported in [4] ranges from $1\,\text{mm}$ to $4\,\text{mm}$, while the max tip error ranges from $2\,\text{mm}$ to $13\,\text{mm}$. Other studies measured mean tip errors on the order of $3\,\text{mm}$ and max tip error between $8\,\text{mm}$ to $9\,\text{mm}$ [5], [10]. The shapes estimated by solving the forward kinematics for these noisy kinematic values are denoted as S_{kin}.

The kinematics errors simulated above should not be considered only as arising from erroneous encoder readings, but as a result of a variety of mechanical and modelling errors including discrepancy in straight/curved length, curvature, relative tube stiffness, non-linear elasticity, unmodeled tube tolerances, friction etc. Even through several of the mechanical parameters, e.g. length, curvature, and relative stiffness can be measured preoperatively to a good degree of precision, non-linear behaviours, friction and imprecise concentricity are likely to be the major causes of discrepancy. Therefore, the performed simulations should be considered as a demonstrator of the robustness of the algorithm to all those unknown parameters that are essentially lumped together into "kinematics/joint errors". Given that concentric tube robots are envisioned as patient-specific devices, demonstrable robustness of the shape estimation algorithm to all factors will be important in practical deployment scenarios.

B. Robustness to Unmodelled Robot Deformations

The second simulation is structured similar to the first one, except that deflections due to increasing forces applied to the robot are introduced. These "forces" cause the deformation of each ground truth robot-shape. As in the first scenario, 145 discrete sets of joint variables, which cause the robot to perform a spiral-like trajectory, are generated together with the synthetic fluoroscopic sequence composed by in-vivo data. However, only one Gaussian noise profile for $N_1(0, \sigma_1)$, $N_2(0, \sigma_2)$ is used, corresponding to the middle one of the first simulation, i.e. $\sigma(4)$. Again, the shapes calculated by solving the forward kinematics for the noisy kinematic values are S_{kin}.

We model the increasing unknown "force/load" acting on the robot via a virtual anchor point located at the following coordinates in the robot frame $(150; 150; 100)$. The position of the anchor point was chosen in order to generate deformations of the robot that are outside the image plane of the C-arm. These deformations are the hardest for vision algorithms to reconstruct from a single view due to the degenerate nature of perspective projection.

The unknown "force/load" may represent a collision of the robot with part of the patient anatomy. A unit vector $\mathbf{v}(j)$ defines the orientation of the deformation of the j^{th} point of the ground truth robot-shape. This vector is calculated considering the virtual anchor point, and the j^{th} point of the robot-shape. The displacement of the j^{th} point of the robot-shape is:

$$\mathbf{dis}(h, j) = \mathbf{v}(j)\left[\frac{hjl}{c(N-1)(q-1)}\right] \quad (17)$$

where l is the length of the robot in [mm], q is the number of discrete points along the robot-shape, $h = 0, ..., N-1$, with $N = 10$ load scenarios, and $j = 0, ..., q-1$ is the index of the points along the robot-shape. The index $j = 0$ describes the robot base while $j = q-1$ describes its tip. Thus, for the load scenario $h = 0$ the ground truth robot shape is not affected by any force, while for $h = N-1$ the "force/load" causes a maximum displacement at the tip of the robot equal to $\frac{1}{c}$ of its length. In our implementation $c = 12$, which allows the investigation of a wide range of potential deflections and leads to kinematics inaccuracies similar to the ones experienced in [1] for the same simulation. The formula is used just to generate the ground truth shapes. In other words, it is a representation of a deformed shape but not a dynamic model of the robot.

To ensure that the deformation model does not introduce non-physical behaviours such as extension or compression of the tubes, a post-processing step maintains constant the length of the deflected robot. For elongated robots, this is achieved by removing points starting from the tip of the deformed shape until its length is equal to the length of the corresponding ground truth non-deformed robot, namely l. For shortened robots, points are added iteratively along the deformed-robot-backbone tangent until its length equals l. The differences in length that have to be corrected by the post-processing steps during the simulations are, on average, 1.6% of the original robot length l.

C. Results

For each set of noisy kinematics values of the first simulation and for each load scenario of the second simulation, three shape reconstruction algorithms are compared: (a) The forward kinematics only; (b) The method presented in [1], i.e. by fusing kinematics and vision[2]; (c) The proposed algorithm, which unifies tracking and registration.

The performance of the three algorithms is quantified and the reconstruction and tip position errors, in millimetres, for the first simulation scenario are reported in Fig. 5(a), and Fig. 5(b), respectively. It can be observed that the 3D tip position errors are in general greater than the shape reconstruction errors for each scenario. This is due to the fact that the error in estimating the shape of the robot propagates from the base of the robot to its most distal point, i.e. the tip, where it reaches its maximum value. The errors for the second simulation scenario are shown in Fig. 6. Finally, the 2D robot tracking and tip tracking errors are reported in Fig. 7(a) and Fig. 7(b).

Overall, the performance in terms of reconstruction and tip position errors, as well as tracking and tip errors achieved by the proposed method in both simulation scenarios reached

[2]In this evaluation, the algorithm presented in [33] is used for tracking the robot in the fluoroscopic sequences.

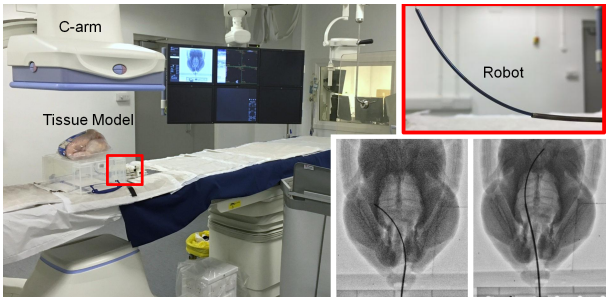


Fig. 8. The experimental setup, composed of a concentric tube robot, a monoplane C-arm and a tissue model (left). Close-up of the robot and two fluoroscopic images of the setup (right).

the best accuracy, with an average shape reconstruction error of 1.22 mm, and tip error of 4.23 mm. The results show the importance of unifying tracking and registration to perform shape sensing.

It is also worth noting that when large uncertainties on the kinematics modelling and unmodelled deformations are present, the results of the proposed approach were inferior to the results of our previous work [33]. This is mainly due to the deterioration of the performance of the tracking algorithm, as can be observed in Fig. 7(a) and Fig. 7(b). In fact, the proposed method introduces a dependency between the tracking and shape reconstruction steps, based on the kinematics of the robot. In our previous work, these tasks are executed independently and the tracking is based only on image information.

In summary, we have demonstrated that our algorithm unifies the tracking and registration tasks successfully to handle both cases of kinematic noise and unmodelled forces acting on the robot. The next section will show that tracking also improves in experimental settings.

V. Experimental Evaluation

The experimental setup is composed of a concentric tube robot, a monoplane C-arm and a chicken acting as a tissue model (see Fig. 8). The tissue model is placed in between the robot and the C-arm detector to simulate the presence of human bones and tissue. The design specifications of the concentric tube robot, which is composed of two tubes of fixed curvature are reported in Table II as Robot 2.

The monoplane C-arm is a GE Innova 4100 for interventional radiology (GE Medical Systems, Buc, France) and can be described by a pinhole camera model [42]. The camera calibration of the C-arm is performed offline to estimate its intrinsic and extrinsic parameters using a customised calibration grid and known 2D/3D landmarks. This calibration grid, a checkerboard pattern, has radio-opaque markers at the corners of the checkerboard boxes and allows the utilisation of state-of-the-art camera calibration algorithms [43]. Manual alignment of the robot with the C-arm is performed to coarsely register their coordinate systems. The origin of the robot in the C-arm coordinate system is subsequently estimated by acquiring an X-ray computed tomography image (CT-scan).

TABLE III
Overall Shape Reconstruction Errors ($\mu \pm \sigma$)

Shape Errors [mm]	Kinematics	Kin&Vision	Track&Rec
In Free Space	3.61 ± 0.86	2.85 ± 1.69	$\mathbf{2.40 \pm 1.09}$
External Loads	3.79 ± 0.74	2.00 ± 1.16	$\mathbf{1.82 \pm 1.09}$
All Experiments	3.67 ± 0.82	2.57 ± 1.57	$\mathbf{2.21 \pm 1.11}$

TABLE IV
Overall Tip Position Errors ($\mu \pm \sigma$)

Tip Errors [mm]	Kinematics	Kin&Vision	Track&Rec
In Free Space	9.82 ± 4.72	10.38 ± 9.60	$\mathbf{7.56 \pm 5.07}$
External Loads	9.17 ± 4.00	7.58 ± 5.15	$\mathbf{6.68 \pm 4.16}$
All Experiments	9.61 ± 4.45	9.47 ± 8.44	$\mathbf{7.28 \pm 4.75}$

Two data sets are collected to experimentally evaluate the performance of the proposed algorithm. Each data set is discretised in motion steps. During every step, monoplane fluoroscopic images (512×512 image size) at a fixed C-arm position of $0°$ RAO (Right Anterior Oblique) are recorded to capture the robot motions while an X-ray computed tomographic image (CT-scan) is acquired at the end of each step for collecting the ground truth via 3D manual segmentation. Then the robot centrelines, which represent the ground truth shapes of the robot for the evaluation, are extracted by thinning the manually segmented mesh. The position of the robot in each fluoroscopic image is manually annotated to generate the 2D ground truth positions. The joint values for the robot kinematics are recorded during every motion step.

A. Evaluation in Free Space

For the first data set, the robot performs in free space a spiral-like trajectory, which is discretised in 25 motion steps.

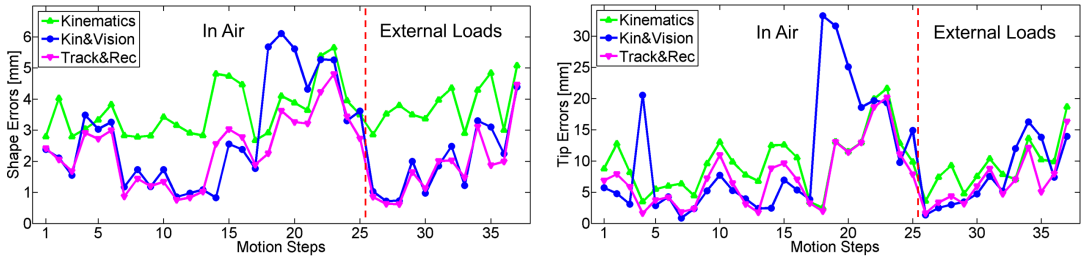
B. Evaluation with External Loads

For the second set of data, increasing loads are applied to the robot in order to simulate unmodelled contact with the tissues. The trajectory is discretised in 12 steps. In steps #26, #29, #32 and #35 the robot is unloaded while in #27, #30, #33 and #36 a force equal to 0.5 N is applied at the robot tip. In the remaining 4 steps, a 1 N force is applied at the robot tip.

C. Results

As for the simulated data, the three shape reconstruction techniques are compared for each motion step of each data set. Tracking of the robot is performed in all fluoroscopic images of both datasets and its results are used as s_{2D}. The tracking algorithm is manually initialised on the first frame. The kinematics of the robot is solved using the recorded kinematics values, generating S_{kin}.

The overall performance of the reconstruction algorithms in terms of shape reconstruction, and tip position errors, is reported in Table III, and Table IV, respectively. The accuracy of the proposed vision-based shape sensing algorithm is 2.21 mm on the shape estimation and 7.28 mm on the tip. Thus, the proposed method achieves a great improvement over the kinematics-only reconstruction of the robot shape, which



(a) (b)

Fig. 9. Shape estimation results for the experiments in air and phantom of each motion step. The red dotted line separates the data of the air to the phantom experiment while the light blue dotted rectangle indicates the steps where an external load is applied to the robot. Motion Steps #13-16 correspond to particularly out of plane robot motions. (a) Shape reconstruction errors, and (b) tip position error.

TABLE V
OVERALL TRACKING ERRORS ($\mu \pm \sigma$)

Tracking Errors [px]	GSEG	Track&Rec
In Air	1.85 ± 2.96	$\mathbf{0.42 \pm 0.84}$
External Loads	0.79 ± 1.95	$\mathbf{0.42 \pm 0.14}$
All Experiments	1.48 ± 2.70	$\mathbf{0.42 \pm 0.68}$

TABLE VI
OVERALL TIP TRACKING ERRORS ($\mu \pm \sigma$)

Tip Tracking Errors [px]	GSEG	Track&Rec
In Air	13.17 ± 21.79	$\mathbf{2.39 \pm 5.48}$
External Loads	7.52 ± 13.97	$\mathbf{2.07 \pm 2.99}$
All Experiments	11.23 ± 19.61	$\mathbf{2.28 \pm 4.77}$

has an accuracy of 3.67 mm and 9.61 mm, respectively. This amounts to a 40% error reduction for the shape estimation, and 24% error reduction for the tip. The vision-based shape sensing algorithm outperforms the method presented in [1] both in terms of shape reconstruction and tip localisation. In fact, the overall accuracy of the vision-only method is 2.57 mm and 9.47 mm for shape and tip, respectively. This leads to an error reduction for shape and tip of 14% and 23% by the proposed method compared to our previous work. Finally, great improvement on the tracking results of the robot in fluoroscopy is also achieved by the proposed method compared to the state-of-the-art tracking method GSEG, as 72%, and 80% error reduction is shown for the tracking error, and tip tracking error, respectively (see Table V, and Table VI).

The detailed shape reconstruction errors, and tip position errors of the compared reconstruction algorithms are shown in Fig. 9(a) and in Fig. 9(b), respectively. In addition, qualitative results of robot tracking in fluoroscopic images and the respective shape reconstruction for the air and phantom experiments are shown in Fig. 10 and in Fig. 11.

It can be observed that for motion steps #18-20 our previous work [1] is characterised by high errors both in shape and tip reconstruction as shown in Fig. 9(a) and in Fig. 9(b). In these steps, the tracking algorithm of [1], which is responsible for detecting the robot in the fluoroscopic images, fails to follow the robot's tube since it is attracted by the neighbouring anatomy. On the other hand, the proposed method can successfully track the robot in this challenging scenario by relying on both image information and robot kinematics via our new unified framework.

VI. DISCUSSION

The results presented in this paper demonstrate that the proposed method outperforms overall both the kinematics modelling of the robot and our previous work on vision-based shape sensing [1]. Few observations on the robustness of the method should, however, be made.

In the proposed method, the deformation range for the control points of the spline has to be defined a priori. This range, which in our implementation reaches a maximum of 6.4 mm, can limit the set of shapes that can be recovered by the algorithm. Therefore, this has to be tuned considering the application and the expected kinematics accuracy of the robot, or adaptively estimated based on the tracking error.

It is worth noting that when there are extreme inaccuracies in the kinematics modelling of the robot, the unified algorithm may fail. This is caused by the discretisation explained above and the intrinsic dependency between the tracking and shape reconstruction step, which relies on the kinematics of the robot. This issue does not occur if the two tasks are executed independently, with tracking being based only on image information as in [1]. This intuition is also supported by the evaluation results, where it can be observed that the proposed approach may even underperform with regards to [1] in terms of shape reconstruction and tip position errors when the kinematics modelling is significantly poor. In addition, simulations with values of c less than 12, i.e. corresponding to larger deformations, would lead to a deterioration of the shape estimation for the method proposed here. In order to consider the kinematics uncertainties, the contribution of the term $\theta^{z}_{tracking}$ could be adaptively increased at the expense of the other terms of the energy, i.e. θ^{ij}_{shape} and θ^{ijz}_{fusion}. This would make the selection of the label for the tracking node less dependent on the kinematics modelling and therefore immune to its inaccuracies. On the other hand, if the kinematics modelling of the robot is really accurate, the contribution of the terms θ^{ij}_{shape} and θ^{ijz}_{fusion} should be increased with respect to $\theta^{z}_{tracking}$. Fine tuning of the energy terms can make the model flexible in the presence of different uncertainties.

Finally, an important improvement introduced by the proposed method compared to our previous work [1] relies on

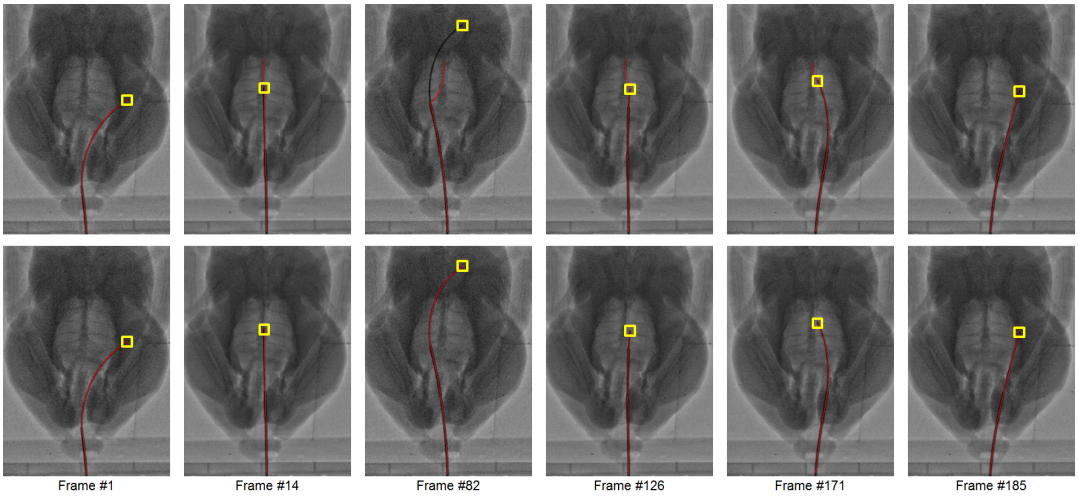


Fig. 10. Tracking of the robot in fluoroscopic images during the in air (frames #1, #14, #82, #126) and external loads (frames #171, #185) experiments: results of GSEG are reported in the first row while the ones from the proposed method are in the second row. The robot's tip is manually annotated in yellow.

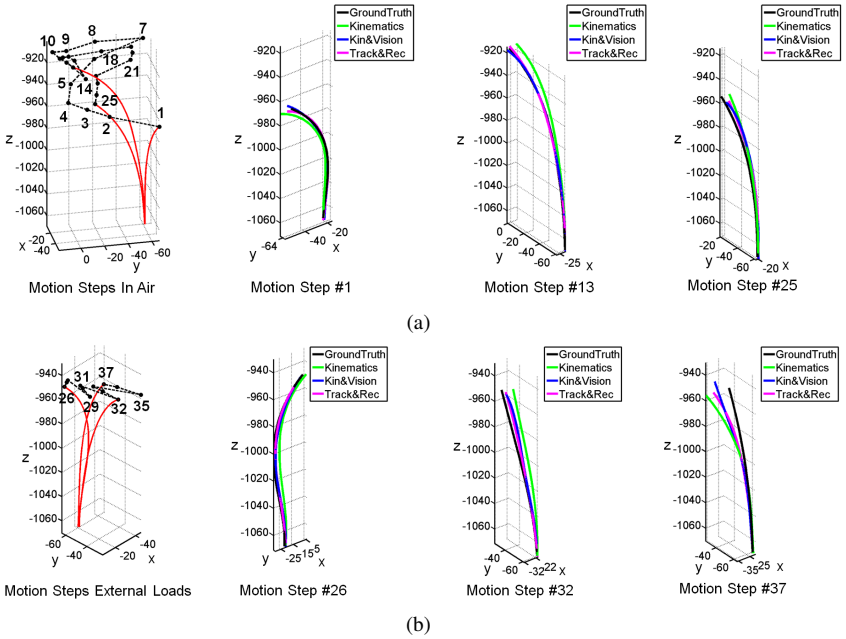


Fig. 11. Ground truth shapes for each motion steps of the experiment in free space (a), and with external loads (b), together with the motion trajectories of the robot tip (black dotted line) are shown in the first columns. The rest of the plots represent the estimated robot-shapes by kinematics, kinematics and vision combined [1] (Kin&Vision), and the proposed method with joint tracking and shape reconstruction (Track&Rec).

achieving more accurate and robust tracking of the robot in fluoroscopic images. This directly reflects on the shape reconstruction and tip position errors of these methods. The improvement is especially significant in fluoroscopic sequences characterised by the presence of anatomical and background structures. These conditions usually make tracking a challenging task to perform. When robot tracking is easy due to increased contrast, the proposed method could under-perform compared to our previous work, which does not unify tracking and kinematics. The behaviour of the algorithm proposed here, however, can be tuned with an image-contrast-based weighting term that gives increased confidence to tracking in easy scenarios.

In case of an occlusion of the robot's concentric tubes by other metallic objects, the estimation of the tube width using our proposed method might be locally affected. This could be solved by integrating constraints on the estimated width (as the tubes width is known) and/or smoothing the estimated width considering the width of its neighbour candidates. In addition, in case the robot is angled in a direction that is normal to the imaging plane, reaching an accurate estimation of the tube endpoint and robot shape can be challenging due to projective ambiguities, e.g. foreshortening. However, in practice, the surgeon would operate the robot avoiding these configurations. In fact, it is dangerous to operate the surgical robot in a configuration where its shape is difficult to

assess due to these projective ambiguities, regardless whether automatic shape estimation is used or not; usually, the effort is to remain "in plane". The same observation is valid for the vessel lumen, which represent the pathway of the endovascular instruments. Numerous publications have reported automatic estimation the optimal C-arm position for reducing tool and anatomy foreshortening and improving visualisation.

Regarding the accuracy of fluoroscope calibration used in the experimental evaluation, an average reprojection error of 0.43 ± 0.14 mm was measured. This precision can be justified by inaccuracies injected during the calibration procedure itself, and uncertainty in the C-arm motion repositioning following the acquisition of a complete "CT" image as a ground truth (the C-arm has to return to 0° RAO angle). This continuous repositioning potentially introduces extrinsic parameter errors but would not be necessary in a clinical setting. In addition, it is assumed that the coordinate systems of the robot and the C-arm can be registered by a 3DoFs displacement-based registration. This was achieved by visually aligning the two coordinate systems considering robot's landmarks in fluoroscopic images. Although this simplified registration procedure may have injected errors on the shape provided by the kinematics, the proposed vision-based approach can compensate for them as the shape estimated is refined considering image information.

It should be noted that the current implementation of the algorithm has not been optimised for high speed computation. However, many components of the algorithm can be parallelised, such as the detection of the SEGlets, the evaluation of the hypotheses, and their labels. Thus, the computational time of the algorithm would benefit from using multi-CPU- and GPU-based implementations. These solutions can be explored for real-time clinical applications. In addition, limiting the cardinality of the tracking hypotheses could further reduce running time.

VII. Conclusion

This paper proposes a new approach to vision-based intraoperative shape sensing of concentric tube robots. The method provides accurate continuous robot shape sensing using monoplane fluoroscopic images, regardless the effect of unknown forces and kinematic uncertainties. This is achieved by simultaneously addressing tracking and shape estimation of the robot, in a unified framework based on MRF. The clinical value of the work was demonstrated through both simulations and experiments. The joint framework proved to be more robust than using kinematics alone or our previous work published at IROS 2015 [1], especially in cases of difficult imaging condition where the tracking is challenging. This work can be easily adapted to biplanar X-ray fluoroscopy systems where the stereo imaging can provide additional robustness to the algorithm.

VIII. Acknowledgment

The authors would like to thank Prof. Pierre E. Dupont from Boston Children's Hospital, Harvard Medical School, for providing the concentric tubes used in the experimental study.

References

[1] A. Vandini, C. Bergeles, F.-Y. Lin, and G.-Z. Yang, "Vision-based intraoperative shape sensing of concentric tube robots," in *IEEE/RSJ Int. Conf. on Intelligent Robots and Systems*, 2015, pp. 2603–2610.

[2] J. Burgner-Kahrs, D. C. Rucker, and H. Choset, "Continuum robots for medical applications - a survey," *IEEE Trans. Robotics*, pp. 1–20, 2015.

[3] V. Vitiello, S.-L. Lee, T. P. Cundy, and G.-Z. Yang, "Emerging robotic platforms for minimally invasive surgery," *IEEE Reviews in Biomedical Engineering*, vol. 6, pp. 111–126, 2013.

[4] P. Dupont, J. Lock, and et al., "Design and control of concentric-tube robots," *IEEE Trans. Robot.*, vol. 26, no. 2, pp. 209–225, 2010.

[5] R. Webster, J. Romano, and et al., "Mechanics of precurved-tube continuum robots," *IEEE Trans. Robot.*, vol. 25, no. 1, pp. 67–78, 2009.

[6] A. Gosline, N. V. Vasilyev, E. Butler, C. Folk, A. Cohen, R. Chen, N. Lang, P. J. del Nido, and P. E. Dupont, "Percutaneous intracardiac beating-heart surgery using metal MEMS tissue approximation tools," *Int. J. Robotics Research*, vol. 31, no. 9, pp. 1081–1093, 2012.

[7] J. Burgner, D. C. Rucker, H. B. Gilbert, P. J. Swaney, P. T. Russell, K. D. Weaver, and R. J. Webster III, "A telerobotic system for transnasal surgery," *IEEE/ASME Trans. Mechatronics*, vol. 19, no. 3, pp. 996–1006, 2014.

[8] R. J. Hendrick, Richard J and Mitchell, Christopher R and Herrell, S Duke and Webster, "Hand-held transendoscopic robotic manipulators: A transurethral laser prostate surgery case study," *Int. J. Robotics Research*, vol. 34, no. 13, pp. 1559–1572, 2015.

[9] F.-Y. Lin, C. Bergeles, and G.-Z. Yang, "Biometry-based concentric tubes robot for vitreoretinal surgery," *IEEE Int. Conf. Engineering in Medicine and Biology*, pp. 5280–5284, 2015.

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

[11] J. Lock, G. Laing, M. Mahvash, and P. Dupont, "Quasistatic modeling of concentric tube robots with external loads," in *IEEE/RSJ Int. Conf. on Intelligent Robots and Systems*, 2010, pp. 2325–2332.

[12] C. Bergeles and P. Dupont, "Planning stable paths for concentric tube robots," in *IEEE/RSJ Int. Conf. Intelligent Robots and Systems*, 2013, pp. 3077–3082.

[13] L. G. Torres, C. Baykal, and R. Alterovitz, "Interactive-rate motion planning for concentric tube robots," in *IEEE Int. Conf. Robotics and Automation*, 2014, pp. 1915–1921.

[14] A. Bajo and N. Simaan, "Kinematics-based detection and localization of contacts along multisegment continuum robots," *IEEE Trans. Robot.*, vol. 28, no. 2, pp. 291–302, 2012.

[15] S. Tully, G. Kantor, M. Zenati, and et al., "Shape estimation for image-guided surgery with a highly articulated snake robot," in *IEEE/RSJ Int. Conf. on Intelligent Robots and Systems*, 2011, pp. 1353–1358.

[16] S. Song, Z. Li, H. Yu, and H. Ren, "Electromagnetic positioning for tip tracking and shape sensing of flexible robots," *IEEE Sensors Journal*, vol. 15, no. 8, pp. 4565–4575, 2015.

[17] R. Roesthuis, S. Janssen, and S. Misra, "On using an array of fiber bragg grating sensors for closed-loop control of flexible minimally invasive surgical instruments," in *IEEE/RSJ Int. Conf. Intelligent Robots and Systems*, 2013, pp. 2545–2551.

[18] S. Ryu and P. Dupont, "FBG-based shape sensing tubes for continuum robots," in *IEEE Int. Conf. Robotics and Automation*, 2014, pp. 3531–3537.

[19] R. J. Roesthuis, N. J. van de Berg, J. J. van den Dobbelsteen, and S. Misra, "Modeling and steering of a novel actuated-tip needle through a soft-tissue simulant using Fiber Bragg Grating sensors," in *IEEE Int. Conf. Robotics and Automation*, 2015, pp. 2283–2289.

[20] B. Kim, J. Ha, F. C. Park, and P. E. Dupont, "Optimizing curvature sensor placement for fast, accurate shape sensing of continuum robots," *IEEE Int. Conf. Robotics and Automation*, pp. 5374–5379, 2014.

[21] R. J. Roesthuis and S. Misra, "Steering of multisegment continuum manipulators using rigid-link modeling and fbg-based shape sensing," *IEEE Transactions on Robotics*, vol. 32, no. 2, pp. 372–382, 2016.

[22] M. W. Hannan and I. D. Walker, "Real-time shape estimation for continuum robots using vision," *Robotica*, vol. 23, no. 05, pp. 645–651, 2005.

[23] V. Chitrakaran, A. Behal, D. Dawson, and I. Walker, "Setpoint regulation of continuum robots using a fixed camera," *Robotica*, vol. 25, no. 5, pp. 581–586, 2007.

[24] D. B. Camarillo, K. E. Loewke, C. R. Carlson, and J. K. Salisbury, "Vision based 3-D shape sensing of flexible manipulators," in *IEEE Int. Conf. Robotics and Automation*, 2008, pp. 2940–2947.

[25] J. M. Croom, D. C. Rucker, J. M. Romano, and R. Webster, "Visual sensing of continuum robot shape using self-organizing maps," in *IEEE Int. Conf. Robotics and Automation*, 2010, pp. 4591–4596.
[26] R. Reilink, S. Stramigioli, and S. Misra, "3D position estimation of flexible instruments: marker-less and marker-based methods," *Int. J. Comput. Assist. Radiol. Surg.*, vol. 8, no. 3, pp. 407–417, 2013.
[27] A. Vandini, A. Salerno, C. J. Payne, and G.-Z. Yang, "Vision-based motion control of a flexible robot for surgical applications," in *IEEE Int. Conf. Robotics and Automation*, 2014, pp. 6205–6211.
[28] J. Burgner, S. D. Herrell, and R. J. Webster, "Toward fluoroscopic shape reconstruction for control of steerable medical devices," in *ASME Dynamic Systems and Control Conference*, 2011, pp. 791–794.
[29] A. Vandini, S. Giannarou, and et al., "3D robotic catheter shape reconstruction and localisation using appearance priors and adaptive C-arm positioning," in *Augmented Reality Environments for Medical Imaging and Computer-Assisted Interventions*, 2013, pp. 172–181.
[30] E. J. Lobaton, J. Fu, L. G. Torres, and R. Alterovitz, "Continuous shape estimation of continuum robots using X-ray images," in *IEEE Int. Conf. Robotics and Automation*, 2013, pp. 717–724.
[31] Y. Otake, R. J. Murphy, M. D. Kutzer, R. H. Taylor, and M. Armand, "Piecewise-rigid 2D-3D registration for pose estimation of snake-like manipulator using an intraoperative X-ray projection," in *SPIE Medical Imaging*, 2014, pp. 90 360Q–90 360Q–6.
[32] C. Bergeles, A. Gosline, N. V. Vasilyev, P. Codd, and et al., "Concentric tube robot design and optimization based on task and anatomical constraints," *IEEE Trans. Robot.*, vol. 31, no. 1, pp. 67–84, 2015.
[33] A. Vandini, B. Glocker, M. Hamady, and G.-Z. Yang, "Robust guidewire tracking under large deformations combining segment-like features (SEGlets)," *Medical Image Analysis*, 2017, (pre-print).
[34] Y. Cao and P. Wang, "An adaptive method of tracking anatomical curves in X-ray sequences," in *Int. Conf. Medical Image Computing and Computer Assisted Intervention*, 2012, pp. 173–180.
[35] N. Otsu, "A threshold-selection method from gray-level histograms," *IEEE Trans. Systems, Man, and Cybernetics*, vol. 9, no. 1, pp. 62–66, 1979.
[36] T. van Walsum, S. A. M. Baert, and W. Niessen, "Guide wire reconstruction and visualization in 3DRA using monoplane fluoroscopic imaging," *IEEE Trans. Med. Imag.*, vol. 24, no. 5, pp. 612–623, 2005.
[37] H. Heibel, B. Glocker, M. Groher, M. Pfister, and N. Navab, "Interventional tool tracking using discrete optimization," *IEEE Trans. Med. Imag.*, vol. 32, no. 3, pp. 544–555, 2013.
[38] T. W. Sederberg and S. R. Parry, "Free-form deformation of solid geometric objects," *ACM SIGGRAPH*, vol. 20, no. 4, pp. 151–160, 1986.
[39] S. Lee, G. Wolberg, K.-Y. Chwa, and S. Y. Shin, "Image metamorphosis with scattered feature constraints," *IEEE Trans. Visualization and Computer Graphics*, vol. 2, no. 4, pp. 337–354, 1996.
[40] V. Lempitsky, C. Rother, S. Roth, and A. Blake, "Fusion moves for markov random field optimization," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 32, no. 8, pp. 1392–1405, 2010.
[41] H. Ishikawa, "Transformation of general binary MRF minimization to the first-order case," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 33, no. 6, pp. 1234–1249, 2011.
[42] S. Gorges, E. Kerrien, M.-O. Berger, Y. Trousset, J. Pescatore, R. Anxionnat, and L. Picard, "Model of a vascular C-arm for 3D augmented fluoroscopy in interventional radiology," in *Int. Conf. Medical Image Computing and Computer Assisted Intervention*, 2005, pp. 214–222.
[43] J.-Y. Bouguet, "Camera calibration toolbox for matlab," 2004.

Alessandro Vandini received his Bachelor's and Masters degree in Computer Engineering from University of Modena and Reggio Emilia, Italy in 2008 and 2011, respectively. He joined the Hamlyn Centre for Robotic Surgery, Imperial College London, as a Ph.D. candidate with a fully funded scholarship in April 2012, and graduated in 2016. His Ph.D. research focused on surgical vision applied to endovascular and endoscopic procedures. This included tracking of surgical tools in intraoperative imaging modalities and vision-based shape sensing of surgical continuum robots.

Christos Bergeles (M'11) received the M.Sc. degree in Electrical and Computer Engineering from the National Technical University of Athens, Greece, in 2006, and the Ph.D. degree in Robotics from ETH Zurich, Switzerland, in 2011. He was a post doctoral research fellow at Boston Children's Hospital, Harvard Medical School, Massachusetts, and the Hamlyn Centre for Robotic Surgery, Imperial College, United Kingdom. He is now a Lecturer (Assistant Professor) at the Translational Imaging Group, Centre for Medical Image Computing, UCL, United Kingdom. Dr. Bergeles received the Fight for Sight Award in 2014, and the ERC Starting Grant in 2016. His main research area is miniaturised surgical robots guided by light-field imaging.

Ben Glocker is a Lecturer (Assistant Professor) in Medical Image Computing at Imperial College London. He holds a PhD from TU Munich, and was a post-doc at Microsoft Research and a research fellow at the University of Cambridge. Ben is the deputy head of the BioMedIA group and his research is in the area of biomedical image computing and medical computer vision with a focus on semantic understanding of images using machine learning and artificial intelligence. He received the Francois Erbsman Prize and Werner von Siemens Excellence Award in 2007, and an honorary mention for the ERCIM Cor Baayen Award 2013 which recognizes promising young researchers in the field of Informatics and Applied Mathematics. Since 2016 he is a member of the World Economic Forum's Young Scientists Community.

Petros Giataganas (M'13) received the M.Sc. degree in Electrical and Computer engineering from the University of Patras, Greece, in 2011, the M.Res. degree in Medical Robotics and Image-Guided Intervention and the Ph.D. degree in Medical Robotics from the Hamlyn Centre for Robotic Surgery, Imperial College London, London, U.K., in 2012 and 2017 respectively. He is currently a Research Assistant with the Hamlyn Centre for Robotic Surgery, Imperial College London. His main research interests include medical instrumentation, robotics, intraoperative surgical microscopy and image-guided interventions.

Guang-Zhong Yang (F'11) is Director and co-founder of the Hamlyn Centre, and Deputy Chairman of the Institute of Global Health Innovation (IGHI), Imperial College London. He holds a number of key academic positions at Imperial College: Director and Founder of the Royal Society/Wolfson Medical Image Computing Laboratory; co-founder of the Wolfson Surgical Technology Laboratory; and Chairman of the Centre for Pervasive Sensing. He is a recipient of the Royal Society Wolfson Research Merit Award and the I.I. Rabi Award from the International Society for Magnetic Resonance in Medicine. His main research interests are in medical imaging, sensing and robotics. He is a Fellow of the Royal Academy of Engineering and fellow of IEEE, IET, AIMBE, IAMBE, MICCAI, and City&Guilds. Finally, he is a recipient of the Royal Society Research Merit Award, and is listed in The Times Eureka "Top 100" in British Science.