Eye-in-Hand Visual Servoing Enhanced With Sparse Strain Measurement for Soft Continuum Robots

Xiaomei Wang, Ge Fang, Kui Wang, Xiaochen Xie, Kit-Hang Lee, Justin D. L. Ho, Wai Lun Tang, James Lam, *Fellow, IEEE*, and Ka-Wai Kwok, *Senior Member, IEEE*

***Abstract*—In the feature/object tracking of eye-in-hand visual servoing, 2D motion estimation relying only on image plane feedback is easily affected by vision occlusion, blurring, or poor lighting. For the commonly-used template matching method, tracking performance greatly depends on the image quality. Fiber Bragg gratings (FBGs), a type of high-frequency flexible strain sensor, can be used as an assistant device for soft robot control. We propose a method to enhance motion estimation in soft robotic visual servoing by fusing the results from template matching and FBG wavelength shifts to achieve more accurate tracking in applications such as minimally invasive surgery. Path following performance is validated in a simulated laparoscopic scene and LEGO-constructed scene, demonstrating significant improvement to feature tracking and robot motion, even under external forces.**

***Index Terms*—Eye-in-hand visual-servo, FBG, motion estimation, soft robot control.**

I. Introduction

VISION plays an important role in robotic sensing, as it provides a non-invasive and intuitive measurement modality [1]. Visual servoing is a control strategy that utilizes visual feedback to close the control loop of robotic systems. It is able to resist distance-related limitations in robotic end-effector control, with applications extending to the manufacturing industry, military field, automobile steering and even aircraft landing [2]. Visual servoing can be classified into eye-in-hand and eye-to-hand configurations based on the camera's location. Eye-in-hand cameras are embedded at the robot end-effector and enable more flexible viewing of the workspace. Using this approach for rigid-link robots, tele-operated object/feature tracking and obstacle avoidance in imperfectly-modeled environments can be achieved [1], [2]. Applications also include robotic minimally invasive surgery (MIS), where surgical instruments and specific tissues or organs can be tracked intra-operatively [3], reducing the burden of manual endoscope control.

With the rising prevalence of soft robotics, researchers also sought to implement visual servoing to enhance their feedback control. Due to the limited selection of sensors that are compatible with their highly deformable structures, soft robots [4], [5] can take advantage of compact and self-contained camera feedback. Wang *et al*. [6] first achieved eye-in-hand visual servo control on cable-driven soft robots, and extended their work to handle static physical constraints [7]. However, their work focused on analytical kinematic modelling without considering visual sensing accuracy. Additionally, they only applied the visual servoing to a simplified non-surgical scene containing distinct features. Similar scenarios were also used in a study that performed visual servoing with concentric-tube robots [8]. Greer *et al*. [9] applied visual servoing on series pneumatic artificial muscles, using a commercial image processing system (Sightline Applications Inc.) for feature identification. However, they did not consider the effect of poor feature qualities in camera view or external disturbances.

Our previous work [10] enabled accurate path following in the camera view with LEGO-constructed scenes. We used a typical template matching algorithm [11] which enables visual feature recognition for motion estimation. However, camera feedback is sometimes unreliable since poor illumination levels or sudden and unpredicted motions may deteriorate the image quality [12]. This will inevitably reduce the motion estimation accuracy. As soft robots are inherently compliant and adaptive in scenarios with high safety requirements (e.g. endoscopy and laparoscopy [13], [14]), enhancing feature tracking and motion estimation is essential for maintaining precise control, and therefore, efficiency during interventional operations.

Researchers have proposed several ideas to use additional sensors to assist vision-based localization or motion estimation, among which inertial measurement units (IMUs) are a prevalent choice. Inertial sensors offer a high tracking rate without the requirement of external landmark references [12]. Since they can only keep track of orientation at discrete nodes, the continuous morphology of the robot must be predicted by extrapolation. IMUs are embedded with magnetometers to determine 3D orientation and compensate gyro drift. However, the accuracy of magnetometers is sensitive to ferromagnetic materials, such as surgical instruments. Special design considerations should be taken into account because an IMU would be placed at the robot tip and packed with the camera and LED. In interventional

Manuscript received September 10, 2019; accepted January 9, 2020. Date of publication January 28, 2020; date of current version February 17, 2020. This letter was recommended for publication by Associate Editor A. A. Stokes and Editor K.-J. Cho upon evaluation of the reviewers' comments. This work was supported in part by the Research Grants Council of Hong Kong under Grants 17202317, 17227616, and 17206818 and in part by the Innovation and Technology Commission under Grant UIM/353. *(Corresponding author: Ka-Wai Kwok.)*

The authors are with the Department of Mechanical Engineering, The University of Hong Kong, Hong Kong 999077, China (e-mail: wangxmei@connect.hku.hk; fangge@hku.hk; kuiwang@connect.hku.hk; xcxie@connect.hku.hk; briankhl@hku.hk; jdlho@connect.hku.hk; at361836@hku.hk; james.lam@hku.hk; kwokkw@hku.hk).

This letter has supplementary downloadable material available at https://ieeexplore.ieee.org, provided by the authors.

Digital Object Identifier 10.1109/LRA.2020.2969953

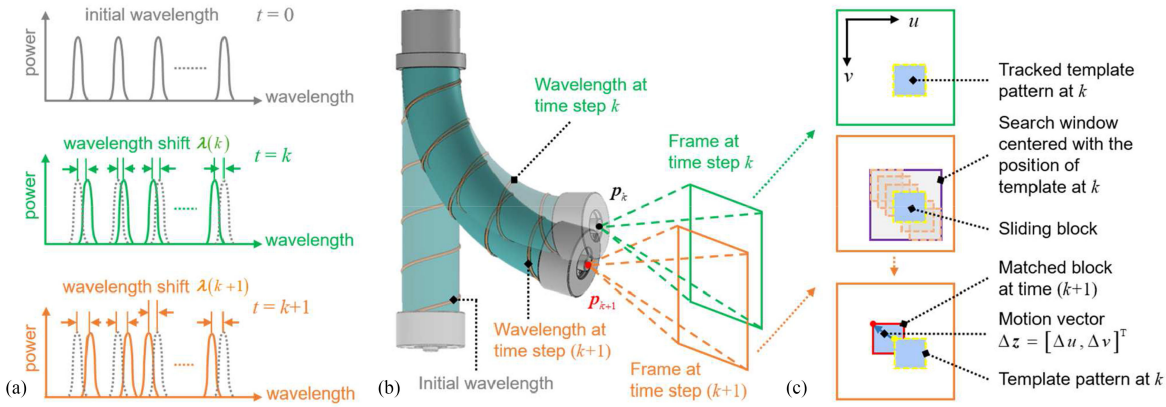

Fig. 1. (a) Example FBG wavelength shift changes from $\lambda(k)$ to $\lambda\ (k+1)$, (b) due to the corresponding robot deformation/bending, the difference of camera coordinate frames at k and $k+1$ can be seen; (c) Incremental motion in the image plane acquired based on the displacement from the template pattern (yellow dotted block) to the matched block (red) that is found by block sliding in the search window (purple block).

operations, additional water-proofing for the IMUs will be required. Other sensors such as electromagnetic (EM) trackers are restricted to the workspace of the EM field generator and may suffer from EM interference.

Optical fibers with fiber Bragg gratings (FBGs) have intrinsic advantages in strain-related state detection and are one of few other sensing modalities particularly suited for soft robots. FBGs can measure strain along optical fiber at high frequencies (>100 Hz) based on the change of reflected wavelengths. The fiber can be combined with soft elastic substrates to form a standalone sensor [15] or attached directly to soft manipulators [16]. FBGs have been investigated for shape and tip sensing of continuum manipulators, even in surgical tools like biopsy needles [17], [18]. Different from sensors like IMUs, optical fibers could reflect morphology information and are unaffected by many environmental effects such as EM fields and water submersion due to its optical nature. Both model-based and learning-based position/shape estimation approaches were investigated in recent work [19], [20]. As a monocular camera already enables 2D motion estimation with slow robot motion and abundant features, it could be employed as the ground truth to train FBGs as a motion estimator, therefore enhancing the estimation results obtained by the camera in scenarios with poor vision.

In this letter, we propose a 2D motion sensing modality to enhance the visual servo performance under feature-deficient and force-disturbed conditions. This work is the first to consider and solve the feature "drifting" problem in soft robot visual servoing. We equip a soft continuum robot with a monocular camera and a helically-wrapped FBG fiber along its body, thus fusing both visual and strain sensing data. The learning-based method is inherently advantageous in soft robot sensing using optical fibers [21], circumventing modeling uncertainties in soft materials. The major contributions are:

i) Development of an enhanced eye-in-hand tracking algorithm by incorporating camera-based 2D motion estimation with sparse strain data collected by optical fiber to improve the real-time tracking performance in dim/feature-lacking environments;
ii) Design of a prototype endoscopic soft robot which is integrated with a helically wrapped FBG fiber. The learning-based motion estimation approach alleviates the need for precise FBG placement on the soft robot;
iii) Experimental validation of the enhanced motion estimation method by path tracking in laparoscopic and LEGO-constructed scenes even under disturbance.

II. Methods

The monocular camera enables 2D motion estimation in the image plane between two adjacent frames. The FBG fiber wrapped on the robot could also reflect the robot configuration and be trained to estimate the motion (Fig. 1). By fusing the estimations from camera and strain measurement, a controller can enable precise end-effector motion in the image plane. The overall control architecture is illustrated in Fig. 2.

A. Task Space Definition

The endoscopic camera mounted at the robot end-effector allows eye-in-hand visual servoing. The actuator input at time step k (at equilibrium state) is represented as $\boldsymbol{q}(k) \in U^m$, where m denotes the dimension of actuation space. The robot deformation/bending induces corresponding differences of camera coordinate frames at time step k and $(k+1)$ (Fig. 1b). The task space is defined in the 2D camera frame (Fig. 1c), with the incremental 2D displacement denoted as $\Delta\mathbf{z}(k) \in \mathbb{R}^2$. The frame is always perpendicular to the robot tip normal. The control objective is to generate an actuation command $\Delta\boldsymbol{q}(k)$, achieving the desired movement $\Delta\mathbf{z}^*(k)$.

B. Motion Estimation on Image Plane

To estimate the 2D motion between two successive frames, a target block of image intensity features is defined as reference for comparison (Fig. 1c). We assume that the orientation change of the robot is small within a short time interval (20∼30 Hz). The change of camera orientation attributed to rotation was found to be <5° between successive frames, which corresponds to a movement error of <3 pixels for template matching calculation [10]. In the cases of continuous robot movements, the camera frames would follow a more-or-less planar motion.

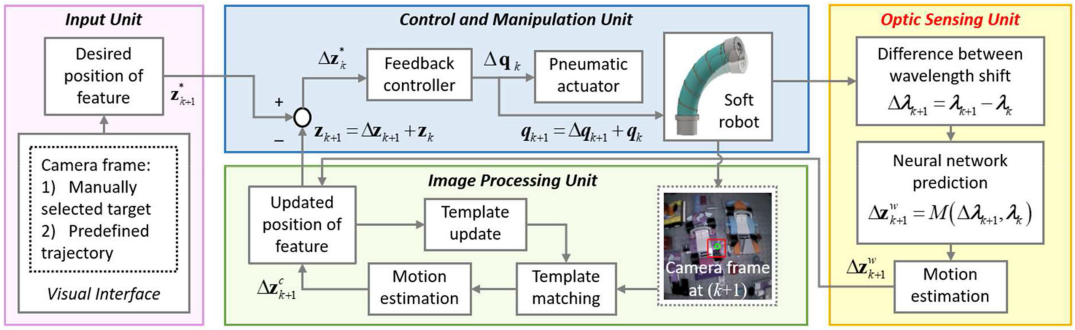

Fig. 2. Control architecture of the proposed control method. Parameters $\boldsymbol{q}$ and $\boldsymbol{z}$, respectively, denote the actuation command and the position of the captured feature in camera view. The input unit provides the positional command in the image view. A model-free feedback controller is responsible for generating the inverse solution to the actuator, causing the object movement in the camera view as well as the change of FBG wavelength. The actual displacement after the execution of actuation command is obtained by the weighted sum of FBG-based motion estimation and image processing.

The translational displacement in this image plane can be estimated using block template matching method (Fig. 1c), named "*matchTemplate*" in OpenCV [11]. Details of template matching algorithm with equation explanations can be found in our previous work [10].

Although template matching has been widely used to estimate relative motions, the light and feature conditions would affect the estimation results to a great extent. The improvement on image-based algorithms could not overcome these intrinsic weaknesses. With the strain feedback of FBGs that reflects the change of robot configuration, a model combining the strain data and camera motion could be trained to assist the real-time camera-based motion estimation.

C. Learning-Based Motion Estimation Combined With FBG

1) Data-Driven Motion Estimation With FBG: FBGs in the optical fiber should reflect the robot configuration during robot motion. The bases for fiber placement are: 1) The FBGs should be distributed along the places with large strain changes; 2) The strain changes or wavelength shifts of FBGs should uniquely map with the robot configuration; 3) The addition of fiber should not affect the mechanical properties (e.g. bending stiffness and elasticity) of soft robot; 4) Surface contact on the cylindrical robot body should not dominate the FBG feedback. Based on these considerations, we wound the fiber helically on the robot and sealed it using silicone adhesive. The wrapped optical fiber with L FBGs provides L wavelength shifts which can be represented by $\boldsymbol{\lambda} = [\lambda_1\ \lambda_2 \cdots \lambda_L]^T \in \mathbb{R}^L$. Wavelength shift $\boldsymbol{\lambda}(k)$ at time step k is the difference between current wavelength vector and the original wavelength vector $\boldsymbol{\lambda}_0$ (corresponds to initial robot configuration) (Fig. 1a). For each two adjacent time steps k and $(k+1)$, we could obtain the 2D motion vectors in the camera frame $\Delta\mathbf{z}(k+1) \in \mathbb{R}^2$ (Section II-A, B), and the wavelength shift $\boldsymbol{\lambda}(k)$ and $\boldsymbol{\lambda}(k+1)$. We represent their wavelength difference as $\Delta\boldsymbol{\lambda}(k+1) = \boldsymbol{\lambda}(k+1) - \boldsymbol{\lambda}(k)$ (Fig. 1a). In the camera view with abundant features, the estimation results from template matching during slow and smooth manipulation could be regarded as a ground truth, which is used to train the FBG-related estimation model.

Training: The pre-training procedure would collect the captured endoscopic image at each time step, as well as the wavelength shift vector $\boldsymbol{\lambda}$. The actuation sequence $\boldsymbol{Q} = [\boldsymbol{q}(1)\ \boldsymbol{q}(2) \cdots\ \boldsymbol{q}(N)]$ (N is the sampling number) to the robot for data exploration is predefined to cover the whole workspace. The wavelength difference sequence could therefore be obtained as

$$\Delta\boldsymbol{\Lambda} = \left[\Delta\boldsymbol{\lambda}(1)\ \Delta\boldsymbol{\lambda}(2) \cdots\ \Delta\boldsymbol{\lambda}(N)\right]$$

where $\Delta\boldsymbol{\lambda}(i) = \boldsymbol{\lambda}(i) - \boldsymbol{\lambda}(i-1),\ i = 1, 2, \ldots, N$. The images captured on each step are saved to calculate the 2D motion $\Delta\boldsymbol{z}(i) = \boldsymbol{z}(i) - \boldsymbol{z}(i-1)$ offline, forming the motion sequence

$$\Delta\boldsymbol{Z} = \left[\Delta\boldsymbol{z}(1)\ \Delta\boldsymbol{z}(2) \cdots\ \Delta\boldsymbol{z}(N)\right].$$

Using the feedforward neural network in the deep learning toolbox of MATLAB, with $[\Delta\boldsymbol{\Lambda}\ \boldsymbol{\Lambda}]^T$ as input and $\Delta\boldsymbol{Z}$ as output, we can train a motion estimation model represented as

$$\Delta\boldsymbol{z}(i) = M\left(\Delta\boldsymbol{\lambda}(i), \boldsymbol{\lambda}(i)\right),\ i = 1, 2, \ldots, N. \tag{1}$$

Prediction: To increase the processing speed, we convert the model generated in MATLAB to a dynamic link library (DLL) that could be called in the C++ environment. The wavelength receiving via User Datagram Protocol (UDP) and the motion estimation are both accomplished in Qt Creator, with the motion estimation at the k^{th} step obtained by

$$\begin{aligned}\Delta\boldsymbol{z}_w(k) &= M\left(\Delta\boldsymbol{\lambda}(k), \boldsymbol{\lambda}(k)\right)\\ &= M\left(\left(\boldsymbol{\lambda}(k) - \boldsymbol{\lambda}(k-1)\right), \boldsymbol{\lambda}(k)\right),\ k = 1, 2, \ldots\end{aligned} \tag{2}$$

As this model is not related to real-time image processing once finished the training procedure, it could be regarded as an independent 2D motion estimator using FBG with a frequency of at least 30 Hz.

2) Camera-FBG-Combined Motion Estimation: We use the template matching algorithm for camera-based motion estimation. For each update of the sliding block in time step k (Fig. 1c), there is a coherence variable β characterizing the similarity between the sliding block and the template pattern in the last step $(k-1)$. The new template position $\boldsymbol{z}(k)$ is finalized by finding the sliding block that possesses the maximum coherence with the template at $\boldsymbol{z}(k-1)$. We define the 2D motion estimation from step $(k-1)$ to k by image processing as $\Delta\boldsymbol{z}_c(k)$, with a coherence $\beta(k)$. Another motion estimation obtained from

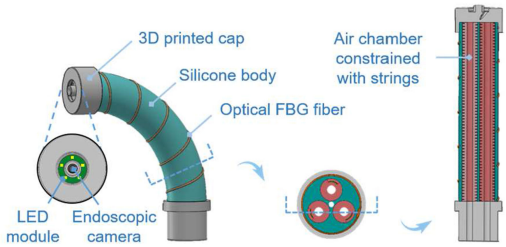

Fig. 3. Pneumatic-driven 3-chamber robot is used. An optical fiber with multiplexing 16 FBGs is helically wrapped around the manipulator for sensing feedback of robot configurations in real time. A monocular endoscopic camera and a LED module are fixed on the tip cap of robot. Cross-section and axial views of the robot show the silicone chambers constrained individually with helical Kevlar strings.

the FBG-related model in (2) is $\Delta \boldsymbol{z}_w(k)$. For the camera-based estimation, the quality of features limits the accuracy. On the other hand, FBG is highly sensitive to temperature and tiny strains, sometimes may involve noise. To avoid their defects for more reliable estimations, we designed a weighted estimator, represented as

$$\Delta \boldsymbol{z}_p(k) = \alpha\beta(k) \cdot \Delta \boldsymbol{z}_c(k) + (1 - \alpha\beta(k)) \cdot \Delta \boldsymbol{z}_w(k) \quad (3)$$

where α is a constant parameter to adjust the ratio of coherence in the camera-estimated portion. It is a scaling factor to normalize the fusion results of two motion estimations. This parameter is tuned by a heuristic procedure before robot execution. In our experiment, we initialized it as 0.7. Its value does not bring notable differences to the tracking accuracy within a small range of variation (0.6∼0.8).

D. Visual Servo Based on 2D Motion Estimation

With the above enhanced estimation as feedback, we implement a model-free controller that is inspired by the optimal control method in [22]. The forward kinematic mapping from actuation space to task space is represented as

$$\dot{\boldsymbol{z}}_p = \boldsymbol{J}\dot{\boldsymbol{q}} \quad (4)$$

where $\boldsymbol{z}_p$ and $\boldsymbol{q}$ are the absolute 2D position of template to be tracked and the actuator input respectively, $\boldsymbol{J}$ is the Jacobian matrix. The inverse kinematics can be discretized to

$$\Delta \boldsymbol{q} = \boldsymbol{J}^{-1} \Delta \boldsymbol{z}_p \quad (5)$$

The core of obtaining $\Delta \boldsymbol{q}$ in (5) for accurate motion is to find a proper Jacobian matrix $\boldsymbol{J}$. For our robot, the 3 degrees of freedom (DoFs) of the actuator are in the same unit and independent of each other. The initialization of $\boldsymbol{J}$ could be achieved by actuating the 3 DoFs in turn with an incremental amount $\Delta q_i, i = 1, 2, 3$ (i.e. the actuation commands are $[\Delta q_1\ 0\ 0]$, $[0\ \Delta q_2\ 0]$ and $[0\ 0\ \Delta q_3]$ successively), and measuring the corresponding displacements $\Delta \boldsymbol{z}_{pi}$. The initial Jacobian matrix could be constructed as

$$\boldsymbol{J} = \left[\boldsymbol{J}_1\ \boldsymbol{J}_2\ \boldsymbol{J}_3\right] \quad (6)$$

where $\boldsymbol{J}_i = \Delta \boldsymbol{z}_{pi} / \Delta q_i$. The obtained Jacobian matrix could also be updated online, relying on continuously solving a quadratic programming problem as follows

$$\begin{aligned} \text{minimize} \quad & \|\Delta \boldsymbol{J}(k+1)\| \\ \text{subject to} \quad & \Delta \boldsymbol{z}_p(k) = \boldsymbol{J}(k+1)\,\Delta \boldsymbol{q}(k) \\ & \boldsymbol{J}(k+1) = \boldsymbol{J}(k) + \Delta \boldsymbol{J}(k+1) \end{aligned} \quad (7)$$

where $\Delta \boldsymbol{J}(k+1)$ is the variable to be optimized, $\boldsymbol{J}(k)$ is the Jacobian matrix at time k, $\boldsymbol{J}(k+1)$ is the estimation for Jacobian at time $(k+1)$. After obtaining the latest Jacobian, the command to actuators could be calculated by

$$\Delta \boldsymbol{q}(k+1) = \boldsymbol{J}(k+1)^{-1} \Delta \boldsymbol{z}_p^*(k+1) \quad (8)$$

where $\Delta \boldsymbol{z}_p^*(k+1)$ is the desired motion in camera view.

As illustrated in Fig. 2, the positional command is obtained by the difference of the desired and current positions of the feature (Section II-A). The robot is actuated accordingly with the inverse solution generated from the model-free feedback controller (Section II-D). The change of robot configuration causes the object movement in the camera view as well as the change of FBG wavelength. The actual 2D motion will be measured by fusing the estimates of image processing (Section II-B) and optical sensing (Section II-C).

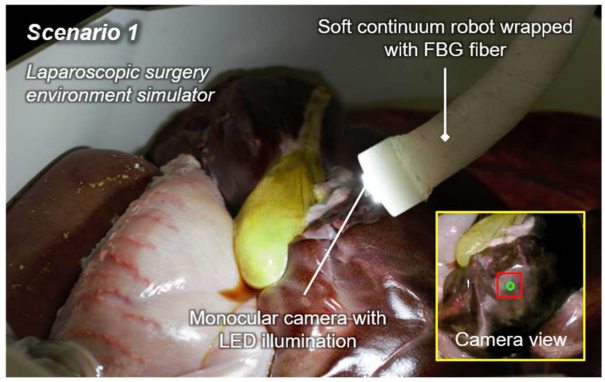

Fig. 4. Endoscopic scene in abdominal surgery simulated with swine viscera. The robot is pre-bended to approximate insertion angle of a laparoscope. EM tracking coils are attached on the tip as the ground truth of end-effector pose.

III. Experiments, Results & Discussion

A. Soft Endoscopic Robot With FBG Sensor Device

The soft robot used in this study is constructed from a moulded silicone rubber (Ecoflex50, Smooth-on Inc.) segment, a monocular endoscopic camera, FBG fiber and 3D printed fixation components (Fig. 3). The continuum robot comprises of three cylindrical fluidic chambers spaced 120° apart from each other. Each chamber is constrained radially by a helically wrapped Kevlar strings in order to only allow expansions on axial direction. Different combinations of three inflation pressures capacitate the omnidirectional bending of the robot (∼100°). An endoscopic camera (OV6930, Depth of view: 8 to 150 mm, Shanghai E-vision Optoelectronic Co., Ltd.) and a module of LED illumination are fixed on the tip cap. A central cavity inside the robot body is reserved to house the cables of camera and LEDs. With a 90° diagonal field of view, the camera captures images of 400 × 400 pixels, indicating that a pixel translates

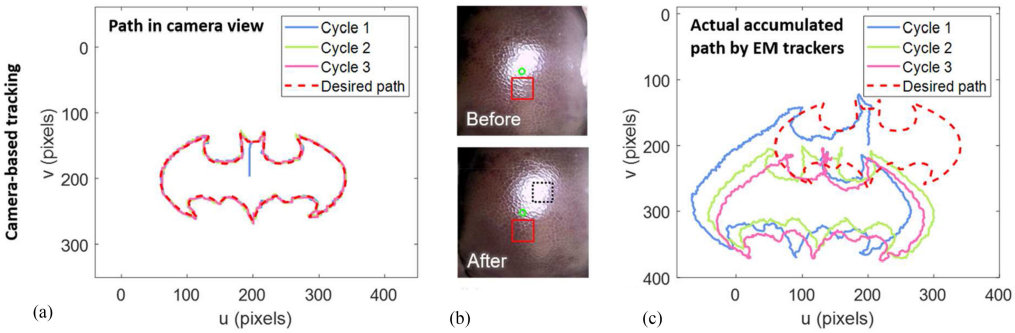

Fig. 5. Robot following of a predefined "Batman" path in the scene of Fig. 4. (a) The motion estimated by image processing of the endoscopic view *alone*; (b) Bright reflective spot displaced along the tissue due to the robot motion. Features in the red block were selected by the user *before* the motion displacement. *After*, such a block was expected to keep matching/tracking at the same square of features (black dotted block indicates the position of selected red block *Before*), acting as a static reference for robot to "draw" the path; (c) Actual path of end-effector recorded by EM tracking coils, which project on the same *u-v* coordinates. The recorded deviation along those 3 cycles was caused by the error of such a red block matching/tracking.

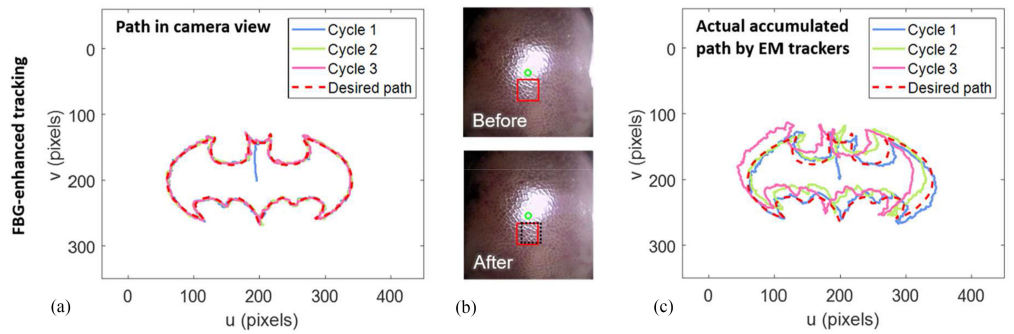

Fig. 6. Tracking performance of the same "Batman" path in the scene of Fig. 4, with the motion estimated by FBG-enhanced method. (a) The motion estimated by the FBG-enhanced algorithm in the endoscopic view; (b) The deviation between the selected feature after 3 cycles (difference between the red block and black dotted block) is obviously reduced, resulting in more accurate tracking in (c) recorded by EM trackers.

to 0.16° field of view. The 3D printed (Stereolithographic) cap could be worn on the distal (relative to the base of robot) part of robot. A single-core optical fiber with 16 FBGs (8 mm length for each FBG) is helically wrapped and adhered (by Sil-Poxy, Smooth-on Inc.) along the silicone body. The bottom of the robot is fixed to a 3D printed base that remains stationary. The outer diameter of the robot is 19 mm and the total length is 95 mm (including the tip cap). With the optical fiber ($421.5) and the camera ($780), the price of soft manipulator is ∼$1,210. The non-consumable FBG interrogator costs ∼$20,000, while the actuation components cost ∼$400.

B. Experimental Setup

Two different scenarios are constructed, as illustrated in Fig. 4 and Fig. 8 respectively. The soft manipulator described in Section III-A is fixed downward, viewing the workspace scene built from swine viscera to simulate the endoscopic scene in laparoscopy (Fig. 4), or LEGO (Fig. 8). The distances between the endoscopic camera and the scene surface are 3 cm and 15 cm, respectively.

The precision of visual servo control depends on two main aspects, which are the camera-based motion estimation and the robot control strategy. The tracking errors could also be considered from two aspects. The first kind ignores the drift of tracked template among different image frames, and the motion estimation is supposed to be reliable. It only reflects the controller performance (i.e. Figs. 5–6a, 9–10a). Another kind includes the accuracy of motion estimation (i.e. Figs. 5–6c, 9–10c) and should be evaluated by a more stable ground truth. As the main purpose of this letter is to improve the motion estimation accuracy, the validation tasks are designed to compare the performances of camera-based motion estimation in [10] with the new FBG-enhanced method, using the same robot controller. During the tests, two 6D electromagnetic (EM) trackers (#610059, NDI Aurora) are laterally attached on the robot tip to record the pose of the end-effector, which is used as the evaluation ground truth. After each actuation step, the 2D motion vectors in the camera view could be calculated by projecting the 3D motion vectors on the current image plane, which is always perpendicular to the end-effector normal direction. The actual path in the camera view could be reconstructed by accumulating these 2D motion vectors.

To obtain the accumulated actual path, we define the 3D positions and 4D quaternions of two EM coils at time step k as $\boldsymbol{p}_i(k)$, and $\boldsymbol{h}_i(k)$, $i = 1, 2$. The position of end-effector is

$$\boldsymbol{p}\left(k\right) = \left(\boldsymbol{p}_1\left(k\right) + \boldsymbol{p}_2\left(k\right)\right)/2 \tag{9}$$

the rotation matrix $\boldsymbol{R}(k)$ can be calculated from $\boldsymbol{h}_1(k)$ or $\boldsymbol{h}_2(k)$. Suppose the orthogonal basis for original local end-effector coordinate is represented as $\boldsymbol{u}_0 = [1\ 0\ 0]^T$, $\boldsymbol{v}_0 = [0\ 1\ 0]^T$ and $\boldsymbol{n}_0 = [0\ 0\ 1]^T$, then the basis vectors at time step k will be

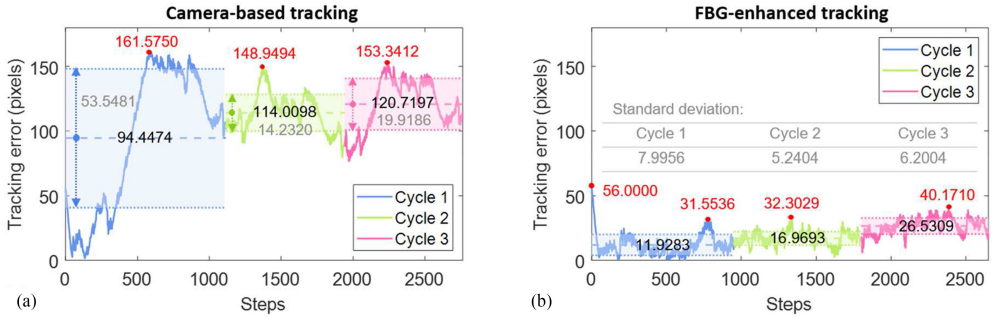

Fig. 7. Tracking error analysis of the "Batman" path in the laparoscopic scene (Fig. 4). (a) Tracking achieved by image processing, e.g. mean error: 94.45 (SD: 53.55) induced in Cycle 1; (b) Tracking achieved by FBG-enhanced estimation. Curves represent the absolute tracking error in Fig. 5c, 6c and red points illustrate the maximum error in each cycle.

$\boldsymbol{u}(k) = \boldsymbol{R}(k) \cdot \boldsymbol{u}_0$, $\boldsymbol{v}(k) = \boldsymbol{R}(k) \cdot \boldsymbol{v}_0$ and $\boldsymbol{n}(k) = \boldsymbol{R}(k) \cdot \boldsymbol{n}_0$. With the 3D motion vector and the current image plane obtained by $\Delta \boldsymbol{p}(k) = \boldsymbol{p}(k) - \boldsymbol{p}(k-1)$ and $\boldsymbol{U}(k)\boldsymbol{V}(k)$ respectively, the projected motion vector on the image plane is

$$\Delta \hat{\boldsymbol{p}}\,(k) = \Delta \boldsymbol{p}\,(k) - \frac{\Delta \boldsymbol{p}^T\,(k) \cdot \boldsymbol{n}\,(k)}{|\boldsymbol{n}\,(k)|^2} \cdot \boldsymbol{n}\,(k) \tag{10}$$

w.r.t. the generator coordinate. The 2D motion vector w.r.t. image plane $\boldsymbol{U}(k)\boldsymbol{V}(k)$ is calculated by

$$\Delta \boldsymbol{z}\,(k) = \left[\frac{\Delta \hat{\boldsymbol{p}}^T\,(k) \cdot \boldsymbol{u}\,(k)}{|\boldsymbol{u}\,(k)|} \;\; \frac{\Delta \hat{\boldsymbol{p}}^T\,(k) \cdot \boldsymbol{v}\,(k)}{|\boldsymbol{v}\,(k)|} \right]^T \tag{11}$$

There would be a constant bias angle θ between coil-obtained $\Delta \boldsymbol{z}(k)$ and camera-based motion estimation $\Delta \boldsymbol{z}_c(k)$, resulting from the mounting angle of EM trackers. According to this calculation method, θ could be obtained by calculating the mean angle difference using a set of sampling data, similarly the multiple that reflects the displacement length could also be calculated. After compensating this bias θ by

$$\Delta \boldsymbol{z}_s\,(k) = \boldsymbol{T} \cdot \Delta \boldsymbol{z}\,(k) = \begin{bmatrix} \cos(\theta) & \sin(\theta) \\ -\sin(\theta) & \cos(\theta) \end{bmatrix} \cdot \Delta \boldsymbol{z}\,(k) \tag{12}$$

the standard motion vector $\Delta \boldsymbol{z}_s(k)$ is obtained to test the accuracy of image-processing-based estimation and FBG- enhanced estimation.

C. Visual Servo in Feature-Deficient Scenes

The improvement of motion estimation is investigated via path following tasks. In our previous letter [10], we have tested a desired path in the "∞" shape. The smooth path is defined by 72 points on the 400×400 px camera frame (with the mean interval of 15.6828 pixels), generating a closed curve with continuous gradient. To increase the tracking difficulty, the path designed in this letter is the outline of Batman logo with several acute angles, consisting of 127 points with the mean interval of 7.9109 pixels. The robot is instructed to track a same template pattern in the laparoscopic scene (Fig. 4) along the Batman path for 3 cycles, using template matching alone (Fig. 5) and the enhanced method proposed in this letter (Fig. 6), respectively.

In the camera view, the tracking trajectory of the target block roughly accords with the Batman path when the 2D motion estimation are achieved by both template matching and

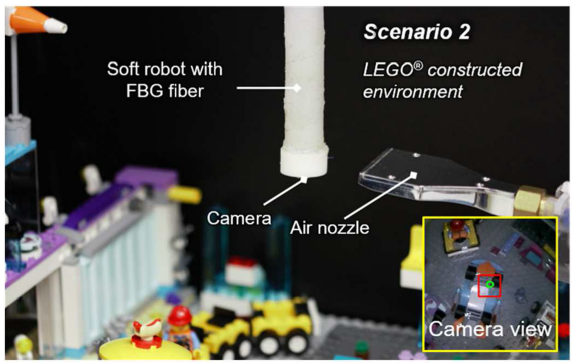

Fig. 8. Setup in a scene of LEGO. Robot is fixed downward viewing the workspace scene. An air nozzle is fixed facing the robot to generate force disturbances. EM tracking coils are attached as the ground truth.

FBG-enhanced method (Fig. 5a, RMSE: 8.5991 pixels, Fig. 6a, RMSE: 8.6391 pixels), benefitting from the well-performed closed-loop controller. However, the accuracy of motion estimation could be reflected in the actual projected path of end-effector (Fig. 5c, 6c), which should also be identical with Fig. 5a, 6a. The 2D estimation error at each step is accumulated and deviates the EM tracker-reconstructed actual path from the desired path. The target template drifts and gradually misses the original target feature (Fig. 5b, 6b). In this task, the to-be-tracked template block is intentionally defined on the feature-deficient liver surface. Fig. 5b includes two camera views captured at the beginning of 1st cycle and the end of 3rd cycle, so does Fig. 6b. The drift of target template could be observed referring to the feature marked by black dotted block. The absolute tracking errors between the actual and desired trajectories are demonstrated in Fig. 7. The root-mean-square error (RMSE) for each cycle gradually increases after three-cycle execution, which are 108.5594, 114.8936 and 122.3500 pixels when using camera-based tracking, and 14.3578, 17.7591 and 27.2450 pixels when using FBG-enhanced tracking. After the enhancement, the overall tracking error is reduced by 82.3%.

It could be seen that the tracking performance in visual servoing would be deteriorated under two conditions, one is in feature-deficient scenes, such as the liver surface in this task; another is the step length of desired motion. As template matching is purely based on the features captured in adjacent camera frames, the estimation would be inaccurate if there are

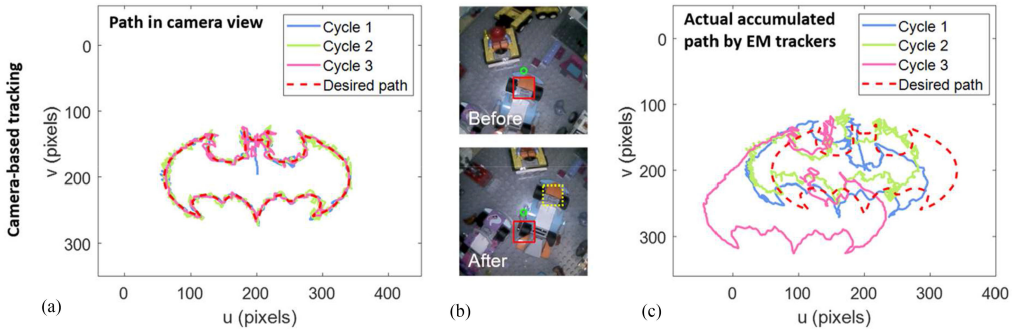

Fig. 9. Performance of robot following on the predefined "Batman" path in the scene of Fig. 8. Force disturbances are applied to the end-effector from the nozzle, with 4 bar pressure at the beginning of the 2nd cycle and 6 bar pressure in the middle of the 2nd cycle, released at the beginning of the 3rd cycle. (a) The motion estimated by image processing of the endoscopic view *alone* vibrates during and after the disturbances applied; (b) Features in red block was selected by user *before* the motion displacement. *After* then, yellow dotted block indicates the position of the selected red block *before*; (c) Actual path of end-effector deviates gradually after 3 cycles.

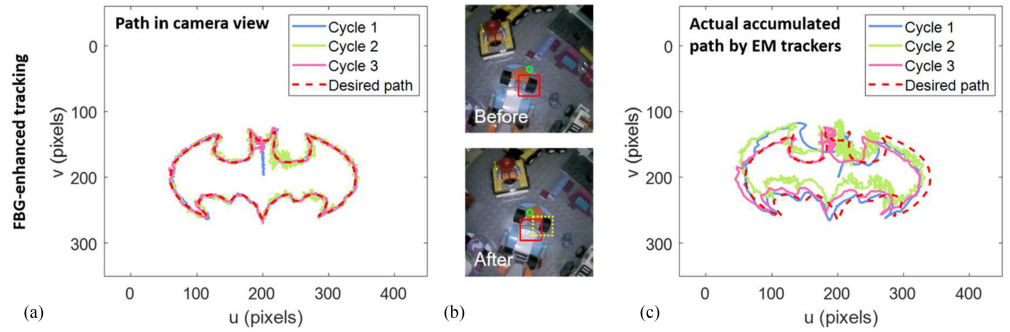

Fig. 10. Performance of robot following on the "Batman" path in the scene of Fig. 8 with the motion estimation by enhanced method. (a) Robot motion in the endoscopic view maintains stable under and after external forces; (b) Feature deviation is well solved; the actual path in (c) demonstrates the improvement.

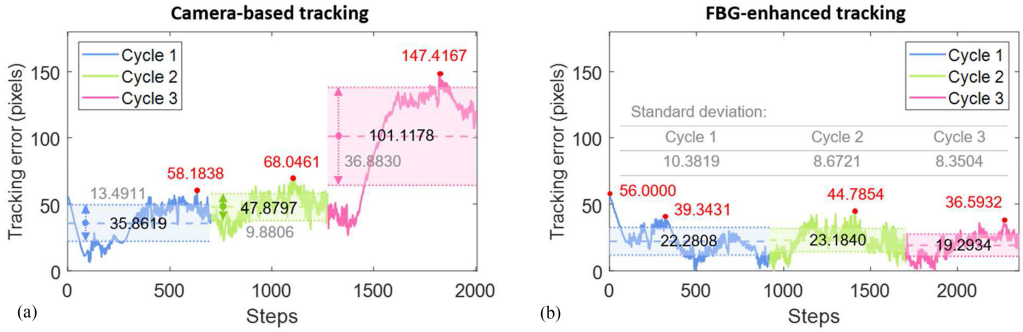

Fig. 11. Tracking error analysis in the LEGO scene (Fig. 8). (a) Tracking achieved by image processing; (b) Tracking achieved by FBG-enhanced estimation. Curves represent the absolute error in Fig. 9c, 10c.

few features to be tracked or the features are repeatable in the camera frame. When the tracking path is defined by dense points, the motion step size has to be small, aggravating the proportion of error in each step.

D. Visual Servo Under External Disturbances

Besides the laparoscopic-simulation task, we also set up an environment using LEGO (Fig. 8) to test another challenge in visual servo for soft robots, external forces. An air nozzle is fixed facing the tip of soft robot in a distance of 5 cm and is connected to an air compressor. The robot is actuated to follow the same Batman path in Section III-C for 3 cycles. A gust of "wind" with 4 bar air pressure is exerted on the robot at the beginning of the 2nd cycle and increased to 6 bar in 0.5 s when half of the 2nd cycle is finished (taking ∼38 s). The external pressure is released at the beginning of the 3rd cycle. The 6-bar pressure is almost the maximum at which the robot could sustain stable tracking. Although the air compressor outputs constant air pressure, the force direction and power applied to the robot varies unpredictably due to the robot motion. The tracking performances using two kinds of motion estimation methods are exhibited in Fig. 9 and Fig. 10.

In the camera view, the actual trajectory when using enhanced motion estimation method is always consistent with the Batman path, although the path becomes unsmooth when the external force is applied (Fig. 10a, RMSE: 9.1586 pixels. max. error: 31.7108 pixels in Cycle 2). However, by the original

camera-based tracking, the path could not return stable quickly when the air pressure is released (Fig. 9a), the max. tracking error in Cycle 3 (32.3634 pixels) is even larger than that of Cycle 2 (26.6751 pixels). One of the reasons is that the step size used in this method is larger than the one in the enhanced estimation method (Fig. 10), to reduce the portion of error in each motion step. If the step size is as small as in the enhanced method, the actual following path (Fig. 9c, RMSE: 107.6258 pixels, max. error: 147.4167 pixels in Cycle 3) will deviate the original path position more seriously. When the force is applied (Cycle 2) and removed (Cycle 3), the controller using the proposed motion estimation enables to maintain a stable tracking performance (Fig. 10c, RMSE: 24.7509 pixels, Max. error: 44.7854 pixels in Cycle 2; RMSE: 21.0204 pixels, Max. error: 36.5932 pixels in Cycle 3).

IV. Conclusion

This letter has presented an enhanced 2D motion estimation method in soft robot visual servo utilizing FBG sensors, and improved the performance of conventional template matching algorithm under dim featured-lacking and force-disturbed conditions. The optical FBG fiber is helically wrapped along the manipulator without requirements of assembly, providing an idea for increasing the capabilities of image processing in soft robotics by integrating with strain sensors. Sparse strain measurement enables to be an independent motion sensor, which could be combined with image processing to improve tracking accuracy.

In the experiments, robot is actuated to follow a template in the endoscopic view along a "Batman" path, in a simulated laparoscopic scenario containing swine viscera. We also validate the motion estimation method in another LEGO constructed scenario, where the robot is tested with external forces. The enhanced estimation method increased tracking accuracy by 82.3% when capturing the dark and feature-deficient liver surface with a close distance (2–4 cm). Under external force, the enhanced method enables to maintain the tracking in a stable state (RMSE: 24.7509 pixels), while the tracking using image processing only would encounter an incremental tracking error (RMSE: 48.8868 pixels in Cycle 2 and 107.6258 pixels in Cycle 3) even after the force is released. High-frequency disturbances were not tested due to the limitations of robot stiffness and pneumatic actuator dynamics. However, the high sensing frequencies of FBG ($>$100Hz) and camera ($\sim$30Hz) allow accurate robot motion estimation even under high-frequency disturbances. Point-to-point tracking performance and the effect of reflective water in the endoscopic view are also tested, as could be seen in the video attached.

In some peculiar conditions such as the case where the path to be followed is predefined by dense target points, the image processing method would meet a tradeoff between tracking accuracy and smooth tip motion (depending on step size). In our future work, we plan to expand this setup for 6D camera localization. We will employ computational mechanics (e.g. Finite element analysis [23]) to explore how the sensor placement and density would affect the pose estimation accuracy. Online-updating sensor fusion algorithms (e.g. linear Kalman filter) will be investigated for improving our method's adaptability to variable visual conditions.

References

[1] K. Hashimoto, "Visual servoing: Real-time control of robot manipulators based on visual sensory feedback," *World Sci. Series Robot. Automated Syst.*, vol. 7, pp. 199–200, 1993.

[2] S. Hutchinson, G. D. Hager, and P. I. Corke, "A tutorial on visual servo control," *IEEE Trans. Robot. Autom.*, vol. 12, no. 5, pp. 651–670, Oct. 1996.

[3] M. Azizian, M. Khoshnam, N. Najmaei, and R. V. Patel, "Visual servoing in medical robotics: A survey. I: Endoscopic and direct vision imaging–techniques and applications," *Int. J. Med. Robot. Comput. Assist. Surgery*, vol. 10, no. 3, pp. 263–274, 2014.

[4] H. Fu *et al.*, "Interfacing soft and hard: a spring reinforced actuator," *Soft Robot.*, 2019, doi: 10.1089/soro.2018.0118.

[5] K.-H. Lee *et al.*, "Nonparametric online learning control for soft continuum robot: An enabling technique for effective endoscopic navigation," *Soft Robot.*, vol. 4, no. 4, pp. 324–337, 2017.

[6] H. Wang W. Chen, X. Yu, T. Deng, X. Wang, and R. Pfeifer, "Visual servo control of cable-driven soft robotic manipulator," in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst.*, 2013, pp. 57–62.

[7] H. Wang, B. Yang, Y. Liu, W. Chen, X. Liang, and R. Pfeifer, "Visual servoing of soft robot manipulator in constrained environments with an adaptive controller," *IEEE/ASME Trans. Mechatronics*, vol. 22, no. 1, pp. 41–50, Feb. 2017.

[8] Y. Lu, C. Zhang, S. Song, and M. Q.-H. Meng, "Precise motion control of concentric-tube robot based on visual servoing," in *Proc. IEEE Int. Conf. Inf. Autom.*, 2017, pp. 299–304.

[9] J. D. Greer, T. K. Morimoto, A. M. Okamura, and E. W. Hawkes, "Series pneumatic artificial muscles (sPAMs) and application to a soft continuum robot," in *Proc. IEEE Int. Conf. Robot. Autom.*, 2017, pp. 5503–5510.

[10] G. Fang *et al.*, "Vision-based online learning kinematic control for soft robots using local Gaussian process regression," *IEEE Robot. Autom. Lett.*, vol. 4, no. 2, pp. 1194–1201, Apr. 2019.

[11] G. Bradski and A. Kaehler, "Learning OpenCV: Computer vision with the OpenCV library. Sebastopol, CA, USA: "O'Reilly Media, Inc."," 2008.

[12] M. Alatise and G. Hancke, "Pose estimation of a mobile robot based on fusion of IMU data and vision data using an extended Kalman filter," *Sensors*, vol. 17, no. 10, 2017, Art. no. 2164.

[13] H. Abidi *et al.*, "Highly dexterous 2-module soft robot for intra-organ navigation in minimally invasive surgery," *Int. J. Med. Robot. Comput. Assist. Surgery*, vol. 14, no. 1, 2018, Art. no. e1875.

[14] F. Cosentino, E. Tumino, G. R. Passoni, E. Morandi, and A. Capria, "Functional evaluation of the endotics system, a new disposable self-propelled robotic colonoscope: In vitro tests and clinical trial," *Int. J. Artif. Organs*, vol. 32, no. 8, pp. 517–527, 2009.

[15] H. Liu *et al.*, "Shape tracking of a dexterous continuum manipulator utilizing two large deflection shape sensors," *IEEE Sensors J.*, vol. 15, no. 10, pp. 5494–5503, Oct. 2015.

[16] R. Xu, A. Yurkewich, and R. V. Patel, "Curvature, torsion, and force sensing in continuum robots using helically wrapped FBG sensors," *IEEE Robot. Autom. Lett.*, vol. 1, no. 2, pp. 1052–1059, Jul. 2016.

[17] C. Shi *et al.*, "Shape sensing techniques for continuum robots in minimally invasive surgery: A survey," *IEEE Trans. Biomed. Eng.*, vol. 64, no. 8, pp. 1665–1678, Aug. 2017.

[18] R. J. Roesthuis, M. Kemp, J. J. van den Dobbelsteen, and S. Misra, "Three-dimensional needle shape reconstruction using an array of fiber bragg grating sensors," *IEEE/ASME Trans. Mechatronics*, vol. 19, no. 4, pp. 1115–1126, Aug. 2014.

[19] W. Zhuang *et al.*, "FBG based shape sensing of a silicone octopus tentacle model for soft robotics," *Optik*, vol. 165, pp. 7–15, 2018.

[20] S. Sefati, R. Hegeman, F. Alambeigi, I. Iordachita, and M. Armand, "FBG-based position estimation of highly deformable continuum manipulators: Model-dependent vs. data-driven approaches," in *Proc. IEEE Int. Symp. Med. Robot.*, 2019, pp. 1–6.

[21] I. M. Van Meerbeek, C. M. De Sa, and R. F. Shepherd, "Soft optoelectronic sensory foams with proprioception," *Sci. Robot.*, vol. 3, no. 24, 2018, Art. no. eaau2489.

[22] M. C. Yip and D. B. Camarillo, "Model-less feedback control of continuum manipulators in constrained environments," *IEEE Trans. Robot.*, vol. 30, no. 4, pp. 880–889, Aug. 2014.

[23] T. L. T. Lun, K. Wang, J. D. L. Ho, K. Lee, K. Y. Sze, and K. Kwok, "Real-time surface shape sensing for soft and flexible structures using fiber Bragg gratings," *IEEE Robot. Autom. Lett.*, vol. 4, no. 2, pp. 1454–1461, Apr. 2019.