


Closed-Loop Pose Control and Automated Suturing of Continuum Surgical Manipulators With Customized Wrist Markers Under Stereo Vision

Baibo Wu, Longfei Wang, Xu Liu, Linhui Wang, and Kai Xu , *Member, IEEE*

Abstract—The use of continuum manipulators in surgical applications can be beneficial because of their inherent safety from their structural compliance. However, the tip pose (position and orientation) accuracy of a continuum manipulator can be low when an external load or disturbance is applied. Closed-loop pose control is preferred to improve the pose accuracy. Considering the requirements stemmed from sterilization and medical device electromagnetic compatibility, the approach of fully utilizing the vision feedback from a stereo endoscopic camera for such a closed-loop pose controller is adopted. Wrist markers are attached to the end-effectors of the surgical continuum manipulators for pose information detection. And these wrist markers are designed in a well-considered way to generate enough variants. The closed-loop pose controller is then designed based on a stereo vision tracking algorithm that detects the position and orientation of the wrist marker. A switching feature is also incorporated into the controller to handle the situation when the closed-loop control is disrupted due to the visual occlusion or the operation transition in dual-arm manipulation. Experimental verification showed that the average errors of the pose feedback are 0.65 mm for the position and 1.05° for the orientation. The experiments of trajectory following under a 120-gram external load and automated dual-arm suturing were also conducted to verify the effectiveness of the proposed controller.

Index Terms—Medical robots and systems, surgical robotics: laparoscopy, motion control.

I. INTRODUCTION

BENEFITING from design compactness, proximal actuation scheme, structural compliance and distal dexterity, continuum manipulators gradually gain attentions in surgical applications such as the SPL (Single Port Laparoscopy) and the NOTES (Natural Orifice Transluminal Endoscopic Surgery) [1], [2]. Different from the articulated rigid-linked manipulators, a continuum manipulator usually consists of two or more

segments and its motion is mainly realized by the segments' bending. The kinematics modeling is mostly based on one of the following approaches: i) constant curvature bending assumption [3], ii) piecewise constant curvature approximation [4], iii) Cosserat rod theory [5], [6]. The first approach is widely used due to its simple formulation and the computational efficiency. However, the tip pose accuracy of a continuum manipulator can be still low, when the manipulator is under a non-negligible external load. In order to improve the motion accuracy, closed-loop control of such a continuum manipulator is highly desired.

Many sensor-based shape reconstruction methods for possible tip pose feedback have been investigated [7]. Using thin FBG (Fiber Bragg Grating) sensors is quite straightforward and accurate [8], [9]. But the FBG sensors and its modulator increase the system cost and the tool fabrication challenge quite considerably. Integrating EM (Electromagnetic) sensors into the continuum manipulators is also possible [10]. However, EM sensors are sensitive to electromagnetic interference, which narrows their uses in surgical applications particularly when high-frequency electrical surgical tools are present. Using the captured images to recover the shape or the tip pose of a continuum manipulator can follow a marker-based approach [11], [12] or a marker-less approach [13], [14]. The image-based approach is considered efficient in terms of the fact that a stereo endoscopic camera is often readily available in a laparoscopic or endoscopic surgical system. Moreover, the integration of the FBG or EM sensors requires additional instrumentation.

This paper hence proposes a closed-loop controller for the tip pose control of continuum surgical manipulators based on specially designed wrist markers and the vision feedback from a stereo endoscopic camera, as depicted in Fig. 1.

Extended from our preliminary proof-of-concept work in [15], substantially more work is included in this paper, forming the paper's main contributions: i) new wrist markers, containing 6-DoF (Degree of Freedom) pose information, are re-designed in a systematic way to generate enough variants for the identification of different surgical tools; ii) the stereo vision algorithm for pose estimation is modified to obtain the marker's position and orientation, while only the tip position was extracted in [15]; iii) the closed-loop controller is modified to realize 6-DoF control with a switching feature to handle the scenarios where the closed-loop control is disrupted due to the visual occlusion or the operation transition in dual-arm manipulation. An automated dual-arm grape suturing experiment together with other experiments was carried out to demonstrate the effectiveness of the proposed approach.

The remainder of this paper is organized as follows. Section II presents the wrist markers designs and the markers' tracking

Manuscript received February 27, 2021; accepted July 5, 2021. Date of publication July 14, 2021; date of current version July 28, 2021. This letter was recommended for publication by Editor Pietro Valdastron upon evaluation of the Associate Editor and Reviewers' comments. This work was supported in part by the National Key R&D Program of China under Grant 2017YFC0110800, and in part by the National Natural Science Foundation of China under Grant 51722507. (Corresponding author: Kai Xu.)

Baibo Wu, Longfei Wang, Xu Liu, and Kai Xu are with the School of Mechanical Engineering, Shanghai Jiao Tong University, Shanghai 200240, China (e-mail: wubaibo@sjtu.edu.cn; longfei.wang@sjtu.edu.cn; xu.liu@sjtu.edu.cn; k.xu@sjtu.edu.cn).

Linhui Wang is with the Department of Urology, Changhai Hospital, Naval Medical University, Shanghai 200433, China (e-mail: wanglinhuic@163.com).

This letter has supplementary downloadable material available at <https://doi.org/10.1109/LRA.2021.3097260>, provided by the authors.

Digital Object Identifier 10.1109/LRA.2021.3097260

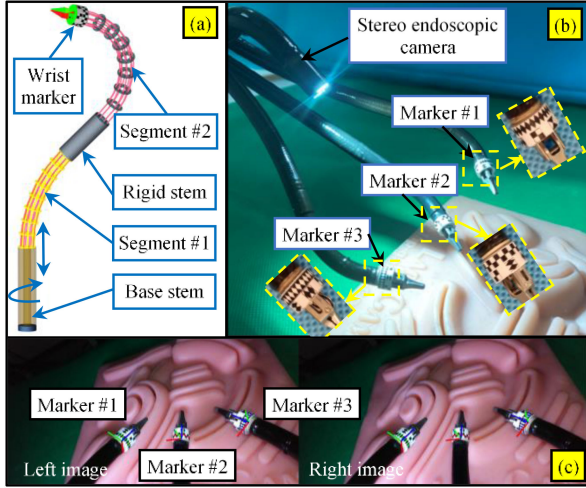


Fig. 1. The proposed closed-loop pose control of surgical continuum manipulators: (a) the manipulator's structure, (b) the setup with three different markers attached to the distal tips of three continuum manipulators, (c) the endoscopic view with the detected marker pose.

algorithm is proposed in Section III. The kinematics and the closed-loop pose controller of the continuum manipulator are presented in Section IV. The experimental verifications are reported in Section V with the conclusions summarized in Section VI.

II. DESIGN OF THE WRIST MARKERS

Surgical tool tracking can follow a marker-based approach or a marker-less one (e.g., using color or geometry features [14]). From a surgical robot developer's point of view, it is preferred to use markers due to the superior feedback accuracy and detection efficiency this approach provides for closed-loop control.

Since the wrist markers shall be wrapped around the wrist's cylindrical surface and the endoscopic camera will never see the entirety of the wrist marker, the marker designs and the recognition algorithms for the ARTag [16] or the ArUco [17] cannot be directly used. Designs of cylindrical markers were investigated in a few existing studies. But the one in [11] only covers half of the cylindrical wrist surface, while the one in [12] has an excessive length of 80 mm. Wrist markers should be carefully designed in a systematic way.

Inspired by the preliminary design that allows real-time marker tracking in [15], the wrist markers are re-designed here for 6-DoF wrist pose and type information. The proposed marker designs are detailed as follows.

A. The Simple Pattern Elements

Firstly, the 26 pattern elements are proposed as shown in Fig. 2. Arbitrary four pattern elements compose a pattern unit. Each wrist marker is constructed by several pattern units.

B. The Cycled Patterns and Distinctive Patterns

The wrist marker contains repeated pattern units (named *cycled patterns*). The cycled patterns have easily identifiable *marker corners* that are shown as green dots in Fig. 3. The marker corners are to determine the position and pointing direction of the wrist marker.

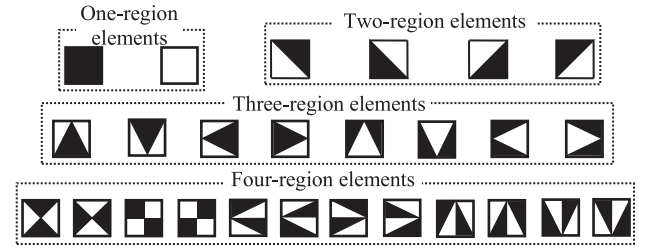


Fig. 2. The 26 pattern elements.

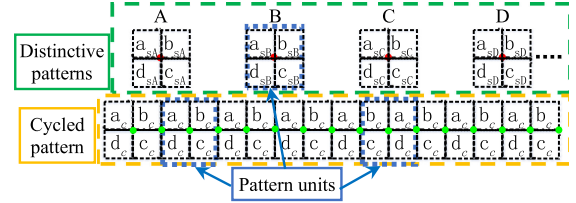


Fig. 3. The proposed structure of a wrist marker.

While forming the pattern units, two criteria were used: i) there should be at least 4 edge lines around the center of a pattern unit; ii) the correlation coefficients between two arbitrary pattern units shall be less than 0.3 to ensure the differentiability of different pattern units. The threshold of 0.3 is chosen empirically, considering a balance between the number of available patterns and the level of distinction of these pattern units.

Because the absolute rolling DoF about the axis of the wrist marker is absent when only the cycled patterns are used, additional information needs to be provided. One way to do so is to replace the cycled patterns with continuously distributed and different pattern units. Another way is to add an extra row of a group of separately distributed and different pattern units (named *distinctive patterns*). For the convenience of the following discussion, the former way is named *single-row structure* while the latter is named *two-row structure*.

As detailed in Section III.A, before a pattern unit can be identified, a lot of corner candidates will be detected in the ROI (Region of Interest). And the template matching approach is applied to check whether a corner candidate is a marker corner of a pattern unit. When the number of possible pattern templates to be compared is large, the image processing time can hardly be shortened.

For the two-row structure, when three surgical manipulators are deployed, three cycled patterns will be known and only six possible pattern units can be seen from the endoscopic view, since each cycled pattern contain two different pattern units ($a_c-b_c-c_c-d_c$ and $b_c-a_c-d_c-c_c$) as illustrated in Fig. 3. Then a corner candidate only needs to be compared to six pattern templates to determine whether it is one marker corner of a specific wrist marker. Since the marker geometry and the marker design is known, highly accurate ROIs will be examined for the detection of the deployed distinctive patterns to determine the rolling angle and the specific wrist type.

For the single-row structure, when three surgical manipulators are used, they will contain $3 \cdot n_{pu}$ pattern units in their wrist markers, where n_{pu} is the number of pattern units in one wrist marker. Then, $6 \cdot n_{pu}$ possible templates should be compared for each corner candidate. This would increase the image processing

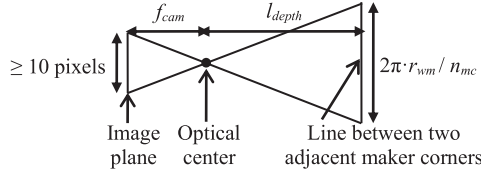


Fig. 4. The pinhole camera model for determining n_{mc} .

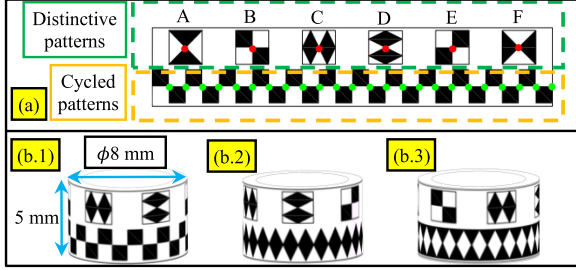


Fig. 5. Designs of the wrist markers: (a) the printed planar pattern; (b) three designs examples that are also shown in Fig. 1.

time substantially. What's more, the single-row structure can generate fewer wrist markers than the two-row structure, which will be discussed later.

Considering the multiple manipulator requirements and computational efficiency, the two-row structure is adopted.

C. The Generation of Wrist Markers

Using arbitrary combinations of four out of the 26 pattern elements, 166 pattern units are generated for the distinctive patterns, while 69 pattern units are generated for the cycled patterns. The reason that fewer pattern units can be used for the cycled patterns is as follows. In the cycled patterns, the two pattern units, $a_c-b_c-c_c-d_c$ and $b_c-a_c-d_c-c_c$ in Fig. 3, should still comply with the aforementioned two criteria, even though they are generated from the same pattern elements.

The working distance l_{depth} of the stereo endoscopic camera in Fig. 1 ranges from 30 mm to 120 mm. The camera has a resolution of 1920×1080 pixel and an average focal length f_{cam} of 1150 pixels after calibration. When n_{mc} marker corners exist in the wrist marker's cycled patterns, the surface distance between two adjacent marker corners is $2\pi \cdot r_{wm} / n_{mc}$, where $r_{wm} = 4$ mm is the marker's radius. It is desired that even when the wrist marker is at the furthest position, the image distance between two adjacent marker corners is still larger than 10 pixels, in order to ensure corner detection quality. Referring to the pinhole camera model as in Fig. 4, n_{mc} should be smaller than $2\pi \cdot r_{wm} / l_{depth} \cdot f_{cam} / 10 = 24.07$. n_{mc} is then set to 24.

In order to facilitate the process of detecting the distinctive patterns as detailed in Section III.B, six distinctive pattern units are evenly arranged right above the corresponding marker corners of the cycled patterns. Theoretically, using three or four pattern units for the distinctive patterns is also possible. The choice of using six is to reduce the chance of marker detection failure due to the visual occlusion.

Then, one wrist marker can be designed, by picking *one* pattern unit from the generated 69 ones for the cycled patterns and picking *six* different pattern units from the 166 ones for the distinctive patterns. Three examples are shown in Fig. 5 and total $1862 = 69 \times \text{floor}(166/6)$ wrist markers can be designed

theoretically, which is far more than the required types of the surgical end-effectors (e.g., the da Vinci Xi system lists 47 types of surgical tools in its catalog).

If the single-row structure is adopted, 12 pattern units should be picked from the generated 69 cycled patterns in order to unequivocally define the orientation and the type of the manipulator. Less than 6 wrist markers can be designed.

III. STEREO VISION TRACKING ALGORITHM

The stereo vision tracking algorithm consists of three steps as explained in the following three sub-sections.

A. Detection of the Marker Corners

The marker corners of the cycled patterns should be detected first. The marker corner detection algorithm is realized via a ray-searching strategy, combining the corner likelihood method [18] and the template matching approach. The detailed algorithm is described as follows.

- 1) Image preprocessing: Firstly, a quadrate ROI is cropped and converted into grayscale. The ROI is centered at an averaged central location of all marker corners from the last image. Then all marker corners are to be identified within this ROI from the current image. The ROI's side-length is set to ten times of the average distance between the adjacent marker corners from last image. The ROI is set big enough to cover all the marker corners, assuming that there is no significant lag in the image feed and no abrupt movements of the manipulators. For the first image, the ROI is the full image itself. When the marker tracking fails for one frame, the ROI is expanded with a ratio that is empirically set at 1.05. The ROI can also be predicted using the kinematics of the surgical manipulator.
- 2) Find the first marker corner: The first marker corner is detected as follows. Firstly, the Corner Likelihood (CL) for each pixel in the ROI is calculated as follows in (1), referring to [18]. The values are stored to speed up later steps.

$$CL = \max(c_{xy}, c_{45})$$

$$c_{xy} = \varepsilon^2 \cdot |I_{xy}| - 1.5 \cdot \varepsilon \cdot (|I_{45}| + |I_{n45}|)$$

$$c_{45} = \varepsilon^2 \cdot |I_{45 \cdot 45}| - 1.5 \cdot \varepsilon \cdot (|I_x| + |I_y|) \quad (1)$$

where ε is set to 2 as in [18]; $I_x, I_{45}, I_y, I_{n45}$ are the first derivatives in the $0, \pi/4, \pi/2, -\pi/4$ directions, respectively; I_{xy} and $I_{45 \cdot 45}$ are the second derivatives in the $0, \pi/2$ and $\pi/4, -\pi/4$ directions, respectively.

The ROI is divided into 9×9 -pixel grids. And the Non-Maximum Suppression method is applied to each grid to obtain a corner candidate with the maximal CL value. The corner candidate needs to have a CL value larger than 0.06.

And the edge direction I_{angle} and its weight I_{weight} for each corner candidate are calculated as follows.

$$I_{angle} = \arctan(I_y/I_x), I_{weight} = \sqrt{I_x^2 + I_y^2} \quad (2)$$

With the I_{angle} and I_{weight} , the edge directions for a candidate corner are estimated using the mean shift method [18] within a 9×9 -pixel neighborhood round each corner candidate.

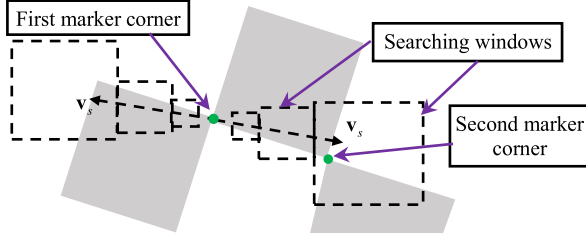


Fig. 6. The ray-searching strategy to detect marker corners.

The template matching approach is applied to check whether a corner candidate is a marker corner. The templates are the specific pattern units in the cycled patterns. Then, the corner candidates are traversed in the order of the CL values until a first marker corner is identified.

3) Find the other marker corners via a ray-searching strategy

A ray-searching strategy, as shown in Fig. 6, is adopted for detecting other marker corners. First, the second marker corner is searched using the searching direction \mathbf{v}_s , searching length and the first marker corner. For the first image, the searching direction \mathbf{v}_s is determined by the edge direction from (2); for the rest of the images, the searching direction is determined by the detected marker corners' distribution in the last image. The CL value is used to locate the second corner candidate at the location with the maximal CL value in a quadrate searching window. Finally, the template matching approach is applied again to check whether the second corner candidate is a marker corner.

Unlike the previous study [15] where a constant searching window was used, the searched window is doubled every step to reach a balance between search speed and the coverage of possible marker corner locations. After all, the marker corners are distributed on an elliptic path in the image. After the first two marker corners are detected, other marker corners are searched in a similar manner in the two opposite directions.

The marker corner search process stops as no marker corners are detected in a 40-pixel searching length. Because when a wrist marker is at a distance of 30 mm from the endoscopic camera, approximately 40 pixels will be between two adjacent marker corners. Finally, the detected marker corners are recorded for use of the next image.

B. Monocular Estimation of the Marker Pose

According to the detected marker corners and the known geometry of the wrist marker, the marker pose with respect to each lens of the stereo endoscopic camera is then calculated, with the coordinates for the wrist marker and the camera lens defined as follows.

- Wrist Marker Coordinate System $\{wm\} = [\hat{\mathbf{x}}_{wm} \hat{\mathbf{y}}_{wm} \hat{\mathbf{z}}_{wm}]^T$, as shown in Fig. 7(b), has its origin coincide with the center of the cycled patterns. The X-axis passes through the specific marker corner that is aligned with the 'A' pattern unit. The Z-axis is parallel to the wrist marker's axis, pointing towards the distinctive patterns.
- Intermediate Wrist Marker Coordinate System $\{wm0\} = [\hat{\mathbf{x}}_{wm0} \hat{\mathbf{y}}_{wm0} \hat{\mathbf{z}}_{wm0}]^T$ in Fig. 7(a) is rotated from $\{wm\}$ by a right-handed rotation such that the X-axis points to the leftmost detected marker corner (or the topmost one when the marker corners are vertically aligned).

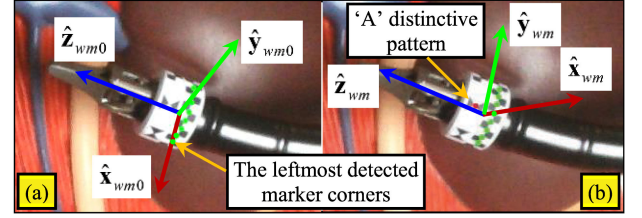


Fig. 7. Monocular estimation of the marker pose: (a) the intermediate wrist marker coordinate $\{wm0\}$; (b) the wrist marker coordinate $\{wm\}$.

- Lens Coordinate System $\{lens\} = [\hat{\mathbf{x}}_{lens} \hat{\mathbf{y}}_{lens} \hat{\mathbf{z}}_{lens}]^T$ has its origin coincide with the lens's center. The XY plane is perpendicular to its optical axis. ${}^{lens}\mathbf{R}_{wm0}$ and ${}^{lens}\mathbf{p}_{wm0}$ are the orientation and position of $\{wm0\}$ in $\{lens\}$, respectively.

As explained in Section II, the intermediate marker pose is first estimated. The spatial coordinates ${}^{wm0}\mathbf{p}_i$ of the i^{th} marker corner is $[r_{wm}\cos(i\cdot\beta) \ r_{wm}\sin(i\cdot\beta) \ 0]^T$, where $\beta = \pi/12$ is division angle of the marker corners. The intermediate marker pose (${}^{lens}\mathbf{R}_{wm0}$ and ${}^{lens}\mathbf{p}_{wm0}$) is estimated using the IPPE (Infinitesimal Plane-Based Pose Estimation) method [19].

Based on the estimated intermediate marker pose, a series of possible distinctive pattern regions are re-projected back to the image, according to the known geometry of the wrist marker and the calibrated camera parameters. This approach is more efficient than defining an orthogonal searching direction for these distinctive pattern units. Once one pattern unit from the distinctive patterns is recognized via the template matching method, the pointing direction of the X-axis of $\{wm\}$ will be immediately obtained, since the marker geometry is known. Namely, the marker pose (${}^{lens}\mathbf{R}_{wm}$ and ${}^{lens}\mathbf{p}_{wm}$) is obtained. In the aforementioned process, the sub-pixel localization [18] was applied to all the marker corners.

C. Optimization for the Marker Pose Under Stereo Vision

The marker pose (${}^{lens}\mathbf{R}_{wm}$ and ${}^{lens}\mathbf{p}_{wm}$) estimation process can be applied to both the left and the right lens to obtain ${}^{ll}\mathbf{R}_{wm}$ and ${}^{ll}\mathbf{p}_{wm}$ as well as ${}^{rl}\mathbf{R}_{wm}$ and ${}^{rl}\mathbf{p}_{wm}$, where $\{ll\}$ and $\{rl\}$ are defined referring to $\{lens\}$. $\{ll\}$ is related to $\{rl\}$ via a homogeneous transformation matrix ${}^{rl}\mathbf{T}_{ll}$. ${}^{rl}\mathbf{T}_{ll}$ can be obtained via a stereo rectification process. In the closed-loop control as detailed in Section IV, the marker pose is expressed in $\{ll\}$.

The marker pose (${}^{ll}\mathbf{R}_{wm}$ and ${}^{ll}\mathbf{p}_{wm}$) shall be refined via an optimization that minimizes the overall re-projection error of all the detected marker corners and pattern centers, as in (3).

$$\min_{{}^{ll}\mathbf{R}_{wm}, {}^{ll}\mathbf{p}_{wm}} \left(\sum_{i=1}^{n_{left}} \| {}^{left}\mathbf{m}_i - {}^{left}\mathbf{P}({}^{left}\mathbf{k}, {}^{ll}\mathbf{p}_{wm} + {}^{ll}\mathbf{R}_{wm} {}^{wm}\mathbf{p}_{n_i}) \| + \sum_{j=1}^{n_{right}} \| {}^{right}\mathbf{m}_j - {}^{right}\mathbf{P}({}^{right}\mathbf{k}, {}^{rl}\mathbf{p}_{wm} + {}^{rl}\mathbf{R}_{wm} {}^{wm}\mathbf{p}_{n_j}) \| \right) \quad (3)$$

where n_{left} , n_{right} are the number of the detected marker corners and pattern centers from the left and the right images, respectively; ${}^{left}\mathbf{m}_i$ and ${}^{right}\mathbf{m}_j$ are the image coordinates of the i^{th} point in the left image and the j^{th} point in the right image, respectively; ${}^{left}\mathbf{P}(\cdot)$ and ${}^{right}\mathbf{P}(\cdot)$ are the re-projection function for the left

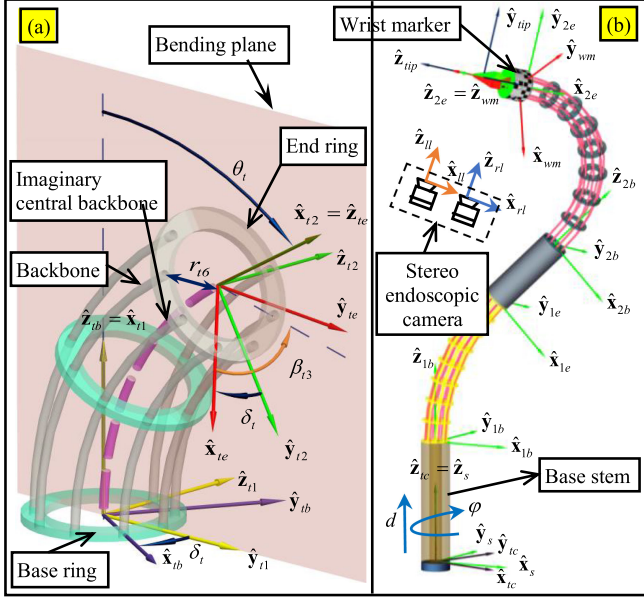


Fig. 8. Nomenclature and coordinates for: (a) the t^{th} continuum segment and (b) the continuum surgical manipulator that is used here from [20].

and right lens, respectively; $^{\text{left}}\mathbf{k}$ and $^{\text{right}}\mathbf{k}$ include the intrinsic and the distortion parameters of the left lens and the right lens, respectively; $^{\text{wm}}\mathbf{p}_{n_i}$ and $^{\text{wm}}\mathbf{p}_{n_j}$ are the 3D coordinates of the n_i^{th} and n_j^{th} points, respectively.

The Eq. (3) is solved using the Ceres Solver (an open-source C++ library from the Google Inc.). The marker pose ($^{\text{ll}}\mathbf{R}_{\text{wm}}$ and $^{\text{ll}}\mathbf{p}_{\text{wm}}$) is finally obtained for the closed-loop control when the optimization in (3) converges.

IV. KINEMATICS AND CLOSED-LOOP POSE CONTROLLER FOR THE CONTINUUM SURGICAL MANIPULATOR

In this study, the continuum surgical manipulator is from a previous study in [20], as shown in Fig. 8. The 6-DoF continuum manipulator is composed of two serially connected continuum segments and a base stem.

The kinematics of the continuum manipulator is summarized in Section IV.A, while the closed-loop pose controller is presented in Section IV.B.

A. Kinematics of the Continuum Manipulator

Referring to Fig. 8, the coordinates are detailed as follows with the details available in [20].

- Trocar Coordinate System $\{tc\} = [\hat{\mathbf{x}}_{tc} \hat{\mathbf{y}}_{tc} \hat{\mathbf{z}}_{tc}]^T$ is attached to the trocar channel with $\hat{\mathbf{z}}_{tc}$ aligned with the base stem.
- The Base Ring Coordinate System of the t^{th} segment $\{tb\} = [\hat{\mathbf{x}}_{tb} \hat{\mathbf{y}}_{tb} \hat{\mathbf{z}}_{tb}]^T$ is attached to the base ring of the t^{th} segment, with the XY plane coinciding with the base ring and the X axis passing through the first backbone.
- Bending Plane Coordinate System 1 of the t^{th} segment $\{t1\} = [\hat{\mathbf{x}}_{t1} \hat{\mathbf{y}}_{t1} \hat{\mathbf{z}}_{t1}]^T$ has its origin coincide with that of $\{tb\}$ and its XY plane coincides with the bending plane.
- Bending Plane Coordinate System 2 of the t^{th} segment $\{t2\} = [\hat{\mathbf{x}}_{t2} \hat{\mathbf{y}}_{t2} \hat{\mathbf{z}}_{t2}]^T$ has its origin located at the center

of the end ring of the t^{th} segment and its XY plane aligned with the bending plane of the t^{th} segment.

- End Ring Coordinate System of the t^{th} segment $\{te\} = [\hat{\mathbf{x}}_{te} \hat{\mathbf{y}}_{te} \hat{\mathbf{z}}_{te}]^T$ is attached to the end ring of the t^{th} segment, with the XY plane coinciding with the end ring and its X-axis passing through the first backbone.
- Tip Coordinate System of the manipulator $\{tip\} = [\hat{\mathbf{x}}_{tip} \hat{\mathbf{y}}_{tip} \hat{\mathbf{z}}_{tip}]^T$ is attached to the surgical end effector. It is translated from $\{2e\}$ by l^{tip}_{2e} along $\hat{\mathbf{z}}_{2e}$, where l^{tip}_{2e} is the distance between the origin of $\{2e\}$ and the origin of $\{tip\}$.

The tip pose can be calculated by the kinematics as follows.

$$\begin{aligned} {}^{tc}\mathbf{R}_{tip}^{kine} &= {}^{tc}\mathbf{R}_{1b} {}^{1b}\mathbf{R}_{1e} {}^{2e}\mathbf{R}_{2b} {}^{2e}\mathbf{R}_{2e} {}^{2e}\mathbf{R}_{tip} \\ {}^{tc}\mathbf{p}_{tip}^{kine} &= {}^{tc}\mathbf{T}_{1b} {}^{1b}\mathbf{T}_{1e} {}^{2e}\mathbf{T}_{2b} {}^{2e}\mathbf{T}_{2e} {}^{2e}\mathbf{p}_{tip} \end{aligned} \quad (4)$$

Jacobian matrix \mathbf{J} of the continuum manipulator is in (5).

$$\dot{\mathbf{x}} = \mathbf{J} \cdot \dot{\boldsymbol{\psi}} = \mathbf{J} \cdot [\dot{\varphi} \dot{\theta}_1 \dot{\delta}_1 \dot{\theta}_2 \dot{\delta}_2]^T \quad (5)$$

where $\boldsymbol{\psi} = [\varphi \ d \ \theta_1 \ \delta_1 \ \theta_2 \ \delta_2]^T$ is the configuration vector for the continuum manipulator.

The inverse kinematics is applied with a singularity robust implementation as in (6) to improve the numerical stability.

$$\mathbf{J}^+ = \begin{cases} \mathbf{J}^T (\mathbf{J}\mathbf{J}^T + \gamma \mathbf{I})^{-1}, & \sigma_{\min} < \xi \\ \mathbf{J}^{-1}, & \sigma_{\min} \geq \xi \end{cases} \quad (6)$$

where σ_{\min} is the smallest non-zero singular value; γ and ξ are small positive values (10^{-4} and 10^{-3} respectively in this paper).

When a wrist marker is attached to the tip of the continuum manipulator, the tip pose can also be expressed through the marker pose ($^{\text{ll}}\mathbf{R}_{\text{wm}}$ and $^{\text{ll}}\mathbf{p}_{\text{wm}}$).

$$\begin{aligned} {}^{tc}\mathbf{R}_{tip}^{vision} &= {}^{tc}\mathbf{R}_{ll} {}^{ll}\mathbf{R}_{wm} {}^{wm}\mathbf{R}_{tip} \\ {}^{tc}\mathbf{p}_{tip}^{vision} &= {}^{tc}\mathbf{R}_{ll} ({}^{ll}\mathbf{R}_{wm} {}^{wm}\mathbf{p}_{tip} + {}^{ll}\mathbf{p}_{wm}) + {}^{tc}\mathbf{p}_{ll} \end{aligned} \quad (7)$$

where ${}^{wm}\mathbf{R}_{tip}$ and ${}^{wm}\mathbf{p}_{tip}$ are the orientation and position of $\{tip\}$ in $\{wm\}$ and are known after the wrist marker installation. The orientation ${}^{tc}\mathbf{R}_{ll}$ and position ${}^{tc}\mathbf{p}_{ll}$ of $\{ll\}$ with respect to $\{tc\}$ are determined by the deployment of the stereo endoscopic camera. This info will usually be known from the control subsystem of the surgical robotic system.

The ${}^{tc}\mathbf{R}_{tip}^{kine}$ and ${}^{tc}\mathbf{p}_{tip}^{kine}$ from (4) and the ${}^{tc}\mathbf{R}_{tip}^{vision}$ and ${}^{tc}\mathbf{p}_{tip}^{vision}$ from (7) are the orientation and position of $\{tip\}$ in $\{tc\}$ obtained via the kinematics model and the stereo vision tracking algorithm, respectively. These two formulations for the tip pose are applicable to any of the deployed continuum manipulators.

B. Closed-Loop Pose Controller

The closed-loop pose controller based on the stereo vision tracking algorithm is proposed as in Fig. 9.

Referring to Fig. 9, the pose error during the k^{th} control cycle (position error ε_p^k and orientation error ε_R^k) is calculated by the target tip pose ($\mathbf{R}_t, \mathbf{p}_t$) and the current tip pose ($\mathbf{R}_c^k, \mathbf{p}_c^k$). The current tip pose will be acquired from the stereo vision tracking algorithm or the pose prediction by the kinematics.

The ε_p^k and the ε_R^k are calculated as follows.

$$\varepsilon_p^k = \mathbf{p}_t - \mathbf{p}_c^k; \varepsilon_R^k = d\theta(\mathbf{R}_t, \mathbf{R}_c^k) \quad (8)$$

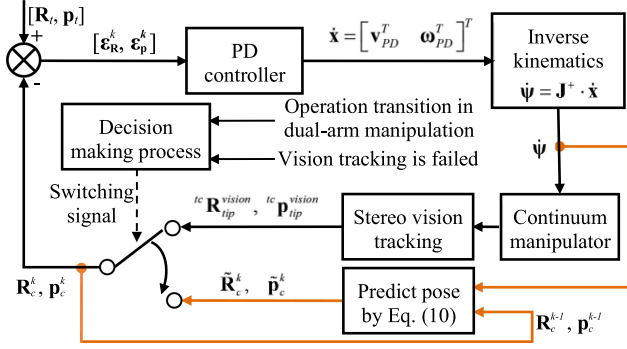


Fig. 9. The proposed closed-loop pose controller.

where $d\theta(\mathbf{R}_1, \mathbf{R}_2) = \text{rot}_{\hat{\theta}}^{-1}(\mathbf{R}_1 \cdot \mathbf{R}_2^T) \cdot \text{rot}_{\hat{\theta}}^{-1}(\mathbf{R}_1 \cdot \mathbf{R}_2^T)$ is the angle error vector between the rotation matrix \mathbf{R}_1 and the rotation matrix \mathbf{R}_2 . $\text{rot}_{\hat{\theta}}^{-1}(\mathbf{R})$, $\text{rot}_{\hat{\theta}}^{-1}(\mathbf{R})$ are the rotation angle and axis for a rotation matrix \mathbf{R} , respectively.

The desired linear velocity \mathbf{v}_{PD} and angular velocity ω_{PD} are then calculated by a PD controller as follows:

$$\begin{aligned} \mathbf{v}_{PD} &= P_v \cdot \varepsilon_p^k + D_v \cdot (\varepsilon_p^k - \varepsilon_p^{k-1}) \\ \omega_{PD} &= P_\omega \cdot \varepsilon_R^k + D_\omega \cdot (\varepsilon_R^k - \varepsilon_R^{k-1}) \end{aligned} \quad (9)$$

Next, the inverse kinematics in (6) is used to calculate the desired configuration velocity $\dot{\psi}$ which is applied to actuate the continuum manipulator.

The stereo vision tracking algorithm as detailed in Section III will detect the tip pose ($^{tc}\mathbf{R}_{tip}^{vision}$ and $^{tc}\mathbf{p}_{tip}^{vision}$) using (7). Meanwhile, the tip pose is also predicted in (10) with the last pose (\mathbf{R}_c^{k-1} , \mathbf{p}_c^{k-1}) and the configuration velocity $\dot{\psi}$ based on the kinematics.

$$\begin{aligned} \begin{bmatrix} \Delta \mathbf{p} \\ \text{rot}_{\hat{\theta}}^{-1}(\Delta \mathbf{R}) \cdot \text{rot}_{\hat{\theta}}^{-1}(\Delta \mathbf{R}) \end{bmatrix} &= \mathbf{J} \cdot (\dot{\psi} \cdot \Delta t) \\ \tilde{\mathbf{p}}_c^k &= \mathbf{p}_c^{k-1} + \Delta \mathbf{p}; \tilde{\mathbf{R}}_c^k = \Delta \mathbf{R} \cdot \mathbf{R}_c^{k-1} \end{aligned} \quad (10)$$

where Δt is the duration of a control cycle in the controller.

A switching mechanism based on a decision-making process is incorporated to select which pose to be feedback. The decision-making process includes two scenarios: 1) whether the vision tracking algorithm for the specific manipulator is failed; 2) whether there is an operation transition (e.g., for a dual-arm manipulation). So, the control mode can be switched between the open-loop mode [20] (as the pose is not measured directly) and the closed-loop mode.

V. EXPERIMENTAL VALIDATION

A series of experimental characterizations were conducted to illustrate the feasibility of the stereo vision tracking algorithm, the accuracy of the pose feedback, as well as the effectiveness of the closed-loop pose controller, even applied for an automated dual-arm grape suturing task.

A. Implementation of the Stereo Vision Tracking Algorithm

The stereo vision tracking algorithm implemented in C++ runs on a laptop with an Intel i7-9750H 2.60GHz CPU. The left and right images of the stereo endoscopic camera are captured at a rate of 60 FPS (Frame Per Second).

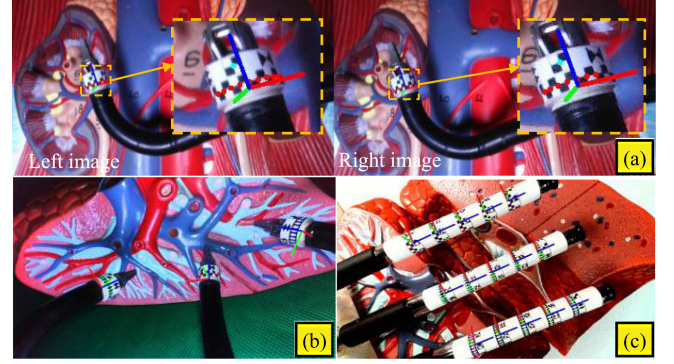


Fig. 10. Tracking different numbers of wrist markers: (a) one wrist marker; (b) three wrist markers; (c) fifteen wrist markers.

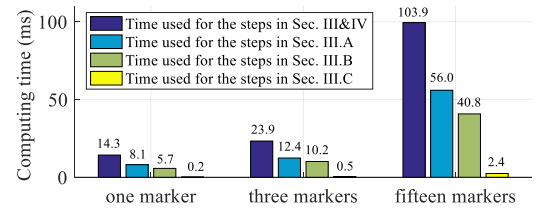


Fig. 11. Computing time of tracking different numbers of wrist markers.

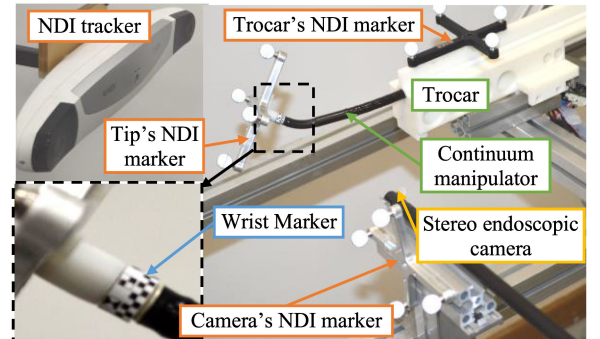


Fig. 12. The experimental setup for the accuracy verification.

Three situations with different numbers of wrist markers were tested as in Fig. 10. The left and right images are shown in Fig. 10(a) while only the left images are shown in Fig. 10(b) and (c). The red, green, and blue lines represent the X-axis, Y-axis, and Z-axis of different $\{wm\}$, respectively.

In Fig. 10(c), fifteen wrist markers were tracked simultaneously. These fifteen markers are generated with five groups of distinctive patterns (each group contains six different pattern units) and three types of cycled patterns.

The average computing time spent on the steps in Sections III and IV is in Fig. 11, with the blue bar as the total time.

B. Accuracy Verification of the Pose Feedback

The experimental setup for the accuracy verification of the pose feedback was shown in Fig. 12. An NDI tracker (Polaris Vega tracker, Northern Digital, Inc.) is used to obtain the position and orientation of $\{tip\}$ and $\{II\}$ with respect to $\{tc\}$ through the three NDI markers.

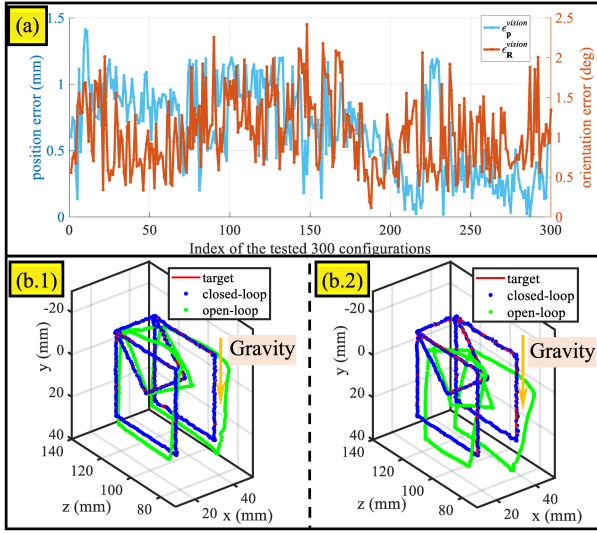


Fig. 13. The experimental results: (a) the vision tracking errors; (b.1) and (b.2) the trajectory-following curve under the 20-gram load and the 120-gram load (the blue and green dots are the tip position measured by the NDI tracker).

The continuum manipulator was commanded to move along the edges of a cube of 20 mm by 40 mm and 40 mm within the workspace, as shown in Fig. 13(b). In this trajectory, two target orientations were applied with $\mathbf{R}_t = \text{rot}_x(\pi/4)$ on the plane of $x_t = 35$ mm and $\mathbf{R}_t = \mathbf{I}_{3 \times 3}$ on the plane of $x_t = 15$ mm. Both the open-loop and the closed-loop controls were implemented under a 20-gram load (the weight of a NDI marker) and a 120-gram load, respectively. To suppress the vision tracking noises, the dead-zone for the closed-loop controller was set with $\varepsilon_P^0 = 0.35$ mm and $\varepsilon_R^0 = 0.5^\circ$.

During the experiments, the target tip pose, the tip pose measured by the NDI tracker, and the tip pose detected by the stereo vision tracking algorithm were recorded simultaneously. The vision tracking errors ($\varepsilon_P^{\text{vision}}$ and $\varepsilon_R^{\text{vision}}$) and the trajectory following errors ($\varepsilon_P^{\text{follow}}$ and $\varepsilon_R^{\text{follow}}$) were calculated as the pose errors between the NDI tracker and the stereo vision tracking algorithm, and the pose errors between the target and the NDI tracker, respectively.

In Fig. 13(a), the vision tracking errors, which are from the experiment of the closed-loop control under 20-gram load, are plotted. The average vision tracking error was 0.65 mm with a standard deviation of 0.34 mm for the position and 1.05° with a standard deviation of 0.47° for the orientation.

To validate the switching mechanism in the proposed controller, the experiment with 120-gram load was conducted again. A hex key was used to occlude the wrist marker occasionally as shown in Fig. 14(a).

The trajectory-following curves were plotted in Fig. 13(b) and Fig. 14(b) with the trajectory-following errors listed in Table I. The results showed that the tip pose error is greatly reduced after using the proposed closed-loop pose controller.

C. Automated Dual-Arm Grape Suturing

An automated dual-arm grape suturing task was conducted. Two wrist markers were attached to the tips of the two manipulators to provide the corresponding tip pose feedback.

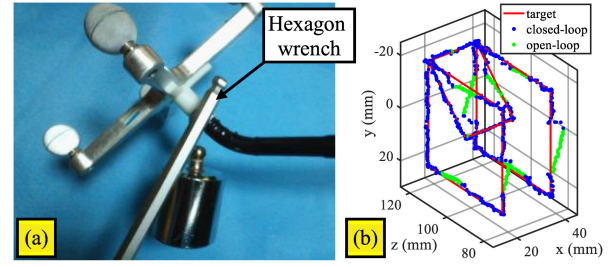


Fig. 14. The experiment for the validation of switching mechanism.

TABLE I
TRAJECTORY-FOLLOWING ERROR

	$\bar{\varepsilon}_P^{\text{follow}}$ (mm)	$\sigma(\varepsilon_P^{\text{follow}})$ (mm)	$\bar{\varepsilon}_R^{\text{follow}}$ (deg)	$\sigma(\varepsilon_R^{\text{follow}})$ (deg)
OC-20g	7.58	3.27	5.36	2.28
CC-20g	0.73	0.38	1.30	0.83
OC-120g	14.81	5.21	7.48	2.59
CC-120g	0.78	0.41	1.46	0.93

Note: The OC and CC are the abbreviations for the open-loop control and the closed-loop control, respectively; $\sigma(\cdot)$ is the standard deviation function.

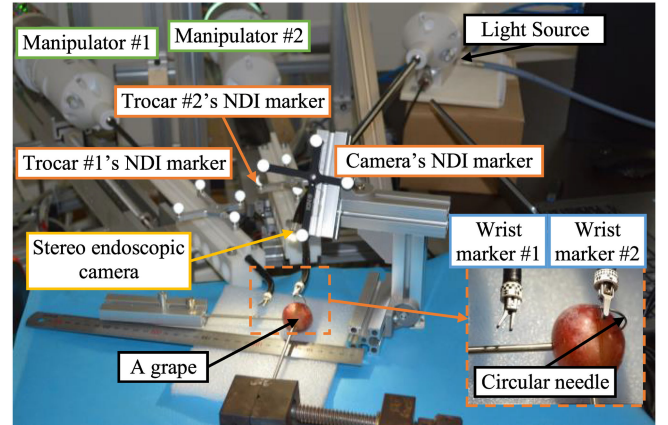


Fig. 15. The experimental setup for the dual-arm grape suturing task.

First, the position and orientation of the stereo endoscopic camera with respect to the trocar #1 and the trocar #2 were obtained by the three NDI markers, as in Fig. 15. The trajectories of the two manipulators were planned off-line with the following actions: grip or release the tail and the tip of the circular needle, position the needle's tip to the insertion point, and penetrate the grape and pull out the needle following a circular motion path.

Finally, the trajectories were commanded to move the two manipulators automatically, as shown in Fig. 16(a) to (f). This is an important step towards future autonomous suturing.

For the moment when the circular needle was gripped simultaneously by the two manipulators as shown in Fig. 16(c), an operation transition strategy was employed to avoid the excessive interactive force between the two manipulators. This strategy was implemented with one manipulator maintained in the closed-loop control mode while the other manipulator switched to the open-loop control mode using the switching mechanism in Fig. 9.

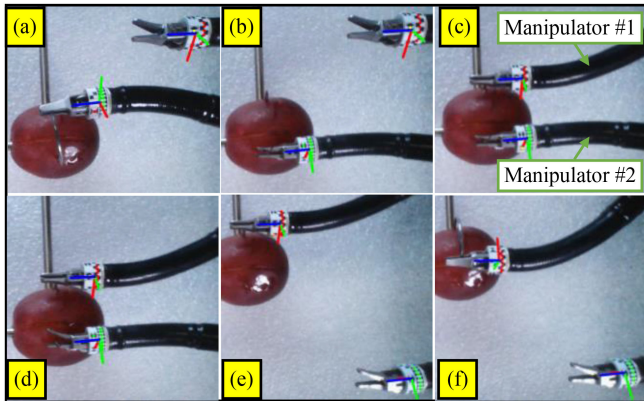


Fig. 16. (a) to (f) Captured images in the automated grape suturing task.

In Fig. 16(c), the manipulator #1 was maintained in the closed-loop control mode while the manipulator #2 was switched to the open-loop control mode. Thus, the manipulator #2 would follow the motion of the manipulator #1 passively due to its intrinsic compliance. When the manipulator #2 released the circular needle as shown in Fig. 16(d), the manipulator #2 was switched back to the closed-loop control mode.

VI. CONCLUSION

A closed-loop pose controller based on the customized wrist markers and the stereo vision from the endoscopic camera is proposed in this paper to improve the tip pose accuracy of the continuum surgical manipulators. Using this controller, an automated dual-arm suturing task was carried out.

The wrist markers are designed in a well-considered way to generate enough variants for the identification of different surgical end-effectors, as well as maintain its computational efficiency. The stereo vision tracking algorithm for the marker pose is extended from [15] to extract full pose information and distinguish the different wrist markers. A switching feature is incorporated in the proposed closed-loop pose controller to handle the scenarios of the visual occlusion and the operation transition in the dual-arm manipulation.

A series of experiments were conducted. The stereo vision tracking algorithm was proved to have the ability of simultaneously tracking multiple wrist markers. The accuracy of the pose feedback was 0.65 mm for the position and 1.05° for the orientation. In the closed-loop trajectory-following experiment under a 120-gram external load, the pose errors were greatly reduced (94.7% of the position errors and 80.5% of the orientation errors) compared the open-loop control, demonstrating the efficacy of the proposed approach.

Future works mainly concern two aspects. Firstly, the closed-loop pose controller shall be integrated into the surgical robot system with the consideration of the vision tool's motion. Secondly, a hybrid force/position controller based on the tip pose feedback can be investigated to regulate the interactive force between the manipulators or with the tissues.

REFERENCES

- [1] J. Burgner-Kahrs, D. C. Rucker, and H. Choset, "Continuum robots for medical applications: A survey," *IEEE Trans. Robot.*, vol. 31, no. 6, pp. 1261–1280, Dec 2015.
- [2] Y. Chen, S. a Zhang, Z. Wu, B. Yang, Q. Luo, and K. Xu, "A review of surgical robotic systems for keyhole and endoscopic procedures: State of the art and perspectives," *Front. Med.*, vol. 14, pp. 382–403, 2020.
- [3] R. J. Webster and B. A. Jones, "Design and kinematic modeling of constant curvature continuum robots: A review," *Int. J. Robot. Res.*, vol. 29, no. 13, pp. 1661–1683, Nov. 2010.
- [4] T. Mahl, A. Hildebrandt, and O. Sawodny, "A variable curvature continuum kinematics for kinematic control of the bionic handling assistant," *IEEE Trans. Robot.*, vol. 30, no. 4, pp. 935–949, Aug. 2014.
- [5] D. C. Rucker and R. J. Webster, "Statics and dynamics of continuum robots with general tendon routing and external loading," *IEEE Trans. Robot.*, vol. 27, no. 6, pp. 1033–1044, Dec. 2011.
- [6] J. Till and D. C. Rucker, "Elastic stability of cosserat rods and parallel continuum robots," *IEEE Trans. Robot.*, vol. 33, no. 3, pp. 718–733, Jun. 2017.
- [7] 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.
- [8] S. C. Ryu and P. E. Dupont, "FBG-based shape sensing tubes for continuum robots," in *Proc. IEEE Int. Conf. Robot. Automat.*, Hong Kong, China, 2014, pp. 3531–3537.
- [9] Y. Chitalia, N. J. Deaton, S. Jeong, N. Rahman, and J. P. Desai, "Towards FBG-based shape sensing for micro-scale and meso-scale continuum robots with large deflection," *IEEE Robot. Automat. Lett.*, vol. 5, no. 2, pp. 1712–1719, Apr. 2020.
- [10] A. M. Franz, T. Haidegger, W. Birkfellner, K. Cleary, T. M. Peters, and L. Maier-Hein, "Electromagnetic tracking in medicine—A review of technology, validation, and applications," *IEEE Trans. Med. Imag.*, vol. 33, no. 8, pp. 1702–1725, Aug. 2014.
- [11] L. Zhang, M. Ye, P.-L. Chan, and G.-Z. Yang, "Real-time surgical tool tracking and pose estimation using a hybrid cylindrical marker," *Int. J. Comput. Assist. Radiol. Surg.*, vol. 12, no. 6, pp. 921–930, Mar. 2017.
- [12] A. Gadwe and H. Ren, "Real-time 6DOF pose estimation of endoscopic instruments using printable markers," *IEEE Sens. J.*, vol. 19, no. 6, pp. 2338–2346, Mar. 2019.
- [13] D. B. Camarillo, C. R. Carlson, and J. K. Salisbury, "Configuration tracking for continuum manipulators with coupled tendon drive," *IEEE Trans. Robot.*, vol. 25, no. 4, pp. 798–808, Aug. 2009.
- [14] D. Bouget, M. Allan, D. Stoyanov, and P. Jannin, "Vision-based and marker-less surgical tool detection and tracking: A review of the literature," *Med. Image Anal.*, vol. 35, pp. 633–654, Jan. 2017.
- [15] H. Yang, B. Wu, X. Liu, and K. Xu, "A closed-loop controller for a continuum surgical manipulator based on a specially designed wrist marker and stereo tracking," in *Proc. IEEE/ASME Int. Conf. Adv. Intell. Mechatron.*, Boston, MA, USA, 2020, pp. 335–340.
- [16] M. Fiala, "Designing highly reliable fiducial markers," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 32, no. 7, pp. 1317–1324, Jul. 2010.
- [17] S. Garrido-Jurado, R. Muñoz-Salinas, F. J. Madrid-Cuevas, and M. J. Marín-Jiménez, "Automatic generation and detection of highly reliable fiducial markers under occlusion," *Pattern Recognit.*, vol. 47, no. 6, pp. 2280–2292, Jun. 2014.
- [18] A. Geiger, F. Moosmann, Ö. Car, and B. Schuster, "Automatic camera and range sensor calibration using a single shot," in *Proc. IEEE Int. Conf. Robot. Automat.*, Saint Paul, MN, USA, 2012, pp. 3936–3943.
- [19] T. Collins and A. Bartoli, "Infinitesimal plane-based pose estimation," *Int. J. Comput. Vis.*, vol. 109, no. 3, pp. 252–286, Jul. 2014.
- [20] S. a Zhang, Q. Li, H. Yang, J. Zhao, and K. Xu, "Configuration transition control of a continuum surgical manipulator for improved kinematic performance," *IEEE Robot. Automat. Lett.*, vol. 4, no. 4, pp. 3750–3757, Oct. 2019.