Visual Servo of a 6-DOF Robotic Stereo Flexible Endoscope Based on da Vinci Research Kit (dVRK) System

Xin Ma, Chengzhi Song, Philip Waiyan Chiu, and Zheng Li, *Member, IEEE*

***Abstract*—Endoscopes play an important role in minimally invasive surgery (MIS). Due to the advantages of less occupied motion space and enhanced safety, flexible endoscopes are drawing more and more attention. However, the structure of the flexible section makes it difficult for surgeons to manually rotate and guide the view of endoscopes. To solve these problems, we developed a 6-DOF robotic stereo flexible endoscope (RSFE) based on the da Vinci Research Kit (dVRK). Then an image-based endoscope guidance method with depth information is proposed for the RSFE. With this method, the view and insertion depth of the RSFE can be adjusted by tracking the surgical instruments automatically. Additionally, an image-based view rotation control method is proposed, with which the rotation of the view can be controlled by tracking two surgical instruments. The experimental results show that the proposed methods control the direction and rotation of the view of the flexible endoscope faster than the manual control method. Lastly, an ex vivo experiment is performed to demonstrate the feasibility of the proposed control method and system.**

***Index Terms*—Surgical robotics, laparoscopy, visual tracking, flexible robots.**

I. Introduction

ENDOSCOPES act as the eyes of the surgeon in MIS. Their safety and flexibility have major influence on the performance of surgeries. Therefore, many novel endoscopes have been developed in the last decade such as magnetic actuated endoscopes [1] and flexible endoscopes [2]. The magnetic actuated endoscopes [3] have realized noncontact control and smaller motion space occupation. But their performances are highly affected by the magnetic and electric fields, thus causing inconvenience during surgeries. Flexible endoscopes have the advantages of smaller occupied space, enhanced safety and higher dexterity [4]. However, most of existing flexible endoscopes used in the laparoscopic surgery are manually controlled, such as the EndoEYE [5] and an articulated universal joint based flexible access robot proposed in reference [6]. These manually controlled flexible endoscopes often bring confusion to camera drivers and increase the learning curve [7]. What's worse, driving manually controlled endoscopes will cause the fatigue of assistant surgeons. To address these challenges, some semi-robotic flexible endoscope systems have been proposed in references [8]–[10]. But only two or three DOFs can be controlled by motors in these semi-robotic flexible endoscope systems [8]–[10]. Tsutomu *et al.* propose a new 4 DOFs robotic-assisted flexible endoscope [11]. In our previous works [12], [13], a 6-DOF robotic monocular flexible endoscope based on dVRK is proposed. Yet, the results of the user study and ex-vivo experiments with the proposed endoscope show that the view control methods of the robotized flexible endoscope need to be further studied.

Recently, the remote-control method is used commonly to control robotic endoscopes [11]. For example, Naviot in [14] adopts two buttons as hand controller; the da Vinci platform, which is introduced by Intuitive [15], combines master tool manipulator with foot pedal tray to remotely control robotic endoscopes. However, these methods have some drawbacks such as discontinuous surgical flow. In 2017, an innovative master interface, which enables the continuous surgical flow, is introduced [16]. Another popular remote-control method is to adjust the view of the endoscope by using facial [17], eye [18] or voice [19] information. In 2003, Nishikawa *et al.* proposed a facial motion-based interface with which surgeons can control endoscopes by facial expressions [14]. Then, an eye motion-based interface is presented [18], in which stand-alone eye trackers are used to obtain the position of the eye gaze to control endoscope. However, it is difficult for surgeons to control their eye or facial motions during surgeries. In 2006, Nathan *et al.* tested the voice-controlled robotic assist scope holder AESOP [19]. It is not perfect either: operating surgeons cannot give verbal orders to assistants and AESOP at the same time. Worse still, manipulating these controllers makes it more complex and difficult for surgeons to perform operations.

Image-based control [20], [21], which can reduce the difficulties of operating endoscopes for surgeons, is another way to adjust the view of endoscopes. In 2005, a semi-robotic flexible

Manuscript received September 9, 2019; accepted December 24, 2019. Date of publication January 14, 2020; date of current version January 23, 2020. This article was recommended for publication by Associate Editor Elena De Momi and Editor Pietro Valdastri upon evaluation of the reviewers' comments. This work was supported by the Hong Kong General Research Grant, with under Project 14212316, 14207017, Early Career Scheme with under Project 24204818, ITF424/17 and the CUHK-SJTU Joint Research Collaboration Fund. *(Xin Ma and Chengzhi Song contributed equally to this work.) (Corresponding author: Zheng Li.)*

X. Ma and C. Song are with the Chow Yuk Ho Technology Centre for Innovative Medicine, The Chinese University of Hong Kong, Hong Kong (e-mail: maxin2000@mail.dlut.edu.cn; songchengzhiinhk@gmail.com).

P. W. Chiu and Z. Li are with the Department of Surgery and the Chow Yuk Ho Technology Centre for Innovative Medicine, The Chinese University of Hong Kong, Hong Kong (e-mail: philipchiu@surgery.cuhk.edu.hk; lizheng@cuhk.edu.hk).

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

Digital Object Identifier 10.1109/LRA.2020.2965863

endoscope system, whose view direction can be adjusted by tracking the markers attached on the surgical instruments, is developed. In the work [11], two DOFs of a flexible endoscope can be automatically controlled based on the target's depth information. The depth information is estimated by using the homography matrix. Then in 2014, surgical instruments tracking interface is applied on the rigid structure robot ViKY [22]. Our previous work extends this technique to a 6-DOF monocular flexible endoscope system [13]. A predictive controller is used to estimate the depth information of the surgical instruments. However, the depth information estimated by the predictive controller is not accurate. Additionally, the insertion depth and the rotation view of the endoscope cannot be adjusted automatically based on the above systems.

In this article, an RSFE integrated with the dVRK is proposed. It can provide the depth information of the end-effectors of surgical instruments by using the stereo vision technic. Then, an image-based endoscope guidance method with depth information is proposed for the RSFE. Furthermore, an image-based view rotation control method is presented. In short, the insertion depth and the rotation view of the existing commercial endoscope systems cannot be adjusted automatically. With the proposed method and system, the depth and the view can be automatically controlled by tracking surgical instruments without causing miscommunication. Main contributions of this paper are as follows:

1) developed a RSFE based on the tendon-driven continuum mechanism (TCM) [23] and integrated it with the dVRK; proposed an image-based endoscope view guidance method with depth information;
2) presented an image-based view rotation control method for the RSFE.

The rest of this paper is organized as follows. In Section II, the modelling of the RSFE is introduced. In Section III, the image-based view guidance method for the RSFE with depth information is detailed. In Section IV, the image-based view rotation control method is proposed. In Section V, the experimental results are shown in detail. At last, Section VI concludes this paper.

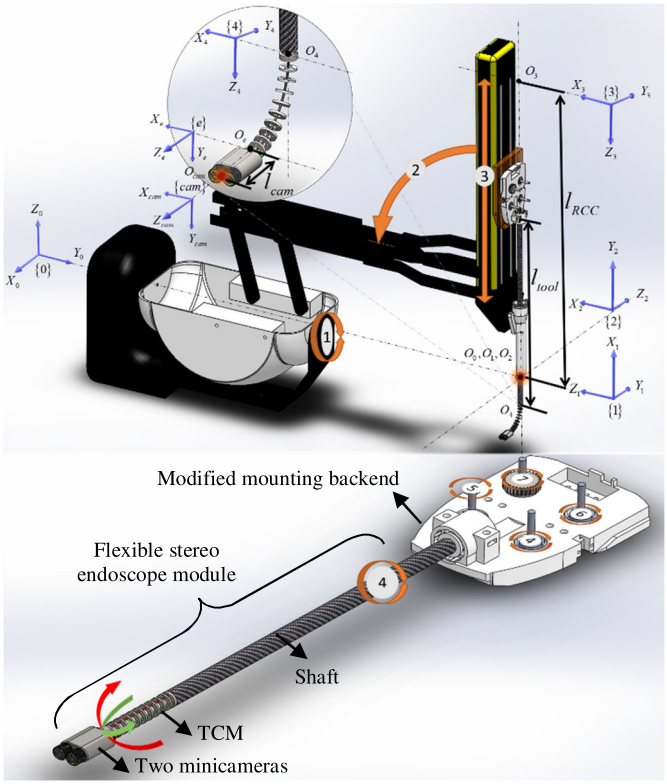


Fig. 1. The structure of the proposed RSFE and the coordinate frames for kinematics modeling.

TABLE I
HYBRID DH TABLE

Frame	Name	Type	a	α	D	θ	φ	k
1	Yaw	1	0	$\pi/2$	0	$q_1+\pi/2$	0	0
2	Pitch	1	0	$-\pi/2$	0	$q_2-\pi/2$	0	0
3	Insertion	2	0	$\pi/2$	q_3-l_{RCC}	0	0	0
4	Roll	1	0	0	l_{tool}	q_4	0	0
e	TCM	3	0	0	0	0	q_5	q_6
cam	Camera	0	0	0	l_{cam}	0	0	0

II. The Modeling of the RSFE

The structure of the RSFE is shown in Fig. 1. In this design, a flexible endoscope module is developed, which is mechanically integrated with the patient side manipulator (PSM) through a modified mounting backend. The flexible endoscope module includes two mini-cameras, a TCM and a shaft. The PSM is a 7-DOF actuated manipulator containing 7 joints, six of which are adopted in our proposed system. Joints 1 to 4 control the pitch, yaw, insertion and roll of the attached flexible endoscope module separately. The 5th joint and 6th joints are coupled together to actuate the orthogonal bending motions of the TCM. The coordinate frames for modeling are shown in Fig. 1. The attachments of frame {0} to frame {4} are the same as those of the original dVRK system [24]. The frame {4} is attached to the distal end of the shaft, which is also set as the base frame for the TCM. The frame {e} is attached to the distal end of the TCM. The frame {cam} is attached to the center of the two mini-cameras. In this article, the kinematic modeling of the proposed RSFE follows the Denavit-Hartenberg (DH) convention. The matrix ${}^{0}_{c}\boldsymbol{T}$ is the relationship between the joint positions and the pose of the frame {c}, which can be computed by the following equation:

$$ {}^{0}_{c}\boldsymbol{T} = {}^{0}_{1}\boldsymbol{T}^{1}_{2}\mathbf{T}^{2}_{3}\boldsymbol{T}^{3}_{4}\boldsymbol{T}^{4}_{e}\boldsymbol{T}^{e}_{c}\boldsymbol{T} \tag{1} $$

where ${}^{e}_{c}\boldsymbol{T}$ is the transformation from the frame {e} to the frame {c}, ${}^{4}_{e}\boldsymbol{T}$ is the transformation from frame {4} to frame {e} and ${}^{j-1}_{j}\boldsymbol{T}$ ($j \in [1,4]$) stands for the transformation from frame $\{j-1\}$ to frame $\{j\}$, which can be obtained with the hybrid table of DH parameters shown in Table I. In Table I, parameter a represents the movement along the X axis relative to the current frame; α represents the rotation around the X axis relative to the current frame; D represents the movement along the Z axis relative to the current frame; θ represents the rotation

around the Z axis relative to the current frame; k represents the curvature of the TCM, and φ represents the bending angle out of bending plane. $q_1, q_2, q_3, q_4, q_5, q_6$ represent the absolute angles of the 1st, 2nd, 3rd, 4th, 5th and 6th joint, respectively. The structure parameters in the system are: $l_{RCC} = 0.4318\,m$, which is the distance from the origin of the 3rd frame to the remote center point; $l_{tool} = 0.4162\,m$, which is the length of the shaft; $l_{cam} = 0.03\,m$, which is the distance from the origin of frame {e} to frame {cam}. Type 1, 2, 3 and 0 represent revolute joint, prismatic joint, flexible joint and no joint, respectively.

In this paper, the direct kinematics of the TCM (${}^4_eT$) can be drawn from the following matrix [25]: eq. (2) shown at the bottom of this page, where $c_{ks} = \cos(k \cdot s)$, $s_{ks} = \sin(k \cdot s)$, $c_\varphi = \cos(\varphi)$, $s_\varphi = \sin(\varphi)$, s stands for the length of the TCM, φ and k are the direction out of bending plane and the curvature of the TCM, separately. Lastly, the transformation from frame {e} to the frame on camera {c} is a constant matrix ${}^e_c\boldsymbol{T}$:

$$ {}^e_c\boldsymbol{T} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & l_{cam} \\ 0 & 0 & 0 & 1 \end{bmatrix} \tag{3} $$

The relationship between the joint motion and the motion of the frame {cam} with respect to the frame {0} is described through the Jacobian ${}^0\boldsymbol{J}_c$. ${}^0\boldsymbol{J}_c(i_1),\ i_1 \in [1,6]$ is the i_1-th column of the ${}^0\boldsymbol{J}_c$. The first 4 columns of the ${}^0\boldsymbol{J}_c$ with reference to the frame {0} (${}^0\boldsymbol{J}_c(i_1),\ i_1 \in [1,4]$) can be obtained by differential inverse kinematics equation from [26]. The 5th and 6th columns of ${}^0\boldsymbol{J}_c$ can be computed by the following equation:

$$ {}^0\boldsymbol{J}_c(i_1), i_1 \in [5,6] = \boldsymbol{Ad}_{g04} \cdot {}^4\boldsymbol{J}_c \tag{4} $$

where $\boldsymbol{Ad}_{g04} = \left[\begin{smallmatrix} {}^0_4\boldsymbol{R} & \boldsymbol{0}_{3\times3} \\ \boldsymbol{0}_{3\times3} & {}^0_4\boldsymbol{R} \end{smallmatrix}\right]_{6\times6}$, ${}^0_4\boldsymbol{R}$ is the rotation matrix obtained from ${}^0_4\boldsymbol{T}$, and ${}^4\boldsymbol{J}_c$ can be computed by the following equation:

$$ {}^4\boldsymbol{J}_c = \begin{bmatrix} \frac{(\mathrm{c}_{ks}-1)\mathrm{s}_\varphi}{k} - \mathrm{s}_{ks} \cdot l_{cam} \cdot \mathrm{s}_\varphi & c_1 \cdot \mathrm{c}_\varphi + s \cdot \mathrm{c}_{ks} \cdot l_{cam} \cdot \mathrm{c}_\varphi \\ \mathrm{s}_{ks} \cdot l_{cam} \cdot \mathrm{c}_\varphi - \frac{(\mathrm{c}_{ks}-1)\mathrm{c}_\varphi}{k} & c_1 \cdot \mathrm{s}_\varphi + s \cdot \mathrm{c}_{ks} \cdot l_{cam} \cdot \mathrm{s}_\varphi \\ 0 & -\frac{(\mathrm{s}_{ks}-k\cdot s\cdot \mathrm{c}_{ks})}{k^2} - s \cdot \mathrm{s}_{ks} \cdot l_{cam} \\ -\mathrm{c}_\varphi \cdot \mathrm{s}_{ks} & -s \cdot \mathrm{s}_\varphi \\ -\mathrm{s}_\varphi \cdot \mathrm{s}_{ks} & s \cdot \mathrm{c}_\varphi \\ 1 - \mathrm{c}_{ks} & 0 \end{bmatrix} \tag{5} $$

where $c_1 = \frac{(\mathrm{c}_{ks} + k\cdot s\cdot \mathrm{s}_{ks} - 1)}{k^2}$.

The relationship between the velocity of the joints ($\dot{\boldsymbol{Q}}$) and the velocity of the frame {cam} ($\dot{\boldsymbol{X}}$) is shown as follows:

$$ {}^0\boldsymbol{J}_c\dot{\boldsymbol{Q}} = \dot{\boldsymbol{X}} \tag{6} $$

III. Vision Guidance With Depth Information

In this paper, two identical endoscope cameras are installed in parallel. Therefore, we assume that the focal lengths of two cameras (f) are the same, and b is the distance between the two cameras.

The relationship between the image coordinate system and camera coordinate system is shown as follows [27]:

$$ \begin{aligned} {}^lu &= f\frac{{}^cx + \frac{b}{2}}{{}^cz}, {}^lv = f\frac{{}^cy}{{}^cz} \\ {}^ru &= f\frac{{}^cx - \frac{b}{2}}{{}^cz}, {}^rv = f\frac{{}^cy}{{}^cz} \end{aligned} \tag{7} $$

where the $({}^cx\ {}^cy\ {}^cz)$ is the 3D coordinate of the point P in the camera frame. $\boldsymbol{p}^l = ({}^lu\ {}^lv)$ and $\boldsymbol{p}^r = ({}^ru\ {}^rv)$ are the coordinates of the point P on the left image and right image, respectively. f is the focal lengths of the two cameras. We set ${}^I\boldsymbol{p} = [{}^lu\ {}^lv\ {}^ru\ {}^rv]^T$. The derivative ${}^I\dot{\boldsymbol{p}}$ is given by:

$$ {}^I\dot{\boldsymbol{p}} = {}^c\boldsymbol{J}_I\dot{\boldsymbol{X}} = {}^c\boldsymbol{J}_I\,{}^0\boldsymbol{J}_c\dot{\boldsymbol{Q}} \tag{8} $$

where

$$ {}^c\boldsymbol{J}_I = \begin{bmatrix} -\frac{s_c}{b} & 0 & {}^lu\frac{s_c}{bf} & \frac{{}^lu\,{}^lv}{f} & \frac{-{}^lu({}^lu+{}^ru)}{2f} - f & {}^lu \\ 0 & -\frac{s_c}{b} & {}^lv\frac{s_c}{bf} & \frac{{}^lv^2}{f} + f & -{}^lv\frac{{}^lu+{}^ru}{2f} & -\frac{{}^lu+{}^ru}{2} \\ -\frac{s_c}{b} & 0 & {}^ru\frac{s_c}{bf} & \frac{{}^ru\,{}^lv}{f} & \frac{-{}^ru({}^lu+{}^ru)}{2f} - f & {}^lv \\ 0 & -\frac{s_c}{b} & {}^lv\frac{s_c}{bf} & \frac{{}^lv^2}{f} + f & -{}^lv\frac{{}^lu+{}^ru}{2f} & -\frac{{}^lu+{}^ru}{2} \end{bmatrix} $$

and $s_c = {}^lu - {}^ru$.

A. Task Function for the Mean of the Image Points on the Left Image

In order to make the surgical instruments at the center of the left image, let $\phi_a \in$ R2 denote a task function defined as the sample mean:

$$ \phi_a = \begin{pmatrix} \frac{1}{m}\sum_{i=1}^m {}^lu_i \\ \frac{1}{m}\sum_{i=1}^m {}^lv_i \end{pmatrix} \tag{9} $$

where m is the number of the surgical instruments detected on the left image and the $({}^lu_i, {}^lv_i)$ is the coordinate of the i-th surgical instrument's tip on the left image. The time derivative

$$ {}^4_e\boldsymbol{T} = \begin{bmatrix} (c_{ks}-1)\cdot {c_\varphi}^2 + 1 & (c_{ks}-1)\cdot s_\varphi \cdot c_\varphi & s_{ks}\cdot c_\varphi & \frac{(1-c_{ks})\cdot c_\varphi}{k} \\ (c_{ks}-1)\cdot s_\varphi \cdot c_\varphi & (1-c_{ks})\cdot {c_\varphi}^2 + c_{ks} & s_{ks}\cdot s_\varphi & \frac{(1-c_{ks})\cdot s_\varphi}{k} \\ -s_{ks}\cdot c_\varphi & -s_{ks}\cdot s_\varphi & c_{ks} & \frac{s_{ks}}{k} \\ 0 & 0 & 0 & 1 \end{bmatrix} \tag{2} $$

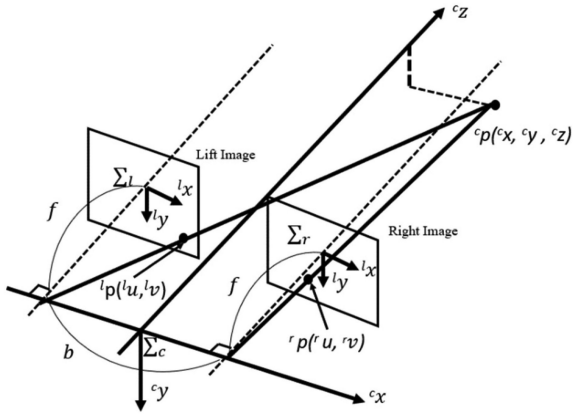


Fig. 2. The model of the stereo vision system. $\sum_l$ and $\sum_l$ represent the left and right image coordinate systems, respectively. $\sum_c$ represents the camera coordinate system.

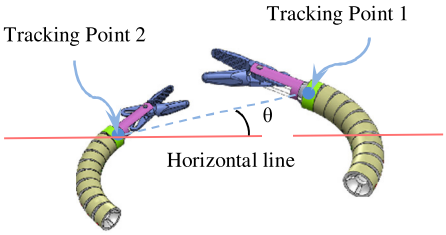


Fig. 3. The process of the angle calculation on the images.

ϕ_a is given by:

$$\dot{\phi}_a = \boldsymbol{J}_a \begin{bmatrix} {}^I\dot{\boldsymbol{p}}_1 \\ {}^I\dot{\boldsymbol{p}}_2 \\ \vdots \\ {}^I\dot{\boldsymbol{p}}_k \end{bmatrix} = \boldsymbol{J}_a \begin{bmatrix} {}^{c1}\boldsymbol{J}_I \\ {}^{c2}\boldsymbol{J}_I \\ \vdots \\ {}^{ck}\boldsymbol{J}_I \end{bmatrix} \dot{\boldsymbol{X}} = \boldsymbol{J}_a \begin{bmatrix} {}^{c1}\boldsymbol{J}_I \\ {}^{c2}\boldsymbol{J}_I \\ \vdots \\ {}^{ck}\boldsymbol{J}_I \end{bmatrix} {}^0\boldsymbol{J}_c\dot{\boldsymbol{Q}} \tag{10}$$

where $\boldsymbol{J}_a = \frac{1}{m}[\,{}^a\boldsymbol{g}_{2\times4} \cdots {}^a\boldsymbol{g}_{2\times4}\,]$ and ${}^a\boldsymbol{g}_{2\times4}$ is the 2×4 matrix repsseated for m times. ${}^a\boldsymbol{g}_{2\times4} = [\begin{smallmatrix}1&0&0&0\\0&1&0&0\end{smallmatrix}]$. ${}^{ci}\boldsymbol{J}_I$ is the ith target's Jacobian for image represented in the camera frame. ${}^I\dot{\boldsymbol{p}}_i$ is the time derivative of the ith ${}^I\boldsymbol{p}$.

B. Task Function for Insertion Depth Control

As shown in Fig. 2, the depth information of each tip of the surgical instruments can be obtained by the following equation:

$${}^cz_i = \frac{bf}{{}^lu_i - {}^ru_i} \tag{11}$$

where ${}^cz_{si}$ is the depth information of the ith surgical instrument's tip. $({}^lu_i, {}^lv_i)$ and $({}^ru_i, {}^rv_i)$ are the coordinates of the ith surgical instrument's tip on the left image and right image, respectively. Let $\phi_b \in$ R1 denote a task function defined as the mean depth of the target, which can be shown as follows:

$$\phi_b = \frac{1}{m}\sum_{i=1}^{m} z_i = \frac{1}{m}\sum_{i=1}^{m} \frac{bf}{{}^lu_i - {}^ru_i} \tag{12}$$

The time derivative of ϕ_b is given by:

$$\dot{\phi}_b = \boldsymbol{J}_b \begin{bmatrix} {}^I\dot{\boldsymbol{p}}_1 \\ {}^I\dot{\boldsymbol{p}}_2 \\ \vdots \\ {}^I\dot{\boldsymbol{p}}_k \end{bmatrix} = \boldsymbol{J}_b \begin{bmatrix} {}^I\boldsymbol{J}_{c1} \\ {}^I\boldsymbol{J}_{c2} \\ \vdots \\ {}^I\boldsymbol{J}_{ck} \end{bmatrix} \dot{\boldsymbol{X}} = \boldsymbol{J}_b \begin{bmatrix} {}^I\boldsymbol{J}_{c1} \\ {}^I\boldsymbol{J}_{c2} \\ \vdots \\ {}^I\boldsymbol{J}_{ck} \end{bmatrix} {}^0\boldsymbol{J}_c\dot{\boldsymbol{Q}} \tag{13}$$

where $\boldsymbol{J}_b = \frac{1}{m}[\,{}^b\boldsymbol{g}_{1\times4} \cdots {}^b\boldsymbol{g}_{1\times4}\,]$ and ${}^b\boldsymbol{g}_{1\times4}$ is the 1×4 matrix repeated for m times. ${}^b\boldsymbol{g}_{1\times4} = [\,-\frac{bf}{({}^lu_i - {}^ru_i)^2}\ 0\ \frac{bf}{({}^lu_i - {}^ru_i)^2}\ 0\,]$.

C. View and Insertion Depth Control for RSFE

In this paper, the optimal control method is used to achieve the image-based control with the depth information. The object function is shown as follows:

$$f(\dot{\boldsymbol{Q}}) = \beta_1((\dot{q}_1)^2 + (\dot{q}_2)^2) + \beta_2((\dot{q}_3)^2 + (\dot{q}_4)^2 + (\dot{q}_5)^2 + (\dot{q}_6)^2) \tag{14}$$

where β_1 and β_2 are weight factors to reduce the movements of the first and the second joints. The $\dot{q}^i$ is the angle changes of the ith joint and $\dot{\boldsymbol{Q}} = (\dot{q}^1, \dot{q}^2, \dot{q}^3, \dot{q}^4, \dot{q}^5, \dot{q}^6)^T$ can be estimated by minimizing the objective function. Here the Levenberg–Marquardt algorithm is used to perform the optimization.

$$\begin{aligned} &\text{minimize} \quad f(\dot{\boldsymbol{Q}}) \\ &\text{subject to} \quad \begin{bmatrix} \dot{\phi}_a \\ \dot{\phi}_b \end{bmatrix} = \begin{bmatrix} \boldsymbol{J}_a \\ \boldsymbol{J}_b \end{bmatrix} \begin{bmatrix} {}^{c1}\boldsymbol{J}_I \\ {}^{c2}\boldsymbol{J}_I \\ \vdots \\ {}^{ck}\boldsymbol{J}_I \end{bmatrix} {}^0\boldsymbol{J}_c\dot{\boldsymbol{Q}} \end{aligned} \tag{15}$$

IV. View Rolling With Vision Guidance

In this section, a novel image-based view rotation control method is proposed. With this method, the rotation of the view of the RSFE can be controlled by tracking two surgical instruments.

A. Image Processing Method

In this paper, green is chosen as the marker's color. Markers are attached on the surface of the surgical instruments' distal end [28]. The image processing method is as follows [11], [29]: (1) Color-based image segmentation method is utilized to segment the green areas on each RGB frame [30], [31]. Once green areas are selected, the RGB frame is converted to gray scale where it is then smoothed by using a Gaussian filter. The connected areas below a fixed size threshold (60 pixels) are removed. (2) shown at the bottom of the previous page, Then, the center coordinates of all selected connected areas are obtained by using the gray centroid method. (3) Lastly, the angle between the line connecting two points and the horizontal line on the images is calculated.

B. Task Function for the View Rotation of the Image

Let $\phi_c \in$ R1 denote a task function defined as the angle (θ) between the horizontal line and the obtained line on the image plane, which can be expressed as follows:

$$\phi_c = \theta = arc\tan\left(\frac{{}^l u_1 - {}^l u_2}{{}^l v_1 - {}^l v_2}\right) \tag{16}$$

where $({}^l u_1, {}^l v_1)$ is the coordinate of the first surgical instrument on the left image, and $({}^l u_2, {}^l v_2)$ is the coordinate of the second surgical instrument on the left image. We set $\Delta u = {}^l u_1 - {}^l u_2$ and $\Delta v = {}^l v_1 - {}^l v_2$. Then the derivative of ϕ_c is given by:

$$\dot{\phi}_c = \boldsymbol{J}_c \begin{bmatrix} {}^I\dot{\boldsymbol{p}}_1 \\ {}^I\dot{\boldsymbol{p}}_2 \end{bmatrix} = \boldsymbol{J}_c \begin{bmatrix} {}^I\boldsymbol{J}_{c1} \\ {}^I\boldsymbol{J}_{c2} \end{bmatrix} {}^0\boldsymbol{J}_c \dot{\boldsymbol{Q}} \tag{17}$$

where

$$\boldsymbol{J}_c = \begin{bmatrix} \dfrac{1}{1+\left(\frac{\Delta u}{\Delta v}\right)^2} & -\dfrac{\Delta u}{(1+\left(\frac{\Delta u}{\Delta v}\right)^2)\Delta v^2} & 0 & 0 \\ -\dfrac{1}{1+\left(\frac{\Delta u}{\Delta v}\right)^2} & \dfrac{\Delta u}{(1+\left(\frac{\Delta u}{\Delta v}\right)^2)\Delta v^2} & 0 & 0 \end{bmatrix}.$$

C. Rotation Control for RSFE

Due to possible assembly and manufacturing error, friction and plastic deformation of the wires with the proposed endoscope system, the view of the endoscope may deviate from its normal direction when it is being rotated. The deviation will, in turn, cause the surgical instruments to move out of the view. To solve this problem, in this paper, the instruments are tracked as the view is being rotated. The optimal control method is used, and the object function is shown as follows:

$$\begin{aligned} f(\dot{\boldsymbol{Q}}) = \beta_1((\dot{q}_1)^2 + (\dot{q}_2)^2) + \beta_2((\dot{q}_3)^2 \\ + (\dot{q}_4)^2 + (\dot{q}_5)^2 + (\dot{q}_6)^2) \end{aligned} \tag{18}$$

where β_1 and β_2 are weight factors to reduce the movements of the first and the second joints, and $\dot{\boldsymbol{Q}}$ can be estimated by minimizing the objective function. Here the Levenberg–Marquardt algorithm is used to perform the optimization.

$$\begin{aligned} &\text{minimize} \quad f(\dot{\boldsymbol{Q}}) \\ &\text{subject to} \quad \begin{bmatrix} \dot{\phi}_a \\ \dot{\phi}_c \end{bmatrix} = \begin{bmatrix} \boldsymbol{J}_a \\ \boldsymbol{J}_c \end{bmatrix} \begin{bmatrix} {}^{c1}\boldsymbol{J}_I \\ {}^{c2}\boldsymbol{J}_I \end{bmatrix} {}^c\boldsymbol{J}_0 \dot{\boldsymbol{Q}} \end{aligned} \tag{19}$$

V. Experiments

A. Experimental Platform

As shown in Fig. 4, the experimental platform consists of four parts: a flexible endoscope module, a PSM, a controller and a computer. The flexible endoscope module includes two mini-cameras, a TCM, a shaft and a backend. The resolution of the two mini-cameras is 640×480 with the 8 mm base length. The diameter of each mini-camera is 7.5 mm, and its frequency is 30 fps. The shaft is made by hollow carbon material with a length of 416 mm. The TCM is 40 mm long and is made with 8 vertebras equally placed along the elastic backbone. Two groups of actuating tendons are fixed on the distal end of TCM through the pilot holes on each vertebra. All tendons are passed through the hollow shaft and anchored on the capstans assembled on the modified mounting backend. The backend is an interface to integrate the flexible endoscope module into the robotic holding system. As the surgeons suggest, a foot pedal tray from dVRK is adopted to receive command from operating surgeons. The 'COAG' button is adopted to switch ON/OFF, the auto tracking function (direction tracking and depth tracking). The 'CAMERA' button is adopted to initiate the distal rolling for image-based view rotation.

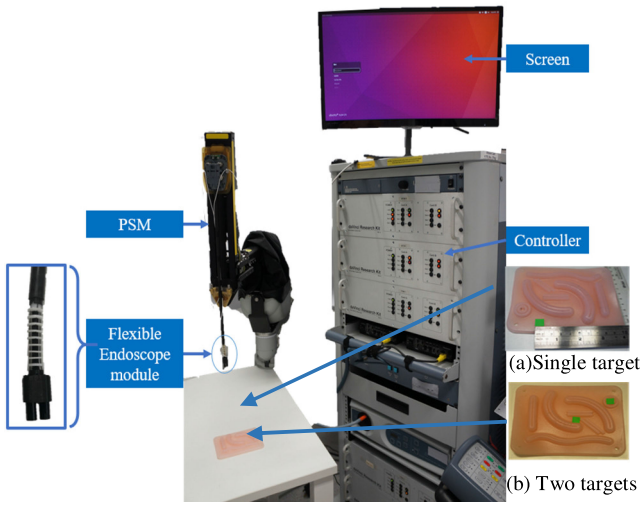
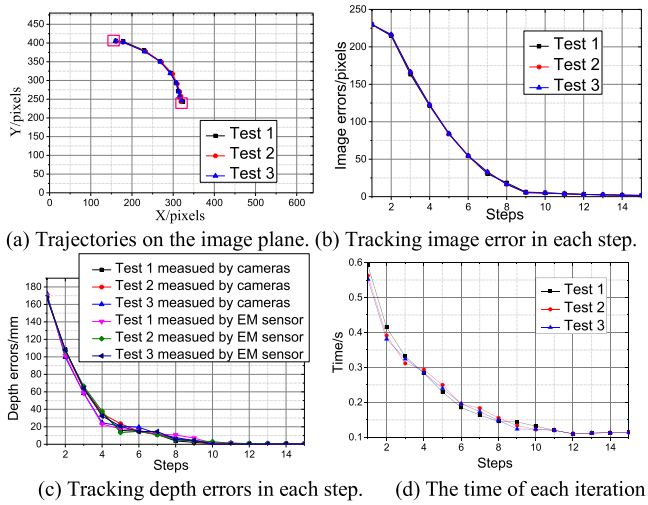


Fig. 4. Experimental platform.

B. Stationary Target Tracking Experiments

In this part, a stationary green marker is tracked by using the RSFE, as shown in Fig. 4(a). The diameter of the marker is 1cm. The value of the permissible image error is set as 10 pixels, and the value of the permissible depth error is set as 3 mm. The desired target image coordinate is the center coordinate of the image (320, 240), and the desired distance between the endoscope and the target is 60 mm. In the meantime, an NDI 3D Guidance trakSTAR system is used to measure the depth information of the target as ground truth. The static position accuracy of the 6 degrees-of-freedom (DOF) EM sensor is 1.4 mm within the measurable ranges. The measurable ranges of the trakSTAR system is 290 mm $\times$ 184 mm $\times$ 64 mm, and the update rate is 80HZ. The tracking tests are repeated for three times.

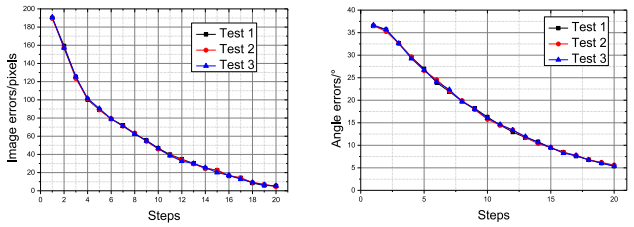
The stationary target tracking trajectories on the left image are shown in Fig. 5(a). The pink triangles are the initial and the desired positions on the image, respectively. The blue, red and black lines are the trajectories of the three tests, respectively. The image errors and depth errors are shown in Fig. 5(b) and Fig. 5(c), respectively. Fig. 5(d) shows the run time of each step. In this paper, the image error is defined as the distance between the target's current image coordinates and the desired image coordinates. The depth error is defined as the distance between the endoscope and the desired target's position in the camera coordinate system. We can see that the stationary target can be moved to the desired image position and the desired depth rapidly and reliably.



(a) Trajectories on the image plane. (b) Tracking image error in each step.

(c) Tracking depth errors in each step. (d) The time of each iteration

Fig. 5. Stationary target tracking results. (a) The pink points are the initial and the desired image positions, respectively. The blue, red and black lines are the trajectories of the three tests. (b) The blue, red and black lines are the image errors of each step in the three tests, respectively. (c) The blue, red and black lines are the target's depth information of each step in the three tests measured by two cameras, respectively. The blue, red and black dotted lines are the target's depth information of each step in the three tests measured by two EM sensors, respectively. (d) The blue, red and black lines are the run time of each step in the three tests, respectively.



(a) Distances on the image plane (b) Angle errors on the image plane

Fig. 6. Rotation control of the view of the RSFE. The blue, red and black lines represent the distance and angle errors in each test.

C. *View Rotating Experiments*

As shown in Fig. 4(b), in order to test the performance of the proposed view rotation control method, two stationary targets are tracked to control the view rotation of the RSFE three times. In this study, the line generated by two stationary targets is adjusted to be parallel to the horizontal line on the image. The initial angle between the obtained line and the horizontal line on the image is 37°. The initial distance between the midpoint of the two targets and the center of the left image is 190 pixels. The desired target image coordinates are the center coordinate of the image (320, 240), and the desired angle is 0°. The value of the permissible image error is set as 10 pixels, and the value of the permissible angle error is set as 5°. The distances from the midpoint of the two targets to the center of the left images in each step are shown in Fig. 6(a). The angle errors on the image are shown in Fig. 6(b). The blue, red and black lines represent the image error and angle error in each test. We can see that the image error decreases from 190 pixels to less than 5 pixels in 20 steps, and the angle error decreases from 33° to less than 5° in 20 steps.

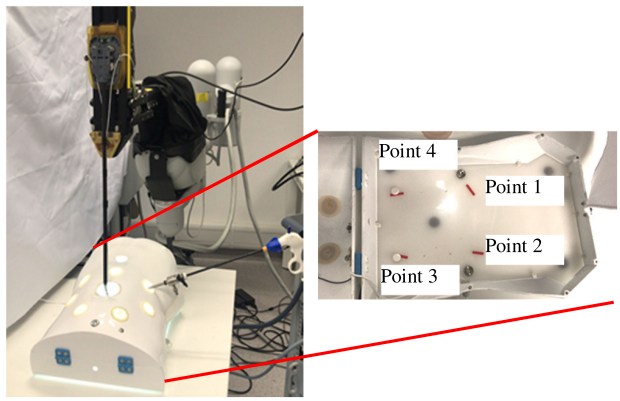
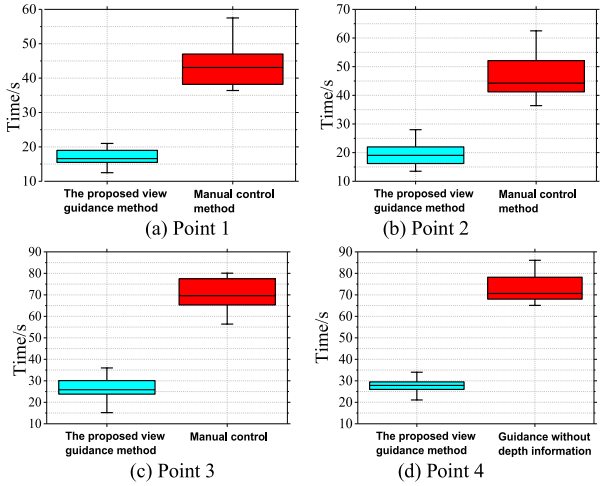


Fig. 7. The comparison experiments between the proposed view guidance method and the manual control method. The four red points are the target points which spread out in a phantom model at varying positions.



(a) Point 1 (b) Point 2

(c) Point 3 (d) Point 4

Fig. 8. The required time with both the proposed view guidance method and the manual control method. The blue boxes are the required time by using the proposed method. The red boxes are the required time by using the manual control method.

D. *Comparison Experiments Between the Proposed View Guidance Method and the Manual Control Method*

In this part, an experiment is designed in which a participant needs to control the RSFE with both the proposed control method and manual control method ten times. The purpose of this study is to assess the feasibility of the proposed control method by determining whether it brings any improvement in terms of speed and convenience over a manual control system. In this study, four red points are the target points which spread out in a phantom model at varying positions, as shown in Fig. 7. Then the participant needs to control the RSFE to make each target to the center of the field of the view. Meanwhile, insertion depth needs to be adjusted to make the target within the field depth of the mini-cameras. Manual control of the system is performed via a foot pedal input. The view guidance tests are repeated for ten times.

The required time of each test with both the proposed view guidance method and the manual control method is shown in Fig. 8. From the results we can see that the proposed image-based

TABLE II
VIEW ROTATING METHOD COMPARISON RESULTS

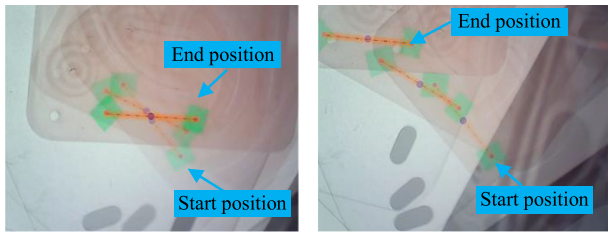
Initial angle (°)	Control time with kinematics-based (s)	Control time with proposed method (s)	Mean error with kinematics-based (pixel)	Standard deviation error with kinematics-based (pixel)	Mean error with proposed method (pixel)	Standard deviation error with proposed method (pixel)
30	5.11	4.65	101.17	85.34	5.35	3.71
40	7.46	6.35	128.63	88.51	4.13	3.60
50	8.34	7.19	142.73	92.39	4.03	4.84
60	9.10	8.09	143.13	94.04	4.24	4.32

view guidance method can move to the desired position faster than the manual control method.

E. Comparison Experiments Between the Proposed View Rotation Control Method and Kinematic Model based View Rotation Control Method

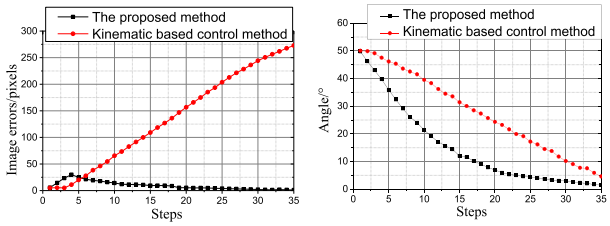
The view rotation of the endoscope can be controlled by both kinematic model-based view rotation control method and the proposed image-based view rotation method. In this part, an experiment is designed to compare the two view rotation control methods. Two methods are used to make the same line parallel to the horizontal line on the image. The line is generated by two stationary points on the image. The initial angle differences between the generated line and the horizontal line on the images are set as 30°, 40°, 50° and 60°, separately. The initial distance between the center of the two stationary points and the center of the image is 5 pixels in all tests. Image errors and control time are selected as criteria to compare the two methods. Image error is defined as the distance error between the generated line and the horizontal line on the images. Due to the manufacturing error, the stationary points in the view will gradually drift and go beyond the field of view with the kinematic model-based method, which will stop the control before reaching parallel with horizontal line. The stopping angle is θ_T. To compare the two methods, control time is defined as the time to reach θ_T for both kinematic model-based and the proposed control methods. From Table II, we can see that compared with the kinematic model-based control method, the proposed control method is faster and more accurate.

Taking the initial angel of 50° as an example, the images obtained from the left endoscope by using the two methods during view rotating are shown in Fig. 9, and the experimental results are shown in Fig. 10. We can see that the distance between the center of the two stationary points and the center of the image can be kept in a smaller value when rotating the view by using the proposed image-based view rotation method. However, due to possible assembly and manufacturing errors of the endoscope system, the distance between the center of the two stationary points and the center of the image may keep increasing by using the kinematic model-based view rotation control method. Additionally, compared with kinematic model-based view rotation control method, by using the proposed method, the angle between the obtained line and the horizontal line decreases faster.



(a) using the proposed image-based view rotation control method (b) using the kinematic model based view rotation control method

Fig. 9. The images obtained from the left endoscope with the two control methods during the view rotating.



(a) The distance error (b) Rotation

Fig. 10. View rotation experiment results with the initial rotation angel of 50°. (a) The red and black lines are the distance error bsy using the proposed image-based rotation control method and inverse kinematic based control method, respectively. (b) The red line is the run time by using the manual control method.

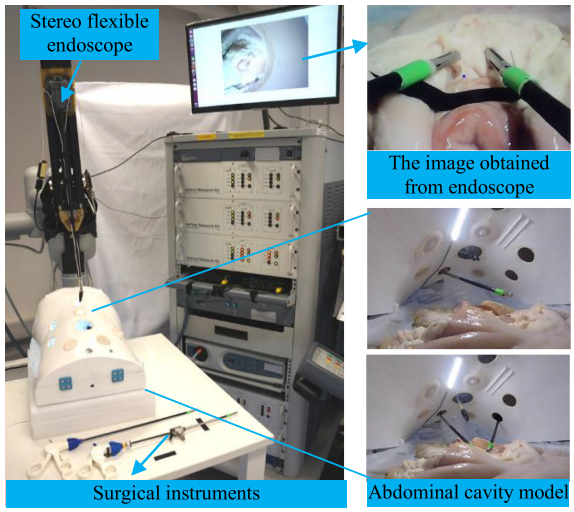


Fig. 11. Porcine Intestine and Stomach Inspection Experiment. The Left Image is the Setup of the Ex Vivo Experiments. The Right Image is the Images in the Abdominal Cavity Model.

F. Ex Vivo Experiments

In this section, an ex vivo test is conducted to demonstrate the feasibility of the RSFE and the proposed control methods.

In this experiment, a surgeon adjusts the view rotation angle of the RSFE with proposed image-based view rotation control method. Then the surgeon is required to observe the porcine intestine and stomach by using the proposed image-based view guidance method. As shown in Fig. 11, the size of the porcine stomach is around 15 cm $\times$ 30 cm, with a thickness ranging from 40 mm to 55 mm. The size of the porcine intestine is around

3 cm × 30 cm. All samples are contained in a test box with a width of 18cm and a length of 40 cm. The RSFE is introduced into the test box via an opening. During the observation, surgical instruments can be well tracked, and all the details of the experimental samples can be captured by the RSFE. The surgeon indicated that the trace function of the proposed endoscope will not interrupt the surgery or prolong the procedure. However, the continuous camera movements do cause the images to blur. Therefore, we set an area in the middle of the images, and when the markers' coordinates are in this area, the proposed endoscope will not track them [13]. Notice that, the proposed RSFE can be easily used in both manual and multi-port robotic laparoscopy. In manual laparoscopies, it can reduce surgeons' fatigue and avoid miscommunication. And in multi-port robotic laparoscopies, it can reduce the occupied space and avoid miscommunication.

VI. Conclusion

In this paper, a Robotic Stereo Flexible Endoscope is developed based on dVRK system. Then an image-based view guidance method with depth information is proposed for the RSFE. With this method, the view of the flexible endoscope can be adjusted by tracking surgical instruments, and the insertion depth of the flexible endoscope can be adjusted to keep the distance between the endoscope and the surgical instrument in the desired working range. Compared with the manual control method, the endoscope can reach the desired view faster and more accurately. Additionally, an image-based view rotation control method for the RSFE is presented, in this paper. With this method, the rotation of the view can be adjusted by tracking two surgical instruments. Compared with the traditional inverse kinematic based control method, the rotation of the view can be controlled faster and more accurately. Lastly, the feasibility of the proposed control method and system is demonstrated by an ex vivo experiment with an abdomen phantom. Our future work aims at reducing motion blur, integrating with augmented reality and applying the proposed system to in vivo experiments. Additionally, the marker attachment method has limitations such as contamination by blood and light source dependence. A learning-based instruments' tips identification method will be further studied.

References

[1] C. D. Natali, M. Beccani, N. Simaan, and P. Valdastri, "Jacobian-based iterative method for magnetic localization in robotic capsule endoscopy," *IEEE Trans. Robot.*, vol. 32, no. 2, pp. 327–338, Mar. 2016.

[2] Z. Li, M. Z. OO, V. Nalam, V. D. Thang, H. Ren, T. Kofidis, and H. Yu, "Design of a novel flexible endoscope—cardioscope," *J. Mechanisms Robot.*, vol. 8, no. 5, Apr. 2016 Art. no. 051014.

[3] T. Cheng, W. Li, C. S. H. Ng, P. W. Y. Chiu, and Z. Li, "Visual servo control of a novel magnetic actuated endoscope for uniportal video-assisted thoracic surgery," *IEEE Robot. Autom. Lett.*, vol. 4, no. 3, pp. 3098–3105, Jun. 2019.

[4] N. Simaan, R. Taylor, and P. Flint, "A dexterous system for Laryngeal surgery," in *Proc. IEEE Int. Conf. Robot. Autom.*, 2004, pp. 351–357.

[5] "Our Products, OLYMPUS MEDICAL Hong Kong, 2019. [Online]. Available: http://olympusmedical.com.hk/products/index.html. [Accessed: 04-Sep. 2019].

[6] J. Shang *et al.*, "An articulated universal joint based flexible access robot for minimally invasive surgery," in *Proc. IEEE Int. Conf. Robot. Autom.*, 2011, pp. 1147–1152.

[7] J. Perrone, C. Ames, Y. Yan, and J. Landman, "Evaluation of surgical performance with standard rigid and flexible-tip laparoscopes," *Surgical Endoscopy*, vol. 19, no. 10, pp. 1325–1328, 2005.

[8] B. P. M. Yeung and T. Gourlay, "A technical review of flexible endoscopic multitasking platforms," *Int. J. Surgery*, vol. 10, no. 7, pp. 345–354, 2012.

[9] L. Ott, F. Nageotte, P. Zanne, and M. D. Mathelin, "Simultaneous physiological motion cancellation and depth adaptation in flexible endoscopy using visual servoing," in *Proc. 2nd Int. Conf. Biomed. Robot. Biomechatronics*, 2008, pp. 568–573.

[10] A. Agustinos, R. Wolf, J. A. Long, P. Cinquin, and S. Voros, "Visual servoing of a robotic endoscope holder based on surgical instrument tracking," in *Proc. 5th IEEE RAS/EMBS Int. Conf. Biomed. Robot. Biomechatronics*, 2014, pp. 13–18.

[11] T. Iwasa *et al.*, "A new robotic-assisted flexible endoscope with single-hand control: Endoscopic submucosal dissection in the ex vivo Porcine stomach," *Surgical Endoscopy*, vol. 32, no. 7, pp. 3386–3392, Jul. 2018.

[12] C. Song, X. Ma, X. Xia, P. W. Y. Chiu, C. C. N. Chong, and Z. Li, "A robotic flexible endoscope with shared autonomy: A study of mockup Cholecystectomy," *Surgical Endoscopy*, 2019. [Online]. Available: https://www.ncbi.nlm.nih.gov/pubmed/31722046.

[13] X. Ma, C. Song, P. W. Chiu, and Z. Li, "Autonomous flexible endoscope for minimally invasive surgery with enhanced safety," *IEEE Robot. Autom. Lett.*, vol. 4, no. 3, pp. 2607–2613, Jan. 2019.

[14] T. Yasunaga *et al.*, "Remote-controlled laparoscope manipulator system, Naviot, for endoscopic surgery," in *Proc. Int. Congr. Series*, 2003, vol. 1256, pp. 678–683.

[15] Intuitive Surgical, Intuitive|da Vinci|Robotic Surgical Systems, 2019. [Online] Available at: https://www.intuitive.com/en-us/products-and-services/da-vinci [Accessed 4 Sep. 2019].

[16] M. Kim, C. Lee, N. Hong, Y. J. Kim, and S. Kim, "Development of stereo endoscope system with its innovative master interface for continuous surgical operation," *BioMed. Eng. OnLine*, vol. 16, no. 1, pp. 81-1–81-16, Dec. 2017.

[17] A. Nishikawa *et al.*, "FAce MOUSe: A novel human-machine interface for controlling the position of a laparoscope," *IEEE Trans. Robot. Autom.*, vol. 19, no. 5, pp. 825–841, Oct. 2003.

[18] U. Schwarz and T. Schmuckle, "Cognitive eyes," *Archives Suisses De Neurologie Et De Psychiatrie*, vol. 153, no. 4, pp. 175–179, 2002.

[19] C.-A. Nathan, V. Chakradeo, K. Malhotra, H. Dagostino, and R. Patwardhan, "The voice-controlled robotic assist scope holder AESOP for the endoscopic approach to the sella," *Skull Base*, vol. 16, no. 03, pp. 123–131, Aug. 2006.

[20] C. S. Bell, K. L. Obstein, and P. Valdastri, "Image partitioning and illumination in image-based pose detection for teleoperated flexible endoscopes," *Artif. Intell. Medicine*, vol. 59, no. 3, pp. 185–196, Nov. 2013.

[21] R. Reilink, S. Stramigioli, and S. Misra, "Image-based flexible endoscope steering," *IEEE/RSJ Int. Conf. Intell. Robots Syst.*, 2010, pp. 2339–2344.

[22] S.-Y. Ko, J. Kim, D.-S. Kwon, and W.-J. Lee, "Intelligent interaction between surgeon and laparoscopic assistant robot system," in *Proc. ROMAN. IEEE Int. Workshop Robot Human Interactive Commun.*, 2005, pp. 60–65.

[23] Z. Li, L. Wu, H. Ren, and H. Yu, "Kinematic comparison of surgical tendon-driven manipulators and concentric tube manipulators," *Mechanism Mach. Theory*, vol. 107, pp. 148–165, 2017.

[24] P. Kazanzides, Z. Chen, A. Deguet, G. S. Fischer, R. H. Taylor, and S. P. Dimaio, "An open-source research kit for the da Vinci surgical system," in *Proc. IEEE Int. Conf. Robot. Autom.*, pp. 6434–6439, 2014.

[25] 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, Oct. 2010.

[26] B. Siciliano, L. Sciavicco, L. Villani, and G. Oriolo, *Robotics: Modelling, Planning and Control*. London: Springer, 2010, Ch. 3.

[27] P. I. Corke, *Visual Control of Robots: High-Performance Visual Servoing*, Taunton, Somerset, England: Research Studies Press, pp. 151–169, 1997.

[28] Z. Jia *et al.*, "Pose measurement method and experiments for high-speed rolling targets in a wind tunnel," *Sensors*, vol. 14, no. 12, pp. 23933–23953, Dec. 2014.

[29] Z. Wang *et al.*, "Image-based trajectory tracking control of 4-DoF Laparoscopic instruments using a rotation distinguishing marker," *IEEE Robot. Autom. Lett.*, vol. 2, no. 3, pp. 1586–1592, Mar. 2017.

[30] F. Garcia-Lamont, J. Cervantes, A. López, and L. Rodriguez, "Segmentation of images by color features: A survey," *Neurocomputing*, vol. 292, pp. 1–27, May. 2018.

[31] L. Qian, A. Deguet, and P. Kazanzides, "ARssist: Augmented reality on a head-mounted display for the first assistant in robotic surgery," *Healthcare Technol. Lett.*, vol. 5, no. 5, pp. 194–200, Nov. 2018.