



Article

Stereo Visual Servoing Control of a Soft Endoscope for Upper Gastrointestinal Endoscopic Submucosal Dissection [†]

Jian Chen ^{1,2,3} , Shuai Wang ³, Qingxiang Zhao ³, Wei Huang ³, Mingcong Chen ⁴, Jian Hu ^{2,3}, Yihe Wang ³ and Hongbin Liu ^{2,3,5,*}

¹ School of Artificial Intelligence, University of Chinese Academy of Sciences, Beijing 100049, China; chenjian2020@ia.ac.cn

² Institute of Automation, Chinese Academy of Sciences, Beijing 100190, China; hujian@ia.ac.cn

³ Centre of AI and Robotics, Hong Kong Institute of Science and Innovation, Chinese Academy of Sciences, Hong Kong; shuai.wang@cair-cas.org.hk (S.W.); qingxiang.zhao@cair-cas.org.hk (Q.Z.); wei.huang@cair-cas.org.hk (W.H.); yihe.wang@cair-cas.org.hk (Y.W.)

⁴ Department of Biomedical Engineering, City University of Hong Kong, Hong Kong; mingcong.chen@my.cityu.edu.hk

⁵ School of Biomedical Engineering and Imaging Sciences, King's College London, London SE1 7EU, UK

* Correspondence: liuhongbin@ia.ac.cn

[†] This paper is an extended version of our paper published in the 45th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC 2023), Sydney, Australia, 24–27 July 2023.

Abstract: Quickly and accurately completing endoscopic submucosal dissection (ESD) operations within narrow lumens is currently challenging because of the environment's high flexibility, invisible collision, and natural tissue motion. This paper proposes a novel stereo visual servoing control for a dual-segment robotic endoscope (DSRE) for ESD surgery. Departing from conventional monocular-based methods, our DSRE leverages stereoscopic imaging to rapidly extract precise depth data, enabling quicker controller convergence and enhanced surgical accuracy. The system's dual-segment configuration enables agile maneuverability around lesions, while its compliant structure ensures adaptability within the surgical environment. The implemented stereo visual servo controller uses image features for real-time feedback and dynamically updates gain coefficients, facilitating rapid convergence to the target. In visual servoing experiments, the controller demonstrated strong performance across various tasks. Even when subjected to unknown external forces, the controller maintained robust performance in target tracking. The feasibility and effectiveness of the DSRE were further verified through ex vivo experiments. We posit that this novel system holds significant potential for clinical application in ESD surgeries.

Keywords: surgical robot; stereo visual servoing control; endoscopic submucosal dissection; soft robot



Citation: Chen, J.; Wang, S.; Zhao, Q.; Huang, W.; Chen, M.; Hu, J.; Wang, Y.; Liu, H. Stereo Visual Servoing Control of a Soft Endoscope for Upper Gastrointestinal Endoscopic Submucosal Dissection. *Micromachines* **2024**, *15*, 276. <https://doi.org/10.3390/mi15020276>

Academic Editor: Kian Jafari

Received: 4 January 2024

Revised: 7 February 2024

Accepted: 14 February 2024

Published: 15 February 2024



Copyright: © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Gastrointestinal (GI) cancer is a prevalent and highly malignant tumor in clinical practice [1]. Stomach cancer, in particular, ranks third in terms of mortality due to its low survival rate, with 782,685 deaths reported in 2018 [2]. Improving the survival rate of GI cancer relies heavily on early-stage treatment [3,4]. Owing to the high compliance, control precision, and cost-effectiveness, flexible endoscopes are increasingly being utilized in a variety of minimally invasive surgical procedures (see Figure 1a). Since the first application in 1988 [5], endoscopic submucosal dissection (ESD) has become the gold standard for the early treatment of GI cancer [6]. ESD enables the en bloc resection of submucosal lesions, resulting in a lower recurrence rate [7]. Currently, most ESD procedures are dependent on systems that utilize a single flexible arm outfitted with a monocular endoscope and various electrosurgical knives designed for specific surgical tasks [8]. However, due to the limited degrees of freedom (DoFs) of existing ESD endoscopes and the high skills

required, improper operations during the procedure may lead to some complications such as tissue perforation and bleeding [9]. Additionally, the narrow visual field offered by traditional endoscopes hinders the surgeon's ability to perform precise, multi-angle excisions, emphasizing the necessity for an ESD system with improved visual capabilities and instrument control [10].

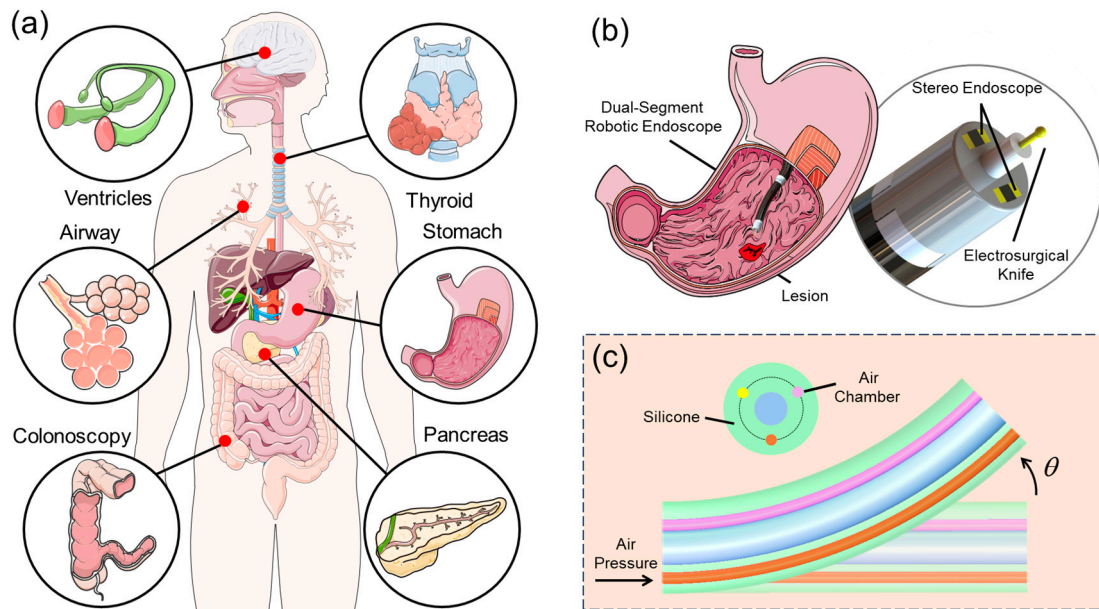


Figure 1. (a) Flexible endoscopes provide effective minimally invasive treatment through different naturally narrow orifices in the human body. (b) DSRE promotes dexterity and flexibility for ESD, and stereo vision contributes to evaluation before and after surgery. (c) Schematic diagram of the DSRE silicone-body soft segment, which would bend when the air pressure in the chamber increases.

To simplify the procedure and improve maneuverability, researchers have proposed different robot arm designs by adapting conventional surgical tools, enabling novices to perform the surgery easily. Chiu et al. [11] compared the feasibility and proficiency of ESD using an insulation-tipped (IT) diathermic knife, demonstrating that the robotic ESD device MASTER is more effective in terms of time cost and exhibits a good grasping and cutting efficiency [12]. Other similar ESD devices include ESD+AWC [13], K-FLEX [14], and TURBT [15]. However, the monocular vision system on traditional ESD endoscopes limits depth perception and lacks the stereoscopic perception of the lesion, which may increase the surgical operation time [16] and compromise the accuracy of resection [17,18]. The application of stereoscopic vision in various miniaturized surgical instruments, such as the da Vinci surgical system [19] and Versius surgical robot [20], enhances surgeons' hand-eye coordination and depth judgment capabilities in minimally invasive surgeries. Figure 1b illustrates a dual-segment robotic endoscope (DSRE) [21] with a stereo vision system, which is used to reach the stomach lesion through the esophagus and perform the excision. As shown in Figure 1c, the DSRE consists of two air-driven soft segments that can be actively bent by tuning the air pressure.

Challenges remain in controlling the end effector to achieve a desired pose in a flexible continuum robot arm, especially given the hyper-redundancy and infinite DoFs [22]. Contact with the target also introduces disturbances at the tip and the circumferential body. Therefore, accurate modeling of the relationship between the actuation inputs, tip pose, and compensation for disturbances is essential. Although modeling in continuum robotics is a well-explored topic [23], the design of an eye-in-hand system like a flexible endoscope is not common. The stereo vision, embedded at the robot tip, contributes to image-based visual servoing (IBVS), enhancing the system's compatibility with the eye-in-hand setup [24]. Table 1 summarizes the related research on multi-segment soft robot control, highlighting

the superiority of visual servoing for robot operation in unstructured environments. The classical IBVS has been widely used to solve the tracking [25], shape control [26], and depth estimation [27] problems of flexible endoscopes. During endoscopic operations, external interference, such as inserting internal instruments, may lead to potential issues with the proportional controller. These issues could manifest as slow convergence [28] and decreased tracking performance [25]. To overcome these uncertainties, control algorithms such as sliding mode control [29], model predictive control [30], and multi-sensor fusion [31] are deployed in the IBVS framework to improve system robustness. The camera not only provides a view of the lesion but also senses the actual motion of the robot arm, serving as a proprioception mechanism for disturbance compensation. In [31], the depth of the endoscopic image is extracted for camera position adjustment in laparoscopic surgeries. In this study, stereo vision is employed to capture a wide view of the lesion and estimate the depth between the robot tip and the desired target, providing sensing information for disturbance compensation.

Table 1. Comparison of the performance of the DSRE with related robot system.

Study	Backbone	Actuator	Model	Controller	Time	Error
Zhao et al. [22]	Continuous	Pneumatic	PCC	Jacobian estimation	0.18 s	0.62 mm
Zhang et al. [31]	Rigid	-	Geometric	VS with optimization	-	15.4629 Pixel
Roshanfar et al. [32]	Continuous	Pneumatic	Cosserat	-	-	MAE 5.79%
Chen et al. [21]	Continuous	Pneumatic	PCC	Classic IBVS	13 iterations	RMSE 1.199 mm
Lau et al. [33]	Discrete	Cable	PCC	Feedforward	0.25 s	4°
Li et al. [34]	Continuous	Cable	PCC	MPC	-	-
Abdulhafiz et al. [35]	Discrete	Cable	CNN	CNN-based VS	85 iterations	SAD 0.058 mm
Greer et al. [36]	Continuous	sPAMs	PCC	Classic IBVS	-	<5% in 2 s
DSRE in this study	Continuous	Pneumatic	PCC	ASVS	0.19 s	13.51 Pixel

This study presents a dual-segment robotic endoscope with a control algorithm based on binocular visual servoing, and preliminarily discusses the application value of stereo vision in ESD surgery. To address the extant clinical challenges, the design requirements of the DSRE are introduced in detail in Section 2, and the fabrication procedure of the robot and the design of the pneumatic control unit are explained, respectively. Section 3 explains the modeling of the relationship between the actuation inputs of the robot and the tip pose, as well as the approach to sensing an object in 3D form. Additionally, Section 3 introduces an adaptive stereo visual servoing (ASVS)-based control scheme. The experiment results in Section 4 demonstrate the effectiveness of the robot and the proposed methods through controller tests and an ex vivo ESD trial. Finally, Section 5 concludes this work. Compared with a previously published conference paper [21], the main improvements in this expanded study include a smaller and more flexible robot for ESD with segments calibration, the ASVS controller with adaptive gain and PD control, and a more comprehensive experimental validation.

2. System Design

2.1. Robot Design

The DSRE is designed to function the same as the distal tip [37] on a traditional endoscope; thus, it was designed with the following three goals in mind: (i) to allow for multiple orientations along a resection path, as a single segment proves inadequate; (ii) to fit within the adult esophagus, which measures 25 cm in length and 2 cm in diameter [38]; and (iii) to facilitate a wide viewing angle during surgery and the efficient operation of a T-type electrosurgical cutting knife.

To this end, both the proximal and distal segment of the DSRE were designed to have three actuators longitudinally aligned with the main body and have a 2 mm through working channel for the electrosurgical knife or biopsy grasper, as shown in Figure 2b. When fabricated with appropriate pretension, this construction allows for approximately 170° and 88° of curvature when the proximal and distal are fully actuated, respectively.

More detailed dimensions of the DSRE are shown in Figure 2b. As shown in Table 2, compared with the previous design [21] and conventional endoscopes, the DSRE's robotic actuator can provide more precise control, while its smaller size and increased bending capabilities expand the operational area within the gastrointestinal tract. Additionally, the stereo vision provides supplementary depth information.

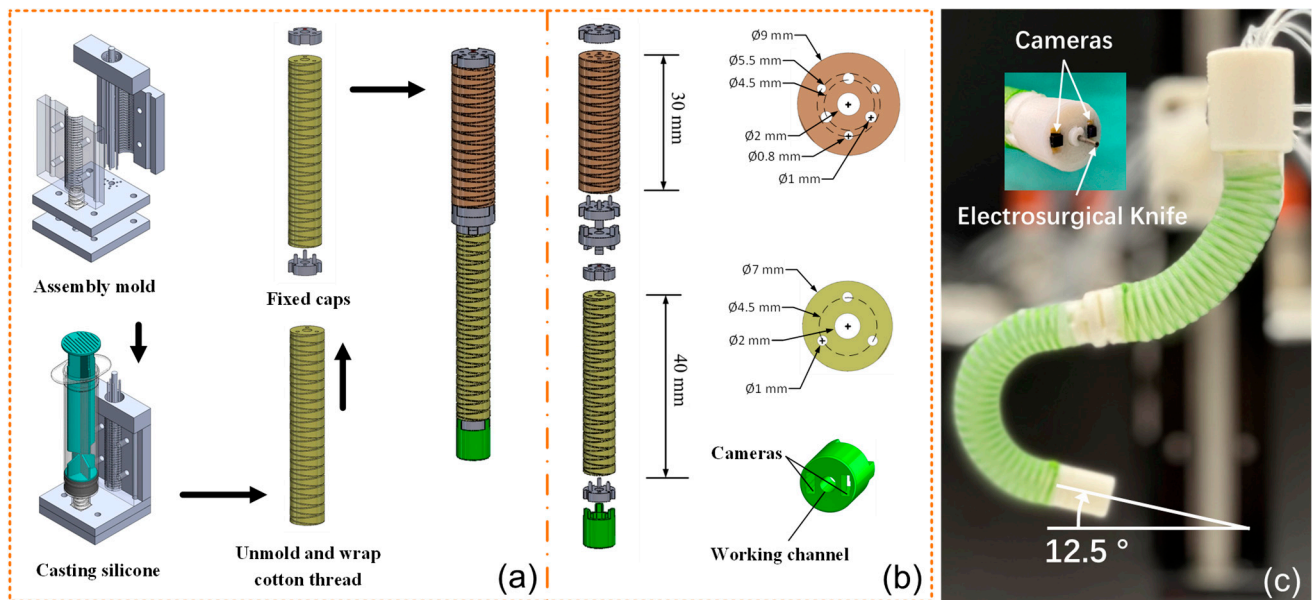


Figure 2. (a) Fabrication process of the DSRE. (b) Detailed dimension. (c) DSRE with two cameras and an electrosurgical knife.

Table 2. Comparison of the DSRE system with previous design [21] and conventional endoscopes.

Systems	Actuation Type	Segments	Body Length (mm)	Diameter (mm) ¹	Max. Angles (°)	Camera	Working Channel (mm)
DSRE	Robotized pneumatic	2	72	P 9, D 7	P170 + D88	2	2
Chen et al. [21]	Robotized pneumatic	2	85	P 10, D 8	P100 + D88	2	s2
CF-XZ1200L/I [39]	Manual cable-driven	1	-	13.2	180	1	3.7
EC-760S-V/L [40]	Manual cable-driven	1	-	12.8	180	1	3.8
EXALT Model D [41]	Manual cable-driven	1	-	15.1	120	1	4.2

¹ P and D denote proximal and distal, respectively.

Two commercial cameras (OV6946, OmniVision, Santa Clara, CA, USA) with a distribution distance of 4.4 mm were attached at the tip, providing a wide view angle during surgery, as shown in Figure 2c. Similarly, a T-type electrosurgical cutting knife was fixed at the tip, acting as the end effector of the robot arm. The resection power and mode could be set on an electrosurgical generator (DGD-300B-2, Beilin, Beijing, China).

2.2. Robot Fabrication

As shown in Figure 2a, by casting the prototypes on 3D-printed molds, the soft segments were constructed of two low-stiffness silicones, Ecoflex 0050 and Dragon Skin 10 AF (Smooth-On, Macungie, PA, USA) mixed in a 1:1 ratio. To restrict elongation and limit radial expansion when pressurized, a nylon thread was attached at the end of each segment and another thread was coiled spirally. The DSRE of two serially connected segments was then affixed to a linear stage to facilitate insertion into the stomach.

2.3. Pneumatic Control Design

Pneumatic actuation is used in the DSRE system for its high precision control, quick response, safety for medical use, and the adaptability of in-body applications. Each air chamber of the robotic arm was linked to a pneumatic regulator (ITV0030, SMC, Tokyo, Japan) by a 60 cm silicone tube (0.3 mm inner, 0.8 mm outer diameter), and command individual pressures continuously to bend the DSRE with a direction angle and a bending angle. These valves were chosen for the following reasons: fast response times (0.1 s with no load), high stability (within 0.5% full span repeatability error), and accurate response (within 1% full span linearity error).

2.4. Robot Calibration

The calibration of soft robots is an essential process that helps determine the mechanical behavior under different loads and conditions, enabling better control and manipulation. To verify the mechanical performance and chamber consistency of the proposed dual-segment soft robot, we initially examined its bending behavior corresponding to each independent air chamber.

Figure 3a illustrates the configurations of the electromagnetic (EM) trackers located at the proximal and distal segments of the DSRE. Notably, to accurately obtain the position and orientation of each segment center, two trackers were symmetrically affixed to a 3D-printed part mounted at the proximal/distal segment's end. We subsequently processed the average of the sensed data to calculate the positions and Euler angles. During testing, the pneumatic regulator was controlled to gradually increase the air pressure with a constant increment of 5000 Pa, while simultaneously recording the bending angles. As shown in Figure 3b, the curves of the bending angles corresponding to each segment's three chambers were roughly equivalent. When reaching the maximum air pressure (0.14 MPa for the proximal segment and 0.1 MPa for the distal segment), the standard deviations of the bending angles across the different chambers were remarkably low, which were only 0.1012 rad and 0.0428 rad, respectively. These results illustrate that the soft robot has a consistent and stable bending performance in all directions.

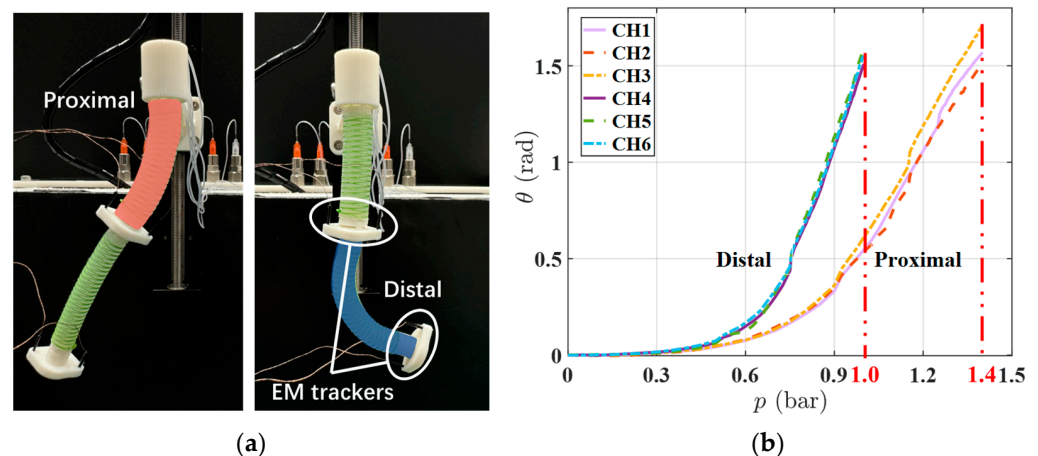


Figure 3. (a) Configuration of the EM trackers on proximal and distal segments. (b) Bending angles of proximal and distal segments corresponding to the air pressure, the red lines correspond to the maximum air pressure exerted on each segment.

3. Methodology

Through geometric analysis and an optimization algorithm, the modeling was first established, and the two endoscopes placed in an embedded sensing system to compensate for the external load effect.

3.1. Kinematics

3.1.1. Shape to Tip Pose

For soft segments with a large slenderness ratio, when the eccentrically fixed chamber is subjected to changing air pressure, the robot will bend and deform. Referring to the PCC model proposed in [42], it is assumed that the bending shape of each soft robot on the DSRE is a circular arc. As shown in Figure 4a, the DSRE can be geometrically parameterized by $\Phi = (z_s \ \theta_s \ \varphi_s \ \theta_e \ \varphi_e)^T$ in the configuration space, where z_s is the overall insert distance provided by a linear stage, θ_s and θ_e are the bending angles, φ_s and φ_e are the direction angles between the bending plane and the oxz plane, l_s and l_e are the length of the proximal and distal segment, and z_e and t_d are the length of the middle cap and tip cap.

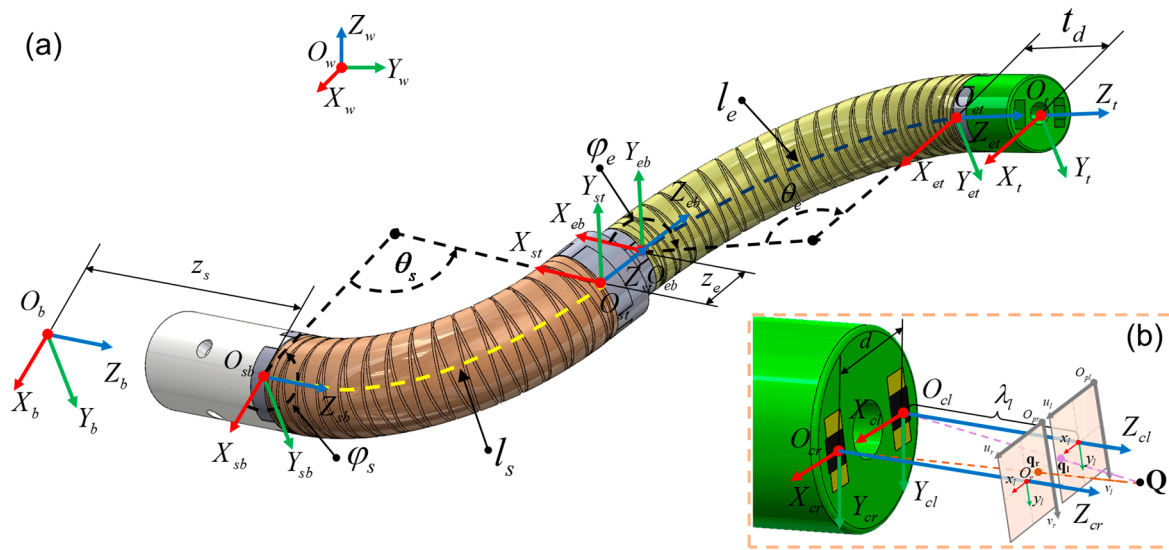


Figure 4. (a) Illustration of coordinate frames. (b) The binocular stereo vision on the tip.

Taking the proximal segment as an example, the coordinate transformation sequence from the $O_{sb} X_{sb} Y_{sb} Z_{sb}$ to $O_{st} X_{st} Y_{st} Z_{st}$ coordinate system is as follows: first rotate the φ_s angle around the z -axis, then translate l_s/θ_s along the x -axis, then rotate the θ_s angle around the y -axis, and, finally, translate l_s/θ_s along the negative direction of the x -axis. The corresponding homogeneous transformation matrix is calculated as follows:

$${}_{sb}^{st}\mathbf{T} = \mathbf{R}_z(\varphi_s) \boldsymbol{\tau}_x\left(\frac{l_s}{\theta_s}\right) \mathbf{R}_y(\theta_s) \boldsymbol{\tau}_x\left(-\frac{l_s}{\theta_s}\right) \quad (1)$$

where $\boldsymbol{\tau}_*$, $\mathbf{R}_* \in \mathbb{R}^{4 \times 4}$, respectively, denote the translation and rotation about the $\{x, y, z\}$ axis.

Considering the proximal segment and the displacements z_s , z_e , and t_d , the overall transformation matrix ${}^b_t\mathbf{T} \in \mathbb{R}^{4 \times 4}$ from the base frame $O_b X_b Y_b Z_b$ to the robot tip frame $O_t X_t Y_t Z_t$ of the DSRE system is as follows:

$${}^b_t\mathbf{T} = \boldsymbol{\tau}_z(z_s) \mathbf{R}_z(\varphi_s) \boldsymbol{\tau}_x\left(\frac{l_s}{\theta_s}\right) \mathbf{R}_y(\theta_s) \boldsymbol{\tau}_x\left(-\frac{l_s}{\theta_s}\right) \mathbf{R}_z\left(\frac{\pi}{3} + \varphi_e\right) \boldsymbol{\tau}_z(z_e) \mathbf{R}_z(\varphi_e) \boldsymbol{\tau}_x\left(\frac{l_e}{\theta_e}\right) \mathbf{R}_y(\theta_e) \boldsymbol{\tau}_x\left(-\frac{l_e}{\theta_e}\right) \boldsymbol{\tau}_z(t_d) \quad (2)$$

where the shift angle between the two segments is 60°

The Jacobian matrix $\mathbf{J}_r \in \mathbb{R}^{6 \times 5}$ is used to analytically establish the approximate relationship between the robot tip velocity $\mathbf{v}_t \in \mathbb{R}^6$ and joint velocity $\dot{\Phi}$:

$$\mathbf{v}_t = \mathbf{J}_r \dot{\Phi} \quad (3)$$

where $\mathbf{J}_r$ can be derived through forward kinematics ${}^b_t\mathbf{T}$. A more detailed calculation of the forward kinematics and the Jacobian matrix can be found in Appendix A.

To reach a given target position of the end of the robot, we need to inversely solve the appropriate joint configuration.

3.1.2. Air Pressure to Shape

After finding the shape parameters, the robot system should set proper air pressures for each segment. The joint angles θ_s , θ_e , φ_s , and φ_e can be calculated from the actuator space variables $\mathbf{q} = (z_s \ p_{s,1} \ p_{s,2} \ p_{s,3} \ p_{e,1} \ p_{e,2} \ p_{e,3})^T$:

$$\begin{aligned}\theta_i &= \frac{2\sqrt{p_{i,1}^2 + p_{i,2}^2 + p_{i,3}^2 - p_{i1}p_{i2} - p_{i2}p_{i3} - p_{i1}p_{i3}}}{3\rho_i} \\ \varphi_i &= \tan^{-1}\left(\frac{p_{i1} + p_{i3} - 2p_{i2}}{\sqrt{3}(p_{i3} - p_{i1})}\right)\end{aligned}\quad (4)$$

where $i \in \{e, s\}$, and the subscripts e and s are used to represent the proximal and distal segment, respectively, ρ_i is the distance between the center of the air chamber and the center of the segment, $p_{i,m}$, $m \in \{1, 2, 3\}$ are the air pressure of the chambers in each segment.

3.2. Depth Estimation

The two cameras mounted at the robot tip could not only view the lesion with a wide angle but also could be employed to calculate the depth of a target (vertical distance from camera to the target). In this study, we employed the semi-global matching (SGM) algorithm [43], renowned for its efficacy in disparity estimation. From the disparity and stereo camera intrinsic information, the targets' 3D position relative to camera frames can be estimated. The SGM algorithm encompasses a pixelwise matching technique based on mutual information (MI) [44] and a semi-global 2D constraint, composed of multiple 1D constraints. It is detailed as follows.

3.2.1. Mutual Information-Based Cost function

Initially, SGM computes the similarity between pixels in two rectified stereo images using a mutual information (MI)-based cost function, which is less sensitive to photometric changes. Thus, the MI value can be used to evaluate the suitability of pixels. The higher the value, the better the match, and vice versa. For the left and right images $\mathbf{I}_l$, $\mathbf{I}_r$ from the end effector of the DSRE, the MI value is $\mathbf{MI}_{\mathbf{I}_l, \mathbf{I}_r} = \mathbf{H}_{\mathbf{I}_l} + \mathbf{H}_{\mathbf{I}_r} - \mathbf{H}_{\mathbf{I}_l, \mathbf{I}_r}$ ($\mathbf{H}$ is the entropy of image, $\mathbf{H}_{\mathbf{I}_l, \mathbf{I}_r}$ is the joint entropy of images). The MI-based cost function would be computed for each pixel and disparity, resulting in a cost volume that the SGM method would use in subsequent steps to compute the final disparity map.

3.2.2. Aggregate Cost

After the initial cost computation, these costs need to be aggregated across multiple paths through the image to enforce the smoothness constraints while preserving disparity discontinuities, which are often present at object boundaries. A path-wise dynamic programming approach was applied in the SGM algorithm. The cost aggregation step involves accumulating the costs for each pixel along several paths through the image. Typically, 8 paths are considered to cover all directions: left-to-right, right-to-left, top-to-bottom, bottom-to-top, and the four diagonals.

For a given pixel $\mathbf{h}$ and disparity d , the aggregated cost $E_s(\mathbf{h}, d)$ along a path $\mathbf{s}$ is calculated as follows:

$$E_s(\mathbf{h}, d) = C(\mathbf{h}, d) + \min \left\{ \begin{array}{l} E_s(\mathbf{h} - \mathbf{s}, d) \\ E_s(\mathbf{h} - \mathbf{s}, d - 1) + K_1 \\ E_s(\mathbf{h} - \mathbf{s}, d + 1) + K_1 \\ \min_i E_s(\mathbf{h} - \mathbf{s}, i) + K_2 \end{array} \right\} - \min_k E_s(\mathbf{h} - \mathbf{s}, k) \quad (5)$$

where $C(\mathbf{h}, d)$ is the matching cost at pixel $\mathbf{h}$ for disparity d computed from the mutual information, $E_s(\mathbf{h} - \mathbf{s}, d)$ is the cost for pixel $\mathbf{h}$ at disparity d accumulated from the previous

pixel along path $\mathbf{s}$, K_1 and K_2 are the penalty for small disparity changes and for large disparity changes, respectively, to allow for disparity discontinuities. The last term is the minimum cost for the previous pixel $\mathbf{h} - \mathbf{s}$ over all disparities. After the cost aggregation step, each pixel will have 8 different accumulated costs, 1 for each path. These costs are then combined to produce a single, aggregated cost for each pixel and disparity:

$$E(\mathbf{h}, d) = \sum_{\mathbf{s}} E_{\mathbf{s}}(\mathbf{h}, d) \quad (6)$$

The final disparity for each pixel is chosen as the disparity that minimizes this aggregated cost.

3.2.3. Disparity Refinement

Once the initial disparity map is computed, it often requires further refinement to improve the accuracy, especially in occluded or low-texture areas. Refinement steps may include subpixel enhancement, occlusion handling, disparity map filtering, and left-right consistency check. Figure 5 illustrates the application of the SGM algorithm on the DSRE system, which was conducted in the MATLAB R2022a (Mathworks Inc. Portola Valley, CA, USA) with the disparity SGM and reconstruct Scene function. The depth information would be further used in the interaction matrix $\mathbf{L}$ in the ASVS controller. This matrix is essential for providing accurate and reliable positional information regarding the target.

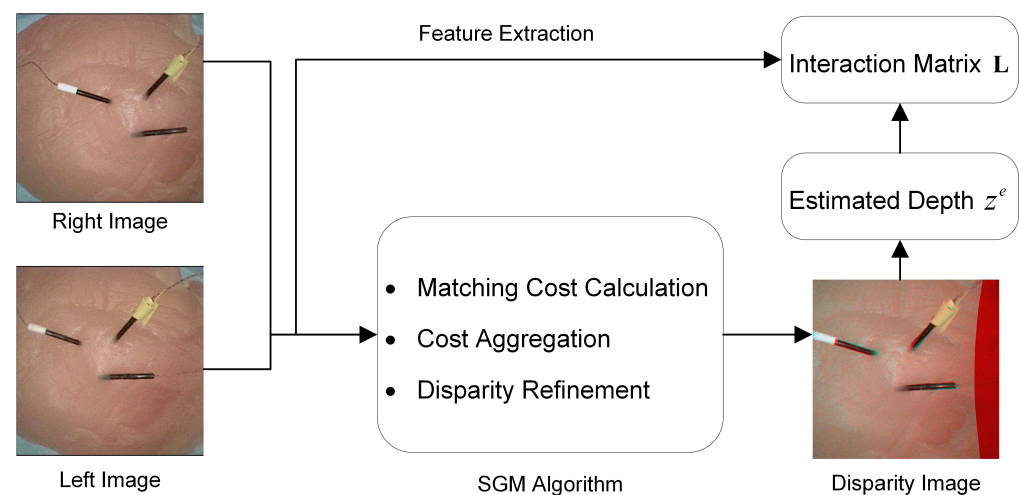


Figure 5. The application of the SGM algorithm on the DSRE system.

3.3. Adaptive Stereo Visual Servoing Control

The visual feedback is directly utilized in the control loop of IBVS. The goal is to match the current image with the desired image. The error signal $\mathbf{e} \in \mathbb{R}^2$ for the controller is derived from the difference between the current image features $\mathbf{s} \in \mathbb{R}^2$ and the desired image features $\mathbf{s}_d$, and $\dot{\mathbf{e}} = \dot{\mathbf{s}}$.

3.3.1. Monocular Vision Model

Take the left camera as an example to demonstrate the camera model for brevity. The midpoint of the connection line between the two cameras coincides with O_t .

As shown in Figure 4b, as a given point $\mathbf{Q} \in \mathbb{R}^3$ in $O_{cl}X_{cl}Y_{cl}Z_{cl}$, its coordinates in the image frame $O_lx_ly_l$ and pixel frame $O_{pl}uv$ are $\mathbf{q}_l = (x_l \ y_l)^T$ and $\mathbf{s}_l = (u_l \ v_l)^T$, respectively. According to the pinhole camera model, the perspective equation can be obtained from the relationship on similar triangles, i.e., as follows:

$$\begin{aligned} u_l &= \frac{\lambda_{xl}x_l}{\lambda_l} + c_{xl} \\ v_l &= \frac{\lambda_{yl}y_l}{\lambda_l} + c_{yl} \end{aligned} \quad (7)$$

where λ_{xl} , λ_{yl} are the focal length in pixels, c_{xl} , c_{yl} are the optical center in pixels, and λ_l is the focal length in millimeters.

The motion of the features $\dot{\mathbf{s}}_l$ on the image plane can be predicted using the interaction matrix and velocity of camera $\mathbf{v}_{cl} \in \mathbb{R}^6$:

$$\dot{\mathbf{s}}_l = \mathbf{L}_l \mathbf{v}_{cl} \quad (8)$$

with (7), $\mathbf{L}_l \in \mathbb{R}^{2 \times 6}$ is given by

$$\mathbf{L}_l = \begin{bmatrix} -\frac{\lambda_l}{z_{cl}^e} & 0 & \frac{x_l}{z_{cl}^e} & \frac{x_l y_l}{\lambda_l} & -\frac{\lambda_l^2 + x_l^2}{\lambda_l} & y_l \\ 0 & -\frac{\lambda_l}{z_{cl}^e} & \frac{y_l}{z_{cl}^e} & \frac{\lambda_l^2 + y_l^2}{\lambda_l} & -\frac{x_l y_l}{\lambda_l} & -x_l \end{bmatrix} \quad (9)$$

where z_{cl}^e is the depth of point $\mathbf{Q}$, which can be estimated online with the stereo vision. Similarly, we can obtain the right camera model with $\dot{\mathbf{s}}_r = \mathbf{L}_r \mathbf{v}_{cr}$.

3.3.2. Stereo Vision Model

Considering the distance d between cameras, the transformation matrix ${}^d_l \mathbf{T} \in \mathbb{R}^{4 \times 4}$ from the left camera frame to the robot tip frame is

$${}^d_l \mathbf{T} = \boldsymbol{\tau}_x(d/2) = \begin{bmatrix} \mathbf{I}_3 & {}^d_l \mathbf{p} \\ 0 & 1 \end{bmatrix}, \quad (10)$$

while the transformation matrix for the right camera is ${}^d_r \mathbf{T} = \boldsymbol{\tau}_x(-d/2)$.

Since the left–right camera is fixed on the tip of the DSRE, the velocity transformation between the camera frame and robot tip frame is given by

$$\begin{aligned} \mathbf{v}_{cl} &= {}^d_l \mathbf{M} \mathbf{v}_t \\ \mathbf{v}_{cr} &= {}^d_r \mathbf{M} \mathbf{v}_t \end{aligned} \quad (11)$$

where the transformation matrix ${}^d_l \mathbf{M} \in \mathbb{R}^{6 \times 6}$ can be calculated from [45] with (10), i.e.,

$${}^d_l \mathbf{M} = \begin{bmatrix} \mathbf{I}_3 & [{}^d_l \mathbf{p}]_{\times} \\ \mathbf{O}_3 & \mathbf{I}_3 \end{bmatrix} \quad (12)$$

where $[{}^d_l \mathbf{p}]_{\times}$ is the skew-symmetric matrix of ${}^d_l \mathbf{p}$.

Let $\dot{\mathbf{s}} = (\dot{\mathbf{s}}_l \quad \dot{\mathbf{s}}_r)^T$ represent the image features motion extracted from the left and right cameras simultaneously, and the overall velocity transformation from the robot tip frame to the pixel frame can be derived from (8), (11), and (12):

$$\dot{\mathbf{s}} = \begin{bmatrix} \mathbf{L}_l & {}^d_l \mathbf{M} \\ \mathbf{L}_r & {}^d_r \mathbf{M} \end{bmatrix} \mathbf{v}_t = \mathbf{L} \mathbf{v}_t \quad (13)$$

where $\mathbf{L} \in \mathbb{R}^{4 \times 6}$ is the overall interaction matrix of the stereo vision system.

Inserting (3) in (13), we obtain the relationship of the joint velocity and image feature:

$$\dot{\mathbf{e}} = \mathbf{L} \mathbf{J}_r \dot{\boldsymbol{\Phi}} \quad (14)$$

$\mathbf{L} \mathbf{J}_r \in \mathbb{R}^{4 \times 5}$ is a singular matrix without an inverse matrix. Commonly used pseudo-inverse methods often have stability problems. When the robot approaches a singular configuration, this method will cause excessive joint velocities and cause safety problems. The damped least squares method [46] not only improves the image feature tracking effect, but also avoids the singularity problem by limiting joint velocities

$$\min \left(\|\mathbf{L} \mathbf{J}_r \dot{\boldsymbol{\Phi}} - \dot{\mathbf{e}}\|^2 + \sigma \|\dot{\boldsymbol{\Phi}}\|^2 \right) \quad (15)$$

where $\sigma \in \mathbb{R}$ is a non-zero damping constant.

The inverse problem could be solved by minimizing the object (14). The joint velocities could be derived further as

$$\begin{aligned}\dot{\Phi} &= \mathbf{L}_a^+ \dot{\mathbf{e}} \\ \mathbf{L}_a^+ &= (\mathbf{L}_r)^T (\mathbf{L}_r (\mathbf{L}_r)^T + \sigma \mathbf{I})^{-1}\end{aligned}\quad (16)$$

3.3.3. Adaptive PD Controller

The classic IBVS tasks usually use a proportional control law [28]. Let the error of the feature satisfy the first-order equation $\dot{\mathbf{e}} = -\eta \mathbf{e}$, so that the error value can decrease rapidly in an exponential speed. Thus,

$$\dot{\Phi} = -\eta \mathbf{L}_a^+ \mathbf{e} \quad (17)$$

Although a proportional controller responds quickly and is simple to deploy, it cannot eliminate the steady-state error (SSE). In this paper, we adopt a PD controller to reduce the SSE and increase the system stability. Adding a derivative term to (17), we obtain

$$\dot{\Phi} = -\eta \mathbf{L}_a^+ \mathbf{e} - \mathbf{K}_d \dot{\mathbf{e}} \quad (18)$$

where $\mathbf{K}_d$ is the non-negative derivative gain.

To enhance the efficiency of the IBVS convergence process, we propose the implementation of an adaptive proportional gain. η is dynamically adjusted in real time based on the infinity norm of the feature error, aiming to reduce the number of iterations necessary for convergence and enhance the system's performance. η is set as follows:

$$\eta = (\eta_u - \eta_l) \frac{\|\mathbf{e}\|_\infty}{\|\mathbf{e}_0\|_\infty} + \eta_l \quad (19)$$

where η_u and η_l are the preset maximum and minimum gains, respectively, $\mathbf{e}_0$ is the initial error to a new target. The detailed flow chart of the ASVS controller is shown in Figure 6.

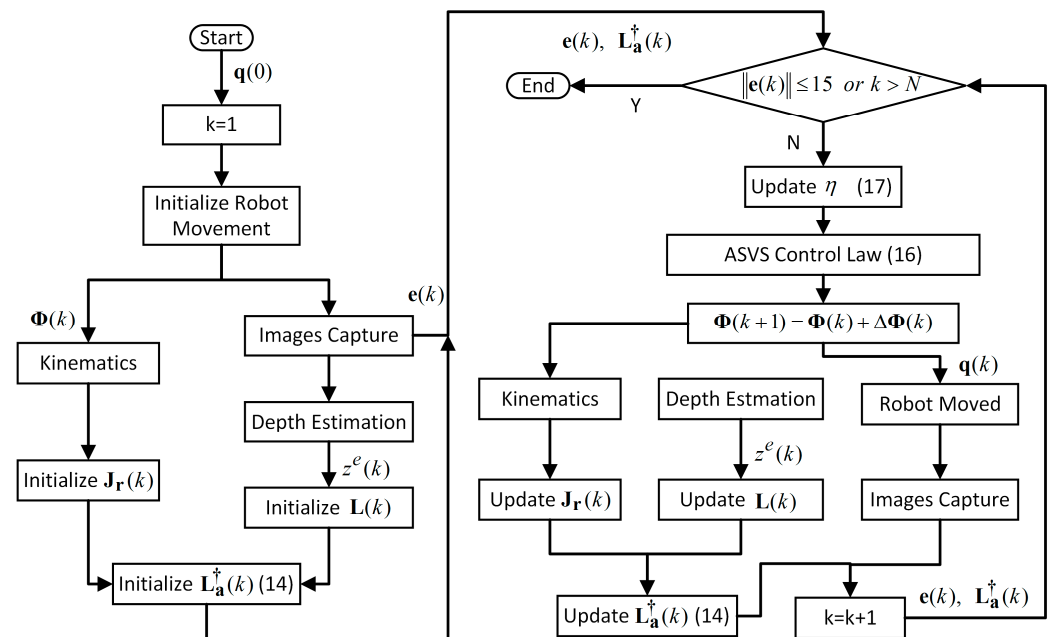


Figure 6. Flow chart of the ASVS controller.

4. Experimental Validation

4.1. Depth Estimation Validation

To validate the efficacy of the proposed camera depth estimation algorithm, a comparison was made between the ground truth, derived from EM trackers, model 10001742 (NDI Aurora, Waterloo, ON, Canada), and the values estimated by the algorithm. The specific setup used for the testing process is illustrated in Figure 7a. Three trackers were fixed on a piece of tissue while two base trackers were mounted on the tip of the distal segment. For this experiment, the ground truth was defined as the distance between the base tracker and the trackers placed within the task space. The stereo camera in Figure 7b was first well calibrated with the method in [47], resulting in a notably low mean reprojection error of merely 0.2 pixels. A comparison was then made between the estimated depth, as calculated by the SGM, and the actual values across different altitudes. As depicted in the error band curves in Figure 8, the RMSE of the estimated depth parameters of different markers were 1.5334 mm, 1.5449 mm, and 1.5468 mm, respectively, when compared with the ground truth. This illustrates that the proposed algorithm is highly accurate, and it could obtain reliable depth estimations in various surgical applications. The estimated depth was further used in the visual servoing control.

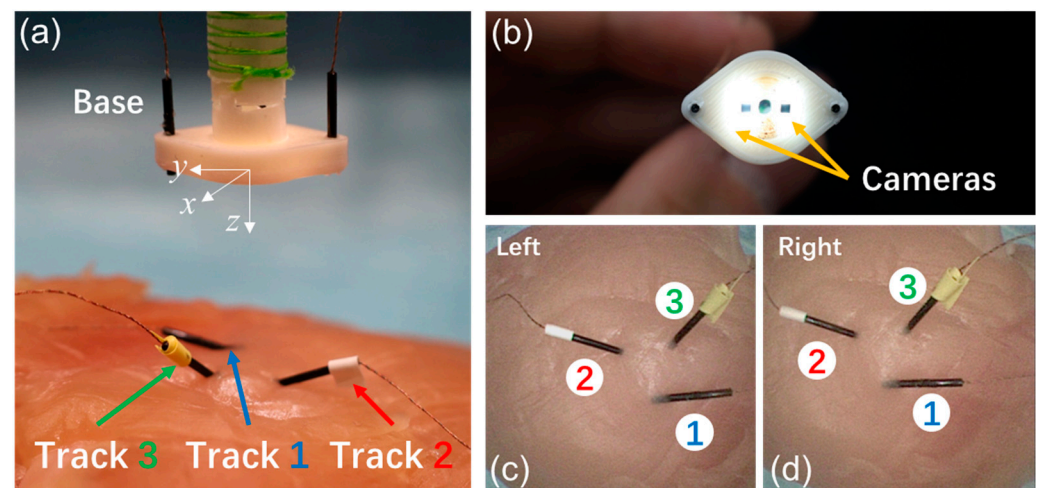


Figure 7. Depth estimation validation setup. (a) A base tracker was used to measure the ground truth of depth. (b) The stereo vision on the tip of DSRE. (c) Image of left camera. (d) Image of right camera.

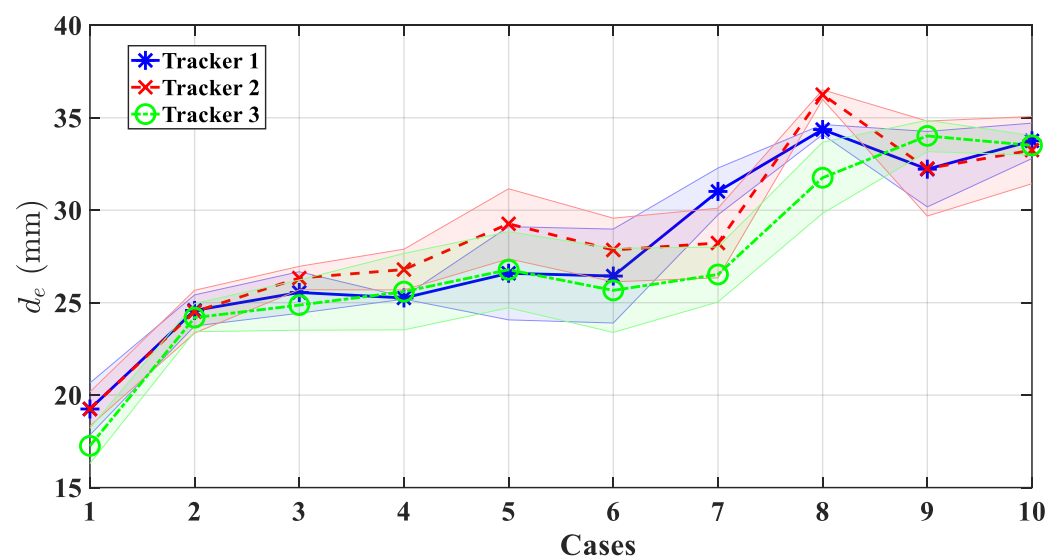


Figure 8. Comparative error band analysis of the camera depth estimation results.

4.2. Controller Performance

The performance of the ASVS was evaluated through the execution of four distinct manipulation tasks in this section. The visual servo controller's parameters during these tests were established as follows: $\mathbf{q}_0 = (0 \ 0.2 \ 0.2 \ 0.2 \ 0.2 \ 0.2 \ 0.2)^T$, $\eta_u = 1$, $\eta_l = 0.5$, $\sigma = 0.5$, $\mathbf{K}_d = \text{diag}\{0.02, 0.06, 0.02, 0.06\}^T$. In this study, the tracking error was quantified by the Euclidean distance between the current coordinates and target coordinates of the extracted image feature in the left and right frames. The control goal was to stabilize the error below 15 pixels. The frame rate of the left and right cameras was 30 FPS to ensure that the calculated air pressures have been applied to the segments to generate the desired robot shape.

As indicated in Figure 9, to simplify the feature detection and extraction process, the stereo vision system was programmed to track the AprilTags [48] in the workspace. Tags with ID 1, 2, 8, and 9 were selected as the tracking targets. The soft robot was controlled to track the target tags to the desired position on the image plane under the air pressure from the pneumatic regulators. A straight beam equipped with a force sensor was fixed on a linear rail guide slide actuator, and the tip caused contact disturbance to the robot. More results can be found in the Supplementary Video S1.

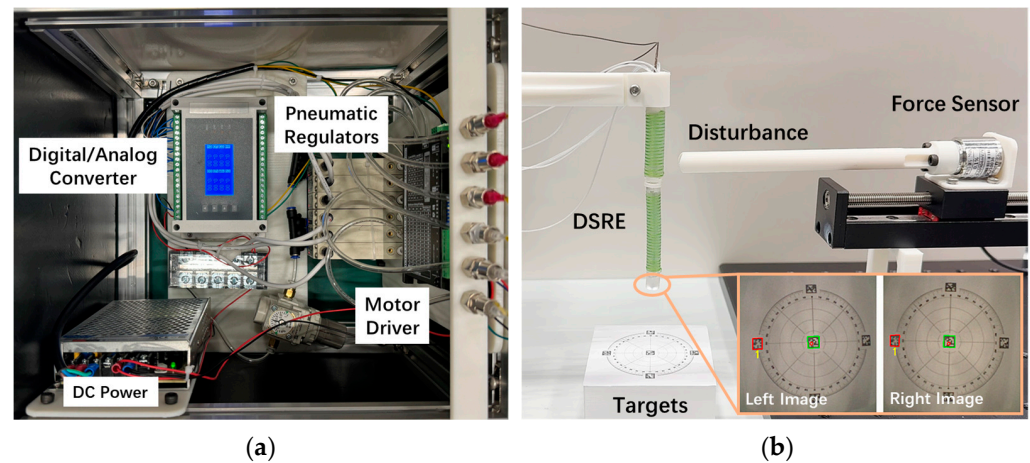


Figure 9. (a) The actuator configuration. (b) Experimental setup for the ASVS controller performance verification.

4.2.1. Static Target Tracking

In this experiment, the regions of interest (ROIs) within the left and right images were delineated as circular areas centered at points $\mathbf{s}_{dl} = (210 \ 200)^T$ and $\mathbf{s}_{dr} = (190 \ 200)^T$ with a radius of 15 pixels, as shown by the red circle and the green dashed circle in Figure 10b. The initial orientation of the DSRE robot was positioned to face the initial target with the AprilTag ID 3. After initialization, the robot was controlled to bring tags with different IDs to the ROI areas. These tags were equidistantly distributed at 90-degree intervals along the circumference of a circle. In each static tag tracking experiment, the adaptively adjusted gains of the ASVS controller were compared with the classic visual servoing method [28] employing fixed gains of $\eta = 0.5$, $\eta = 0.75$, and $\eta = 1$. Figure 10b illustrates the trajectories of various tags in the left and right images, and all ultimately converged within the ROIs. Due to the inherent limitation of robots with inconsistent stiffness in all directions, the DSRE cannot reach the target at some positions, and it will overshoot after passing through these positions, which resulted in the controller's tracking effect on tag 1 being inferior to several others, as depicted by the blue curve in Figure 10b. Figure 10c and Table 3 detail the pixel errors and convergence times for tracking different tags, respectively. As indicated by (17), when the gain was low, the robot's end effector moved at a reduced velocity, resulting in excessive time consumption. Conversely, with high gain, the end effector's velocity increased, potentially leading to overshoot and

oscillation. With the gain set at constant values of 5, 0.75, and 1, the classic controller required an average of 19.4 s, 8.29 s, and 9.04 s, respectively, to achieve the tracking of a static target, reducing the error to below 15 pixels. Meanwhile, the proposed control algorithm necessitated only 4.99 s, corresponding to a reduction in time consumption of 74.28%, 39.81%, and 37.94%, respectively. This result illustrates the robustness and reliability of this method in achieving the accurate control of the soft robot arm and precise tracking of the moving target.

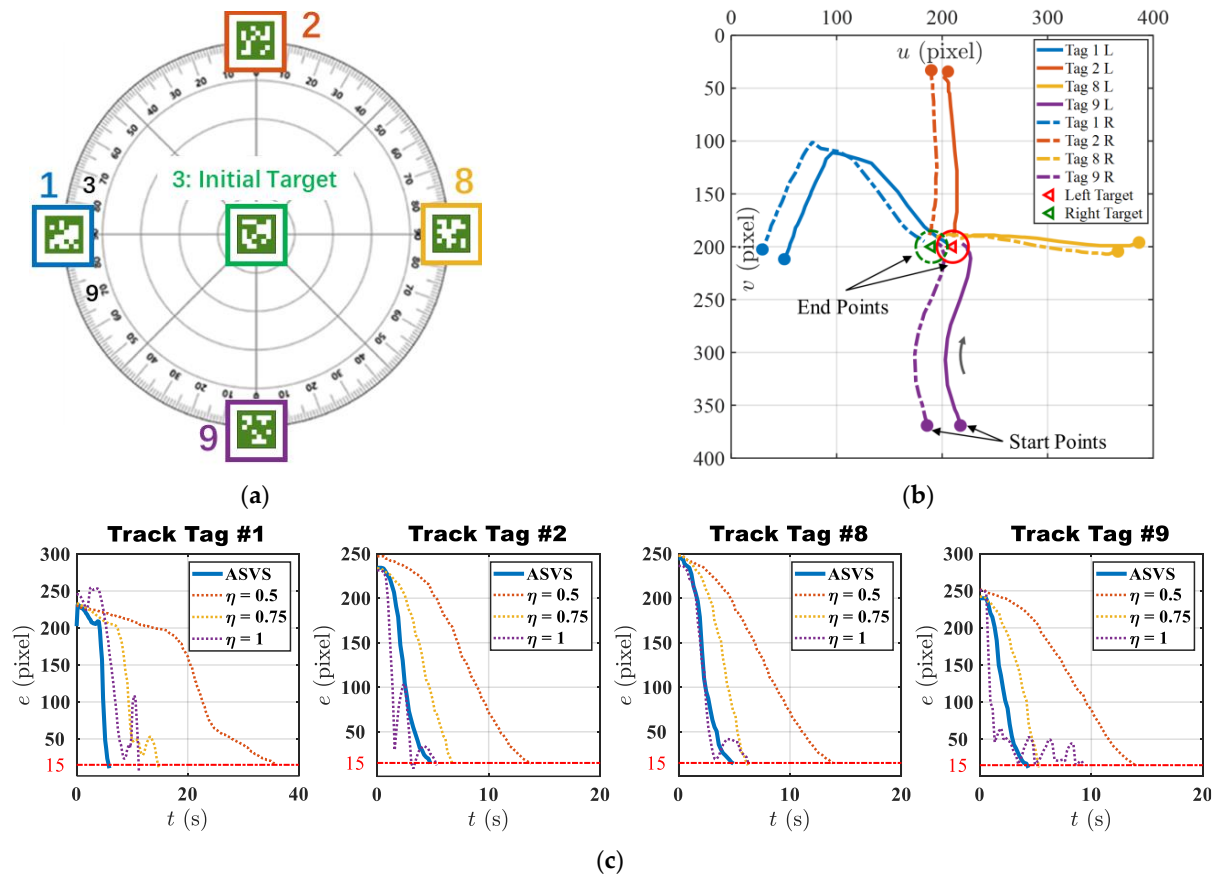


Figure 10. Static target tracking results. (a) Distribution configuration of target markers. (b) Target movement trajectories in the left and right images, the tags moved from the start points and converged to the end points. (c) Tracking errors of different controllers.

Table 3. Convergence times of ASVS and the classic control law with different constant η .

Controllers	Tag #1 (s)	Tag #2 (s)	Tag #8 (s)	Tag #9 (s)	Avg. (s)	Dec. (%)
ASVS	5.89	4.78	4.93	4.36	4.99	-
$\eta = 0.5$	35.87	13.80	13.87	14.05	19.40	74.28
$\eta = 0.75$	14.78	6.86	6.22	5.31	8.29	39.81
$\eta = 1$	11.21	5.33	6.29	9.33	8.04	37.94

4.2.2. Dynamic Target Tracking

This part introduces a tracking experiment designed to evaluate the control performance of the robot under dynamic target changes. As demonstrated in the previous experiment, comparisons were made between the ASVS and classical algorithms using different fixed gains. Figure 11a shows the experimental setup, with tags affixed to an electric linear slide rail, which moved back and forth between two points, A and B, 15 mm apart, at a specified velocity of $v_{\text{tag}} = 0.75 \text{ mm/s}$. The endoscopic images captured during the ex-

periment and the corresponding tracking errors are presented in Figure 11b,c, respectively. The robot was tasked to follow tag 9, while the linear guide remained stationary until the initial stable tracking of the target was accomplished by the robot. As shown in Figure 11b, a significant surge in error occurred when the tag's movement direction changed, but the controller quickly re-captured the tag and subsequently reduced the tracking difference to less than 15 pixels. Table 4 enumerates the tracking errors' root mean square error (RMSE), mean absolute value error (MAE), standard deviation (SD), and maximum error $e_{\max}$ for different controllers. The ASVS outperformed the classic controller across various metrics, maintaining the RMSE and MAE beneath 15 pixels, with an SD of merely 5.14 pixels. This rapid error correction is indicative of the agility and precision of the real-time tracking of dynamic targets.

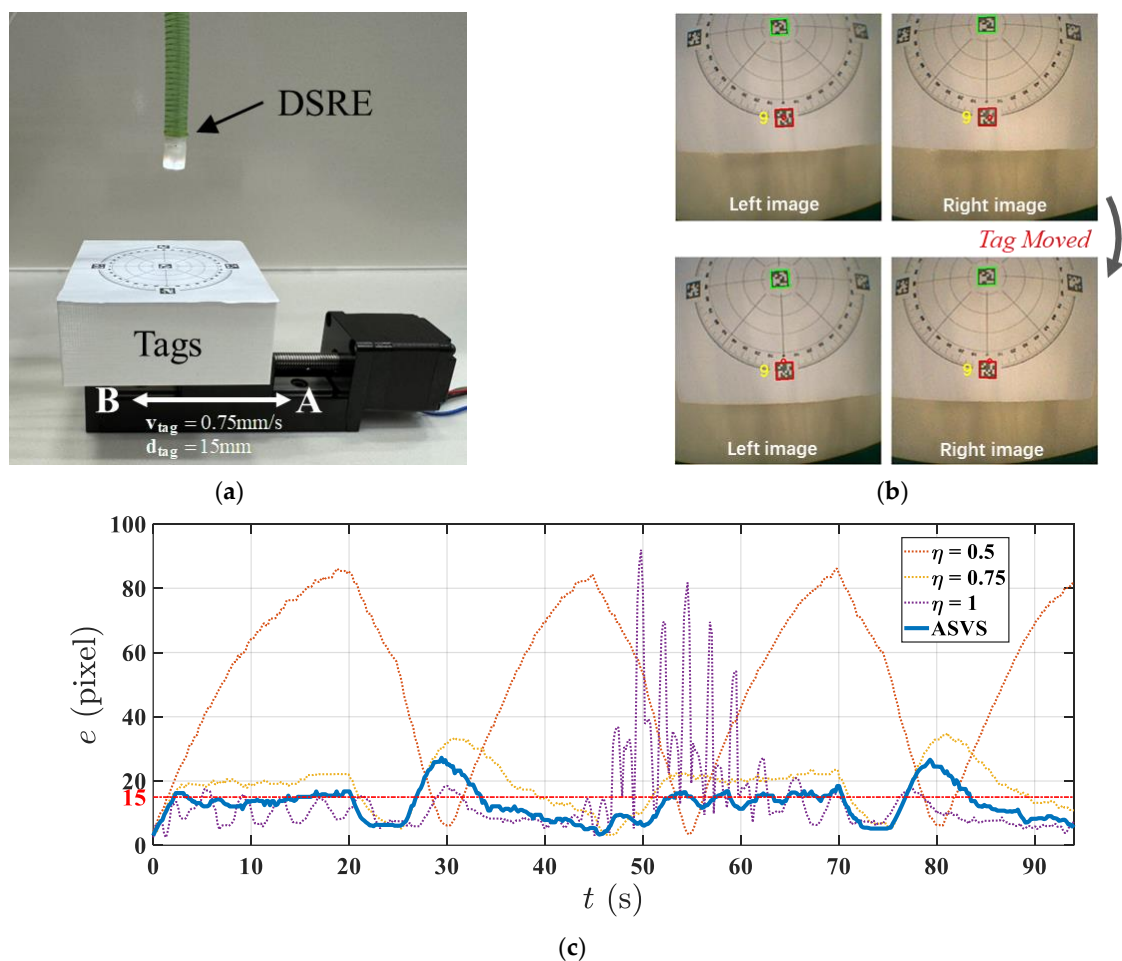


Figure 11. Dynamic target tracking results. (a) Experiment setup. (b) Endoscopic view when the tag moved. (c) Tracking error of different controllers.

Table 4. Control performance of ASVS and the classic control law with different constant η in [28].

Controller	RMSE		MAE		SD		$e_{\max}$	
	Data (pixel)	Dec. (%)	Data (pixel)	Dec. (%)	Data (pixel)	Dec. (%)	Data (pixel)	Dec. (%)
ASVS	13.48	-	12.46	-	5.14	-	27.24	-
$\eta = 0.5$	57.63	76.57	52.06	76.07	24.74	79.22	86.20	68.40
$\eta = 0.75$	19.37	30.41	17.88	30.31	7.44	30.91	34.91	21.97
$\eta = 1$	18.65	27.72	13.88	10.23	12.47	58.78	91.95	70.38

4.2.3. Tracking under Unknown External Disturbance

The performance of dealing with external disturbance was also evaluated. During the test, the robot utilized endoscopic image feedback exclusively. An external force of unknown magnitude and location was applied to the soft robot using a slender beam affixed with a single-axial force sensor, which caused a mutable robot configuration. The control objective was to bring the target tag into the predefined ROIs on the image planes. The robot configuration and experimental results are shown in Figure 12. During the experiment, at 11.6 s, the robot first received an external force of 120 mN, and the tracking error increased from 5.9 pixels to 93.4 pixels. The controller subsequently engaged to resist the external force and tracked the tag to the ROIs within approximately 3.5 s, reducing the error to less than 15 pixels. After the external force was maintained for 5 s, it was removed, and the controller repeated the above error correction process. During the entire test process, the robot was affected by dynamic external forces a total of three times. The maximum external force was 150 mN. The robot could maintain the tracking of the target and the error could converge to five pixels. This shows that image-based feedback control is efficacious in enabling the DSRE robot to adjust its configuration responsively to external environmental conditions, thereby enhancing the tracking precision.

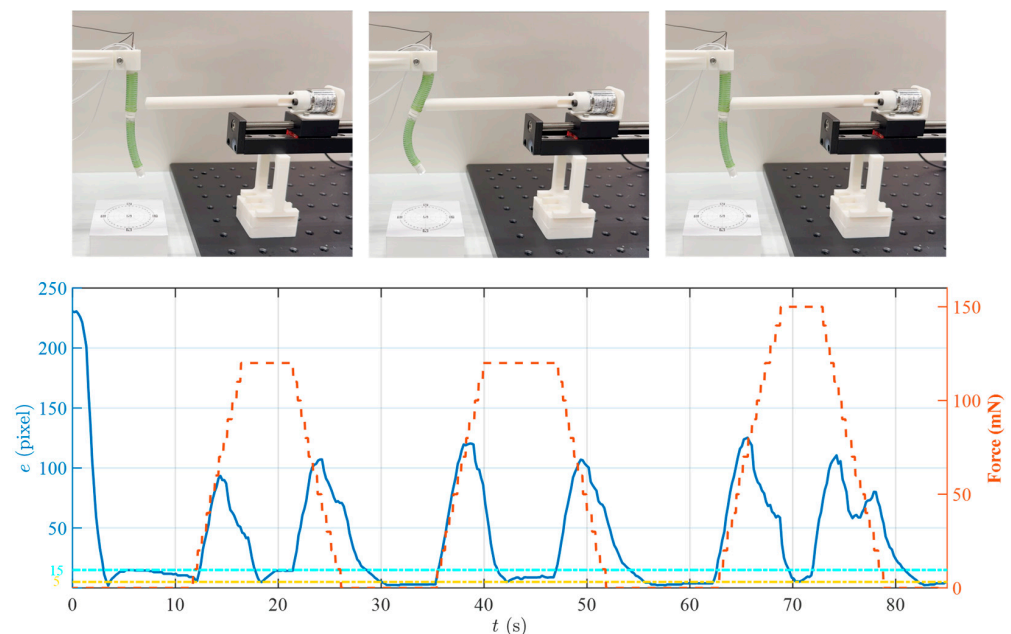


Figure 12. Tracking results under unknown external disturbance. The upper pictures show the robot configurations before, during, and after the external force application. The figure below shows the changes in external force and tracking error results.

4.2.4. Trajectory Tracking

Marking the lesion is an important early step in ESD surgery [7]. The end effector of the soft robot carrying the electrosurgical knife should always be perpendicular to the working surface, marking along the periphery of the target lesion by coagulation. This experiment aims to evaluate the feasibility of robots in the marking process, with the DSRE tracking a triangular reference trajectory. This trajectory guided the tag moving along the pre-defined path in the captured images. The experimental trajectories of the pixel positions on the left and right image planes are shown in Figure 13. The overall RMSE of the left and right image errors was 13.51 pixels, which demonstrates that the controller has a good tracking performance for the trajectories.

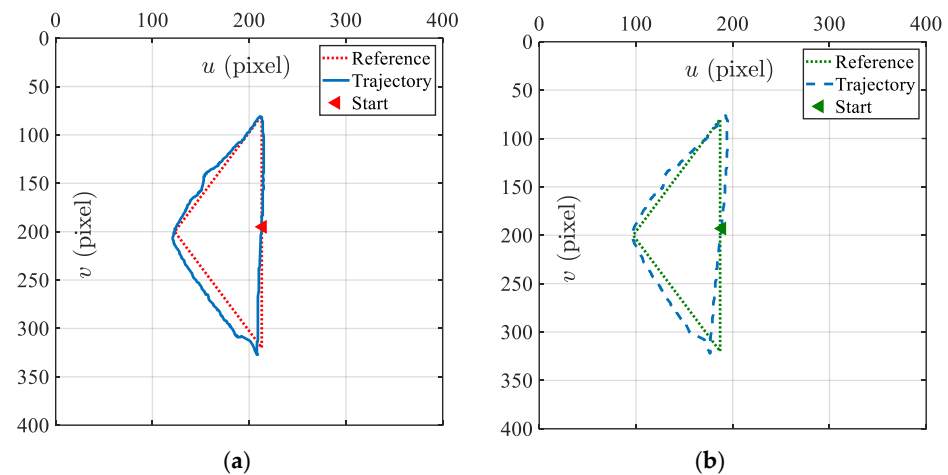


Figure 13. Triangular trajectory tracking results. (a) Target movement in the left image. (b) Target movement in the right image.

4.3. Resection in an Ex Vivo Porcine Stomach

This test was designed to assess the practical efficacy of robotic systems in ESD procedures. As shown in Figure 14, an ex vivo porcine stomach was affixed within a receptacle to simulate surgical conditions. The operator used a gamepad to perform a remote operation program on the DSRE and controlled the robot to complete the ESD surgery in the stomach. Figure 15 shows the entire process of the ESD surgery. According to the DSRE configuration shown in Figure 2c, the angle between the end effector and the horizontal plane was only 12.5° , which illustrates that the robot can cut tissue in multiple directions. Such a small operating angle can improve the cutting efficiency and safety of traditional ESD. Figure 15a,b show the left endoscopic images during ESD marking and cutting, respectively. The operator operated the DSRE to perform electrocoagulation around the lesion and successfully completed a series of marking points. During the cutting process, the operator used tweezers to assist in pulling the target tissue up to expose the cutting point to the robot and used the joystick of the gamepad to adjust the cutting angle of the robot. After the procedure, as illustrated in Figure 15c, the DSRE has completely removed the target tissue mass without causing perforation in the porcine stomach wall. Finally, a piece of tissue measuring approximately 5×4 mm was successfully removed. Therefore, this ex vivo experiment illustrates the feasibility of applying the DSRE in ESD surgery.

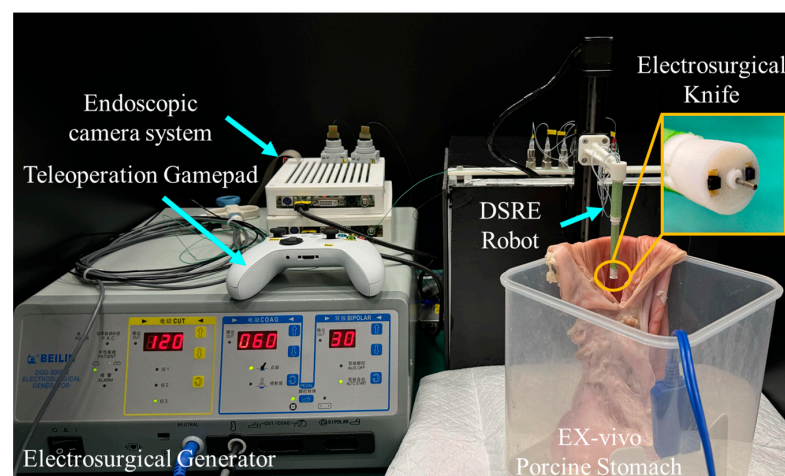


Figure 14. Experimental setup of ex vivo trial in a porcine stomach.

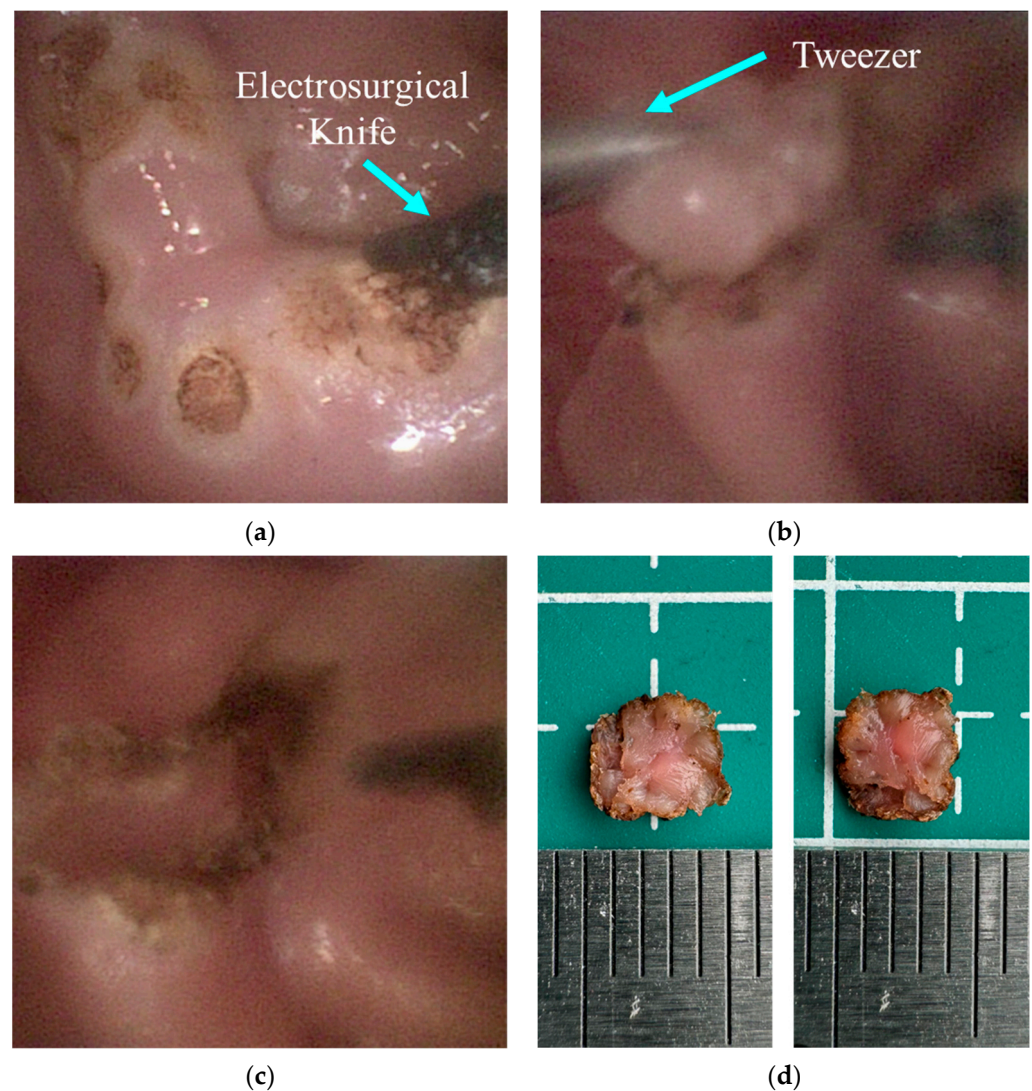


Figure 15. ESD test results in ex vivo porcine stomach. (a) Marking points. (b) Cutting with assistance. (c) Cutting results. (d) Target tissue after dissection.

5. Conclusions

This work presents a novel dual-segment soft robotic endoscope with stereo vision designed to enhance reachability and dexterity during ESD surgery. The proposed system can execute multi-angle cutting operations at a small angle relative to the lesion surface, allowing for efficient en bloc resection. Additionally, the system incorporates two calibrated RGB cameras and a depth estimation algorithm to provide real-time 3D information of the tumor, which is also used to guide the control framework. Based on the adaptively updated gain and PD control laws, the stereo vision servo controller improves the convergence speed and path tracking performance during surgery. The experimental results indicate that the proposed system improves the motion stability and precision. Ex vivo testing further demonstrates its significant potential in endoscopic surgery.

Supplementary Materials: The following supporting information can be downloaded at: <https://www.mdpi.com/article/10.3390/mi15020276/s1>, Video S1: the experimental results of the dual-segment robotic endoscope (DSRE).

Author Contributions: Conceptualization, J.C. and H.L.; funding acquisition, H.L.; methodology, J.C., S.W. and Q.Z.; project administration, H.L.; resources, W.H.; software, J.C. and M.C.; supervision, H.L.; validation, J.C., S.W. and Y.W.; visualization, J.C., S.W. and Q.Z.; writing—original draft, J.C. and S.W.; writing—review and editing, Q.Z., J.H. and H.L. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by the Centre of AI and Robotics, Hong Kong Institute of Science and Innovation, Chinese Academy of Sciences, sponsored by InnoHK Funding, HKSAR, partially supported by the Institute of Automation, Chinese Academy of Sciences, and partially supported by the Sichuan Science and Technology Program (Grant number: 2023YFH0093).

Data Availability Statement: Data are contained within the article.

Acknowledgments: Parts of Figure 1 were created using templates from Servier Medical Art (<http://smart.servier.com/>), accessed on 14 December 2023, licensed under a Creative Commons Attribution 3.0 Generic License.

Conflicts of Interest: The authors declare no conflicts of interest.

Appendix A

The translation and rotation matrices in (2) are given by

$$\begin{aligned} \mathbf{R}_y(\theta) &= \begin{bmatrix} c\theta & 0 & s\theta & 0 \\ 0 & 1 & 0 & 0 \\ -s\theta & 0 & c\theta & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}, \mathbf{R}_z(\varphi) = \begin{bmatrix} c\varphi & -s\varphi & 0 & 0 \\ s\varphi & c\varphi & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}, \\ \boldsymbol{\tau}_x(t) &= \begin{bmatrix} 1 & 0 & 0 & t \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}, \boldsymbol{\tau}_z(t) = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & t \\ 0 & 0 & 0 & 1 \end{bmatrix} \end{aligned} \quad (\text{A1})$$

In order to simplify the formula, use s and c instead of sine and cosine functions, and the forward kinematics in (2) can be extended to

$$\begin{aligned} {}^b_t\mathbf{T} &= \begin{bmatrix} {}^b_t\mathbf{R} & {}^b_t\mathbf{P} \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} T_{11} & T_{12} & T_{13} & T_{14} \\ T_{21} & T_{22} & T_{23} & T_{24} \\ T_{31} & T_{32} & T_{33} & T_{34} \\ 0 & 0 & 0 & 1 \end{bmatrix} \\ &= \begin{bmatrix} -c(\theta_e)\sigma_4 - \sigma_1 & s(\varphi_e)\sigma_8 - c(\varphi_e)\sigma_9 & \sigma_2 - s(\theta_e)\sigma_4 & \frac{l_e(c(\theta_e)\sigma_4 + \sigma_1)}{\theta_e} - t_d(s(\theta_e)\sigma_4 - \sigma_2) + \frac{l_s c(\varphi_s)}{\theta_s} + z_e c(\varphi_s) s(\theta_s) - \frac{l_e \sigma_4}{\theta_e} - \frac{l_s c(\theta_s) c(\varphi_s)}{\theta_s} \\ \sigma_5 & c(\varphi_e)\sigma_{12} - s(\varphi_e)\sigma_{13} & \sigma_6 & t_d \sigma_6 - \frac{l_e \sigma_5}{\theta_e} + \frac{l_s s(\varphi_s)}{\theta_s} + \frac{z_e s(\theta_s) s(\varphi_s)}{\theta_e} + \frac{l_e \sigma_{10}}{\theta_e} - \frac{l_s c(\theta_s) s(\varphi_s)}{\theta_s} \\ -\sigma_3 - c(\theta_e)\sigma_{11} & c(\varphi_e)s(\theta_s)\sigma_{14} + s(\theta_s)\sigma_{15} s(\varphi_e) & \sigma_7 & z_s + t_d \sigma_7 + z_e c(\theta_s) - \frac{l_e \sigma_{11}}{\theta_e} + \frac{l_s s(\theta_s)}{\theta_s} + \frac{l_e(\sigma_3 + c(\theta_e)\sigma_{11})}{\theta_e} \\ 0 & 0 & 0 & 1 \end{bmatrix} \end{aligned} \quad (\text{A2})$$

where

$$\begin{aligned}\sigma_1 &= c(\varphi_s)s(\theta_e)s(\theta_s), \sigma_2 = c(\theta_e)c(\varphi_s)s(\theta_s), \sigma_3 = c(\theta_s)s(\theta_e), \\ \sigma_4 &= c(\varphi_e)\sigma_8 + s(\varphi_e)\sigma_9, \sigma_5 = c(\theta_e)\sigma_{10} - s(\theta_e)s(\theta_s)s(\varphi_s), \\ \sigma_6 &= s(\theta_e)\sigma_{10} + c(\theta_e)s(\theta_s)s(\varphi_s), \sigma_7 = c(\theta_e)c(\theta_s) - s(\theta_e)\sigma_{11} \\ \sigma_8 &= s(\varphi_s)\sigma_{14} - c(\theta_s)c(\varphi_s)\sigma_{15}, \sigma_9 = \sigma_{15}s(\varphi_s) + c(\theta_s)c(\varphi_s)\sigma_{14}, \\ \sigma_{10} &= c(\varphi_e)\sigma_{13} + s(\varphi_e)\sigma_{12}, \sigma_{11} = c(\varphi_e)s(\theta_s)\sigma_{15} - s(\theta_s)s(\varphi_e)\sigma_{14}, \\ \sigma_{12} &= c(\varphi_s)\sigma_{15} - c(\theta_s)s(\varphi_s)\sigma_{14}, \sigma_{13} = c(\varphi_s)\sigma_{14} + c(\theta_s)\sigma_{15}s(\varphi_s), \sigma_{14} = s(\varphi_e + \frac{\pi}{3}).\end{aligned}$$

The pose $\Theta_t \in \mathbb{R}^6$ of the robot end effector can be calculated from (A2)

$$\begin{aligned}\Theta_t &= \begin{bmatrix} {}^b_P \\ {}^b_t\Omega \end{bmatrix}^T = \begin{bmatrix} P_x & P_y & P_z & \Omega_x & \Omega_y & \Omega_z \end{bmatrix}^T \\ P_x &= T_{14}, P_y = T_{24}, P_z = T_{34} \\ \Omega_x &= a \tan 2(T_{32}, T_{33}), \Omega_y = a \tan 2(-T_{31}, \sqrt{T_{32}^2 + T_{33}^2}), \Omega_z = a \tan 2(T_{21}, T_{11})\end{aligned}\quad (A3)$$

The velocity $\mathbf{v}_t \in \mathbb{R}^6$ of the robot end effector is

$$\mathbf{v}_t = \dot{\Theta}_t = [\tau_x \quad \tau_y \quad \tau_z \quad \omega_x \quad \omega_y \quad \omega_z]^T \quad (A4)$$

Thus, the Jacobian matrix $\mathbf{J}_r \in \mathbb{R}^{6 \times 5}$ in (3) is used to represent the approximate linear relationship between the robot tip velocity $\mathbf{v}_t$ and joint velocity $\dot{\Phi}$, which can be derived from the forward kinematics as follows:

$$\mathbf{v}_t = \mathbf{J}_r \dot{\Phi}$$

$$\begin{bmatrix} \tau_x \\ \tau_y \\ \tau_z \\ \omega_x \\ \omega_y \\ \omega_z \end{bmatrix} = \begin{bmatrix} \frac{\partial \tau_x}{\partial z_s} & \frac{\partial \tau_x}{\partial \theta_s} & \frac{\partial \tau_x}{\partial \varphi_s} & \frac{\partial \tau_x}{\partial \theta_e} & \frac{\partial \tau_x}{\partial \varphi_e} \\ \frac{\partial \tau_y}{\partial z_s} & \frac{\partial \tau_y}{\partial \theta_s} & \frac{\partial \tau_y}{\partial \varphi_s} & \frac{\partial \tau_y}{\partial \theta_e} & \frac{\partial \tau_y}{\partial \varphi_e} \\ \frac{\partial \tau_z}{\partial z_s} & \frac{\partial \tau_z}{\partial \theta_s} & \frac{\partial \tau_z}{\partial \varphi_s} & \frac{\partial \tau_z}{\partial \theta_e} & \frac{\partial \tau_z}{\partial \varphi_e} \\ \dots & \dots & \dots & \dots & \dots \\ \frac{\partial \omega_x}{\partial z_s} & \frac{\partial \omega_x}{\partial \theta_s} & \frac{\partial \omega_x}{\partial \varphi_s} & \frac{\partial \omega_x}{\partial \theta_e} & \frac{\partial \omega_x}{\partial \varphi_e} \\ \frac{\partial \omega_y}{\partial z_s} & \frac{\partial \omega_y}{\partial \theta_s} & \frac{\partial \omega_y}{\partial \varphi_s} & \frac{\partial \omega_y}{\partial \theta_e} & \frac{\partial \omega_y}{\partial \varphi_e} \\ \frac{\partial \omega_z}{\partial z_s} & \frac{\partial \omega_z}{\partial \theta_s} & \frac{\partial \omega_z}{\partial \varphi_s} & \frac{\partial \omega_z}{\partial \theta_e} & \frac{\partial \omega_z}{\partial \varphi_e} \end{bmatrix} \begin{bmatrix} z_s \\ \theta_s \\ \varphi_s \\ \theta_e \\ \varphi_e \end{bmatrix} \quad (A5)$$

References

- Asawa, S.; Nüesch, M.; Gvozdenovic, A.; Aceto, N. Circulating Tumour Cells in Gastrointestinal Cancers: Food for Thought? *Br. J. Cancer* **2023**, *128*, 1981–1990. [\[CrossRef\]](#)
- Feng, R.-M.; Zong, Y.-N.; Cao, S.-M.; Xu, R.-H. Current Cancer Situation in China: Good or Bad News from the 2018 Global Cancer Statistics? *Cancer Commun.* **2019**, *39*, 22. [\[CrossRef\]](#)
- Lei, Z.-N.; Teng, Q.-X.; Tian, Q.; Chen, W.; Xie, Y.; Wu, K.; Zeng, Q.; Zeng, L.; Pan, Y.; Chen, Z.-S.; et al. Signaling Pathways and Therapeutic Interventions in Gastric Cancer. *Signal Transduct. Target. Ther.* **2022**, *7*, 358. [\[CrossRef\]](#)
- Huang, R.J.; Hwang, J.H. Improving the Early Diagnosis of Gastric Cancer. *Gastrointest. Endosc. Clin. N. Am.* **2021**, *31*, 503–517. [\[CrossRef\]](#)
- Hirao, M.; Masuda, K.; Asanuma, T.; Naka, H.; Noda, K.; Matsuura, K.; Yamaguchi, O.; Ueda, N. Endoscopic Resection of Early Gastric Cancer and Other Tumors with Local Injection of Hypertonic Saline-Epinephrine. *Gastrointest. Endosc.* **1988**, *34*, 264–269. [\[CrossRef\]](#)
- Maple, J.T.; Dayyeh, B.K.A.; Chauhan, S.S.; Hwang, J.H.; Komanduri, S.; Manfredi, M.; Konda, V.; Murad, F.M.; Siddiqui, U.D.; Banerjee, S. Endoscopic Submucosal Dissection. *Gastrointest. Endosc.* **2015**, *81*, 1311–1325. [\[CrossRef\]](#)
- Esaki, M.; Ihara, E.; Gotoda, T. Endoscopic Instruments and Techniques in Endoscopic Submucosal Dissection for Early Gastric Cancer. *Expert Rev. Gastroenterol. Hepatol.* **2021**, *15*, 1009–1020. [\[CrossRef\]](#) [\[PubMed\]](#)
- Kume, K. Flexible Robotic Endoscopy: Current and Original Devices. *Comput. Assist. Surg.* **2016**, *21*, 150–159. [\[CrossRef\]](#) [\[PubMed\]](#)
- Osawa, K.; Bandara, D.S.V.; Nakadate, R.; Nagao, Y.; Akahoshi, T.; Eto, M.; Arata, J. Stress Dispersion Design in Continuum Compliant Structure toward Multi-DOF Endoluminal Forceps. *Appl. Sci.* **2022**, *12*, 2480. [\[CrossRef\]](#)
- Zorn, L.; Nageotte, F.; Zanne, P.; Legner, A.; Dallemagne, B.; Marescaux, J.; de Mathelin, M. A Novel Telemanipulated Robotic Assistant for Surgical Endoscopy: Preclinical Application to ESD. *IEEE Trans. Biomed. Eng.* **2018**, *65*, 797–808. [\[CrossRef\]](#) [\[PubMed\]](#)
- Chiu, P.W.; Phee, S.J.; Bhandari, P.; Sumiyama, K.; Ohya, T.; Wong, J.; Poon, C.C.; Tajiri, H.; Nakajima, K.; Ho, K.Y. Enhancing Proficiency in Performing Endoscopic Submucosal Dissection (ESD) by Using a Prototype Robotic Endoscope. *Endosc. Int. Open* **2015**, *3*, E439–E442. [\[CrossRef\]](#)

12. Ho, K.-Y.; Phee, S.J.; Shabbir, A.; Low, S.C.; Huynh, V.A.; Kencana, A.P.; Yang, K.; Lomanto, D.; So, B.Y.J.; Wong, Y.Y.J.; et al. Endoscopic Submucosal Dissection of Gastric Lesions by Using a Master and Slave Transluminal Endoscopic Robot (MASTER). *Gastrointest. Endosc.* **2010**, *72*, 593–599. [[CrossRef](#)] [[PubMed](#)]
13. Knoop, R.F.; Wedi, E.; Petzold, G.; Bremer, S.C.B.; Amanzada, A.; Ellenrieder, V.; Neesse, A.; Kunsch, S. Endoscopic Submucosal Dissection with an Additional Working Channel (ESD+): A Novel Technique to Improve Procedure Time and Safety of ESD. *Surg. Endosc.* **2021**, *35*, 3506–3512. [[CrossRef](#)] [[PubMed](#)]
14. Hwang, M.; Kwon, D.-S. K-FLEX: A Flexible Robotic Platform for Scar-Free Endoscopic Surgery. *Int. J. Med. Robot. Comput. Assist. Surg. MRCAS* **2020**, *16*, e2078. [[CrossRef](#)]
15. Sarli, N.; Del Giudice, G.; Herrell, D.S.; Simaan, N. A Resectoscope for Robot-Assisted Transurethral Surgery1. *J. Med. Devices* **2016**, *10*, 020911. [[CrossRef](#)]
16. Wagner, O.J.; Hagen, M.; Kurmann, A.; Horgan, S.; Candinas, D.; Vorburger, S.A. Three-Dimensional Vision Enhances Task Performance Independently of the Surgical Method. *Surg. Endosc.* **2012**, *26*, 2961–2968. [[CrossRef](#)]
17. Sachdeva, R.; Traboulsi, E.I. Performance of Patients With Deficient Stereoacuity on the EYESi Microsurgical Simulator. *Am. J. Ophthalmol.* **2011**, *151*, 427–433.e1. [[CrossRef](#)] [[PubMed](#)]
18. Bogdanova, R.; Boulanger, P.; Zheng, B. Depth Perception of Surgeons in Minimally Invasive Surgery. *Surg. Innov.* **2016**, *23*, 515–524. [[CrossRef](#)]
19. Douissard, J.; Hagen, M.E.; Morel, P. The Da Vinci Surgical System. In *Bariatric Robotic Surgery: A Comprehensive Guide*; Domene, C.E., Kim, K.C., Vilallonga Puy, R., Volpe, P., Eds.; Springer International Publishing: Cham, Switzerland, 2019; pp. 13–27. ISBN 978-3-030-17223-7.
20. Puntambekar, S.P.; Rajesh, K.N.; Goel, A.; Hivre, M.; Bharambe, S.; Chitale, M.; Panse, M. Colorectal Cancer Surgery: By Cambridge Medical Robotics Versus Surgical Robot System—a Single-Institution Study. Our Experience. *J. Robot. Surg.* **2022**, *16*, 587–596. [[CrossRef](#)]
21. Chen, J.; Wang, S.; Zhao, Q.; Chen, M.; Liu, H. A Robotized Soft Endoscope with Stereo Vision for Upper Gastrointestinal Endoscopic Submucosal Dissection (ESD). In Proceedings of the 2023 45th Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC), Sydney, Australia, 24–27 July 2023; pp. 1–6.
22. Zhao, Q.; Lai, J.; Hu, X.; Chu, H.K. Dual-Segment Continuum Robot With Continuous Rotational Motion Along the Deformable Backbone. *IEEE/ASME Trans. Mechatron.* **2022**, *27*, 4994–5004. [[CrossRef](#)]
23. Wockenfuß, W.R.; Brandt, V.; Weisheit, L.; Drossel, W.-G. Design, Modeling and Validation of a Tendon-Driven Soft Continuum Robot for Planar Motion Based on Variable Stiffness Structures. *IEEE Robot. Autom. Lett.* **2022**, *7*, 3985–3991. [[CrossRef](#)]
24. Chen, M.; Huang, Y.; Chen, J.; Zhou, T.; Chen, J.; Liu, H. Fully Robotized 3D Ultrasound Image Acquisition for Artery. In Proceedings of the 2023 IEEE International Conference on Robotics and Automation (ICRA), London, UK, 29 May–2 June 2023; pp. 2690–2696.
25. Li, Y.; Ng, W.Y.; Li, W.; Huang, Y.; Zhang, H.; Xian, Y.; Li, J.; Sun, Y.; Chiu, P.W.Y.; Li, Z. Towards Semi-Autonomous Colon Screening Using an Electromagnetically Actuated Soft-Tethered Colonoscope Based on Visual Servo Control. *IEEE Trans. Biomed. Eng.* **2023**, *71*, 77–88. [[CrossRef](#)]
26. Xu, F.; Zhang, Y.; Sun, J.; Wang, H. Adaptive Visual Servoing Shape Control of a Soft Robot Manipulator Using Bézier Curve Features. *IEEE/ASME Trans. Mechatron.* **2023**, *28*, 945–955. [[CrossRef](#)]
27. Fallah, M.M.H.; Norouzi-Ghazbi, S.; Mehrkish, A.; Janabi-Sharifi, F. Depth-Based Visual Predictive Control of Tendon-Driven Continuum Robots. In Proceedings of the 2020 IEEE/ASME International Conference on Advanced Intelligent Mechatronics (AIM), Boston, MA, USA, 6–9 July 2020; pp. 488–494.
28. Chaumette, F.; Hutchinson, S. Visual Servo Control. I. Basic Approaches. *IEEE Robot. Autom. Mag.* **2006**, *13*, 82–90. [[CrossRef](#)]
29. Jiang, J.; Wang, Y.; Jiang, Y.; Xie, H.; Tan, H.; Zhang, H. A Robust Visual Servoing Controller for Anthropomorphic Manipulators With Field-of-View Constraints and Swivel-Angle Motion: Overcoming System Uncertainty and Improving Control Performance. *IEEE Robot. Autom. Mag.* **2022**, *29*, 104–114. [[CrossRef](#)]
30. Ke, F.; Li, Z.; Xiao, H.; Zhang, X. Visual Servoing of Constrained Mobile Robots Based on Model Predictive Control. *IEEE Trans. Syst. Man Cybern. Syst.* **2017**, *47*, 1428–1438. [[CrossRef](#)]
31. Zhang, C.; Zhu, W.; Peng, J.; Han, Y.; Liu, W. Visual Servo Control of Endoscope-Holding Robot Based on Multi-Objective Optimization: System Modeling and Instrument Tracking. *Measurement* **2023**, *211*, 112658. [[CrossRef](#)]
32. Roshanfar, M.; Taki, S.; Sayadi, A.; Cecere, R.; Dargahi, J.; Hooshier, A. Hyperelastic Modeling and Validation of Hybrid-Actuated Soft Robot with Pressure-Stiffening. *Micromachines* **2023**, *14*, 900. [[CrossRef](#)]
33. Lau, K.C.; Leung, E.Y.Y.; Chiu, P.W.Y.; Yam, Y.; Lau, J.Y.W.; Poon, C.C.Y. A Flexible Surgical Robotic System for Removal of Early-Stage Gastrointestinal Cancers by Endoscopic Submucosal Dissection. *IEEE Trans. Ind. Inform.* **2016**, *12*, 2365–2374. [[CrossRef](#)]
34. Li, C.; He, S.; Xu, Y.; Li, D.; Guan, Y. Model and Control of Hybrid Hard-Soft Robots Using Model Predictive Control. In Proceedings of the 2022 41st Chinese Control Conference (CCC), Hefei, China, 25–27 July 2022; pp. 3705–3710.
35. Abdulhafiz, I.; Nazari, A.A.; Abbasi-Hashemi, T.; Jalali, A.; Zareinia, K.; Saeedi, S.; Janabi-Sharifi, F. Deep Direct Visual Servoing of Tendon-Driven Continuum Robots. In Proceedings of the 2022 IEEE 18th International Conference on Automation Science and Engineering (CASE), Mexico City, Mexico, 20–24 August 2022; pp. 1977–1984.

36. Greer, J.D.; Morimoto, T.K.; Okamura, A.M.; Hawkes, E.W. Series Pneumatic Artificial Muscles (sPAMs) and Application to a Soft Continuum Robot. In Proceedings of the 2017 IEEE International Conference on Robotics and Automation (ICRA), Singapore, 29 May–3 June 2017; pp. 5503–5510.
37. Kadhim, I.; Choi, J. Extension Mechanism of a Flexible End for Colonoscopy. *Med. Devices Evid. Res.* **2020**, *13*, 245–258. [[CrossRef](#)]
38. Ferhatoglu, M.F.; Kivilcim, T.; Ferhatoglu, M.F.; Kivilcim, T. Anatomy of Esophagus. In *Esophageal Abnormalities*; IntechOpen: London, UK, 2017; ISBN 978-953-51-3632-3.
39. Colonoscopes—Gastroenterology. Available online: <https://www.olympus.co.uk/medical/en/Products-and-solutions/Products/Gastroenterology/Colonoscopes.html> (accessed on 6 February 2024).
40. ELUXEO 700 Series Colonoscopes. Available online: <https://healthcaresolutions-us.fujifilm.com/products/endoscopy/diagnostic-gastroenterology/eluxeo-700-series-colonoscopes> (accessed on 6 February 2024).
41. Ehrlich, D.; Muthusamy, V.R. Device Profile of the EXALT Model D Single-Use Duodenoscope for Endoscopic Retrograde Cholangiopancreatography: Overview of Its Safety and Efficacy. *Expert Rev. Med. Devices* **2021**, *18*, 421–427. [[CrossRef](#)] [[PubMed](#)]
42. Webster, R.J.; Jones, B.A. Design and Kinematic Modeling of Constant Curvature Continuum Robots: A Review. *Int. J. Robot. Res.* **2010**, *29*, 1661–1683. [[CrossRef](#)]
43. Hirschmuller, H. Stereo Processing by Semiglobal Matching and Mutual Information. *IEEE Trans. Pattern Anal. Mach. Intell.* **2008**, *30*, 328–341. [[CrossRef](#)] [[PubMed](#)]
44. Viola, P.; Wells III, W.M. Alignment by Maximization of Mutual Information. *Int. J. Comput. Vis.* **1997**, *24*, 137–154. [[CrossRef](#)]
45. Murray, R.M.; Li, Z.; Sastry, S.S. *A Mathematical Introduction to Robotic Manipulation*, 1st ed.; CRC Press: Boca Raton, FL, USA, 2017; ISBN 978-1-315-13637-0.
46. Buss, S. Introduction to Inverse Kinematics with Jacobian Transpose, Pseudoinverse and Damped Least Squares Methods. *IEEE Trans. Robot. Autom.* **2004**, *17*, 16.
47. Zhang, Z. A Flexible New Technique for Camera Calibration. *IEEE Trans. Pattern Anal. Mach. Intell.* **2000**, *22*, 1330–1334. [[CrossRef](#)]
48. Wang, J.; Olson, E. AprilTag 2: Efficient and Robust Fiducial Detection. In Proceedings of the 2016 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), Daejeon, Republic of Korea, 9–14 October 2016; pp. 4193–4198.

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.