

Article

An Uncalibrated Image-Based Visual Servo Strategy for Robust Navigation in Autonomous Intravitreal Injection

Xiangdong He ¹, Hua Luo ^{1,*}, Yuliang Feng ², Xiaodong Wu ¹ and Yan Diao ¹¹ School of Mechanical Engineering, Sichuan University, Chengdu 610065, China² Aier School of Ophthalmology, Central South University, Changsha 410000, China

* Correspondence: luohua@scu.edu.cn

Abstract: Autonomous intravitreal injection in ophthalmology is a challenging surgical task as accurate depth measurement is difficult due to the individual differences in the patient's eye and the intricate light reflection or refraction of the eyeball, often requiring the surgeon to first preposition the end-effector accurately. Image-based visual servo (IBVS) control does not rely on depth information, exhibiting potential for addressing the issues mentioned above. Here we describe an enhanced IBVS strategy to achieve high performance and robust autonomous injection navigation. The radial basis function (RBF) kernel with strong learning capability and fast convergence is used to globally map the uncertain nonlinear strong coupling relationship in complex uncalibrated IBVS control. The Siamese neural network (SNN) is then used to compare and analyze the characteristic differences between the current and target poses, thus making an approximation of the mapping relationships between the image feature changes and the end-effector motion. Finally, a robust sliding mode controller (SMC) based on min–max robust optimization is designed to implement effective surgical navigation. Data from the simulation and the physical model experiments indicate that the maximum localization and attitude errors of the proposed method are 0.4 mm and 0.18°, exhibiting desirable accuracy with the actual surgery and robustness to disturbances. These results demonstrate that the enhanced strategy can provide a promising approach that can achieve a high level of autonomous intravitreal injection without a surgeon.

Keywords: visual servo; neural network; sliding mode control

Citation: He, X.; Luo, H.; Feng, Y.; Wu, X.; Diao, Y. An Uncalibrated Image-Based Visual Servo Strategy for Robust Navigation in Autonomous Intravitreal Injection. *Electronics* **2022**, *11*, 4184. <https://doi.org/10.3390/electronics11244184>

Academic Editors: Mohammad Jafari and Rania Hodhod

Received: 10 November 2022

Accepted: 10 December 2022

Published: 14 December 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 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

Intravitreal injections are very common minor procedures in ophthalmology [1–3], and their automation benefits patients, physicians, and society. Intravitreal injections are usually used in the treatment of endophthalmitis [4], choroidal neovascularization [5], posterior uveitis [6], and so on. The commonly injected drugs include antibiotics [7], antivirals [8], antifungals [9], and anti-VEGF agents [10]. In practice, intravitreal auto-injection requires the injection needle to enter vertically into the inferior temporal quadrant of the human eye, 3.5 to 4 mm from the corneal limbus. With high precision required, inaccurate surgical procedures may lead to conjunctival hemorrhage, conjunctival scarring, intense pain, and even traumatic cataracts [11,12].

Advances in robotics, machine learning, and imaging endow the surgical approach with better clinical outcomes. Pre-clinical and clinical evidence suggests that automatic surgeries may standardize technical operations, increase efficiency, and reduce clinical complications [13]. Cehajic-Kapetanovic et al. used optical coherence tomography (OCT)-guided surgical robots for subretinal injections and achieved high precision. However, the 3D imaging techniques based on OCT, magnetic resonance imaging (MRI), and ultrasound are too expensive and complicated for minor procedures such as intravitreal injections [14–18]. Braun et al. performed simultaneous localization and mapping (SLAM) to conduct retinal vascular cannulation and laser photocoagulation based on the features of

blood vessels in the retina [19]. S. Yang and B. C. Becker et al. also reported similar methods and results [20,21]. However, retinal vascular features are difficult to capture for intravitreal injections because the needle is located far outside the eyeball before the injection.

Visual servo uses machine vision to provide closed-loop position control for a robot end-effector [22]. The main advantage of this technique is that it imitates human vision and takes advantage of a large amount of information from the environment [23]. In recent years, visual servo has found many application scenarios in medical robot navigation, but its practical clinical implementation remains rare [24]. Image-based visual servo (IBVS) does not rely on accurate depth measurement [25,26], so it began to be used in procedures that lose depth perception such as minimally invasive surgery (MIS). C. Molnár et al. used the IBVS method to manipulate the da Vinci robot to reach the pre-placed marker [27]. P. Hynes et al. used the calibration-free IBVS technique to achieve surgical robotic manipulation of sutures with special markers [28]. These IBVS methods utilize the feature points or pre-placed special patterns; nevertheless, the eye has no easily extractable geometric features other than the pupil and iris. The eyelid, the blood vessels on the sclera, and the texture of the iris, on the other hand, are very different due to individual differences. The closest to intravitreal auto-injection is the ophthalmorobotics proposed by F. Ullrich et al. [29], which still requires the assistance of an ophthalmologist.

The main contribution of this work is the ability to achieve automatic and robust navigation of intravitreal injections without calibration and without depth information, using only limited information from the eye images. This proposed method can increase the level of automation of intravitreal injection and also helps to enhance the quality of the procedure as well as save valuable surgeon resources. We present a novel robust navigation method for autonomous intravitreal injection in a constrained environment using a visual servo-controlled robot. The developments presented in this work provide notably more autonomy compared with existing methods. To the best of our knowledge, this is the first intravitreal auto-injection system that is expected to achieve level of autonomy (LoA) 5 (a robotic surgeon that performs an entire process without the need for a human) [30,31]. This method combines a radial basis function (RBF) neural network [32], a Siamese neural network [33], and a robust SMC [34,35] to provide accurate and robust navigation for automatic intravitreal injection in the presence of inaccurate extracted features and noise. As a demonstration, we performed simulations on CoppeliaSim and compared the results with a generic multilayer perceptron (MLP) model to illustrate the advantages of this proposed method. The robustness of the method was also examined by introducing specific disturbances during the experiment. Finally, to verify the effectiveness of the algorithm, experiments were conducted using a simulated eyeball and a robotic injection device.

2. Materials and Methods

2.1. System Overview

The proposed method was based on an eye-in-hand system. In this system, two cameras fixed at the robot end-effector captured images. The controller guided the robot end-effector according to the captured image features, thus enabling the injector fixed at the robot end-effector to reach the injection pose, i.e., end-point closed-loop control [36]. The proposed method can also be used for a single camera, but two cameras can reduce errors by averaging effects.

The eyeball is a complex organ, and an accurate description of its geometry is difficult [37]. Based on the size data of the adult eyeball, it is an uncertain ellipsoid of 20.9–27.1 mm (transverse) × 20.5–26.4 mm (sagittal) × 19.9–27.0 mm (axial). There is an 11–12 mm diameter iris in the front of the eye [38]. The pupil size can vary from 10% to 80% of the iris [39]. Normally, saccades are the fastest and most amplified type of eyeball movement, with a peak speed of 450°/s [40]. However, the anesthetized eye movements are unlikely to reach the sweeping speed. We designed a simplified model: the eye was simplified as a 24 mm sphere, which could be rotated around the center of the sphere in both the X and Y directions perpendicular to the optical axis. Differences in the diameter

of the patient's iris caused differences in the target image features. The iris diameter was fixed at 12 mm in the experiment, while in practice, it needs to be pre-measured by the physician. It is extremely difficult to measure very small rotation angles in patients in the clinic, so we set a reference value of 5° in our experiments, but of course, the setting could be modified. In the simulation experiment, the eye could be directly manipulated to rotate, and in the physical experiment, the eye was pulled by a high-precision servo motor to achieve a rotation of $\pm 4.4^\circ$. If the eye was rotated beyond $\pm 5^\circ$, the entire procedure was forcibly aborted.

The injection needle path was discretized into N time intervals and represented as a set of needle tip poses at each time step $C_t = \begin{bmatrix} R_t & p_t \\ 0 & 1 \end{bmatrix} \in SE(3)$, and the image features were also in a corresponding state s . Following mainstream intravitreal injection requirements, the injection site was recommended to be in the inferotemporal quadrant 3.5 to 4 mm from the limbus, and the injection needle entered vertically at a depth of approximately 12 mm, which defined the final needle pose. There was an increased risk of retinal detachment if the injection site was too close to the iris and an increased risk of traumatic cataract formation if the injection site was too far from the iris [41]. In addition, throughout the process, the injection needle must not collide with the eye beyond the injection site. Since the iris diameter and the corneal diameter are close to each other, both around 12 mm, and the cornea is transparent, we used the iris as an approximate substitute for the cornea for navigation [42].

Many medical robots, such as the robot-assisted endonasal suturing implemented by J. Colan et al., use online path generation, mainly due to complex functional needs and the fact that longer paths can cause more trauma [43]. The intravitreal injection path had no such problems because it was simple to perform, and the needle was in the air rather than in the tissue before reaching the injection site. In addition, the uncertainty of our system was large. We suggest that designing a reference path manually, rather than generating it algorithmically, is more beneficial in ensuring the robustness of the system. We designed the path with the following considerations: (1) The closer the injection site was, the closer the needle tip was to the eye, and vice versa, to avoid accidental collisions. (2) The needle tip was gradually brought closer to the injection site from the inferior temporal direction, avoiding the bulging cornea. (3) After the path was discretized, the closer to the injection point, the denser the number of samples was to ensure accurate injection. (4) When the needle was moving on the reference path, the image features were located as close to the middle of the cameras' field of view (FOV) as possible. (5) The length was as short as possible while ensuring the above premise. The reference path and the corresponding image features we designed are shown in Figure 1.

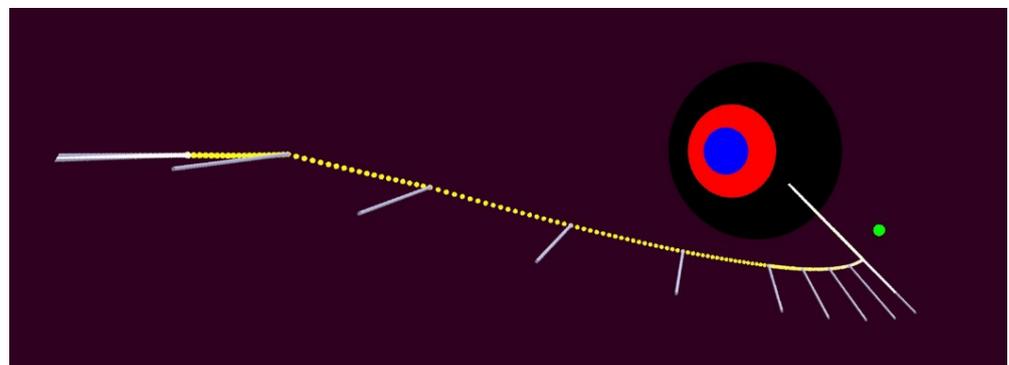


Figure 1. The reference path and the corresponding image features.

2.2. Visual Mapping Model

A visual mapping model that describes the changing relationship between the motion of the robot end-effector and visual features is a key link in the implementation of visual

servo control. The classical interaction matrix approach requires the estimated parameters (depth) and the calibrated model of the camera. Moreover, the model of the robot can contain uncertainties engendered by the movement with sliding [44]. However, due to the reflective and refractive properties of the eyeball, the measuring depth cannot be accurate enough [25,26]. Although IBVS is quite tolerant to errors in measuring depth compared with position-based visual servo (PBVS), inaccurate depth estimates can still greatly affect the actual path of IBVS, which may increase the risk of the final stage of intravitreal injection [36]. In addition, it is also common to use state estimation algorithms such as Kalman filtering to estimate Jacobi online in simple visual servo tasks [45–50]. However, in practical applications, visual servo control is an uncertain nonlinear strong coupling system, and the accurate computation of the feature Jacobian matrix is very complicated [51]. Neural networks can be used to approximate such relations, hence avoiding computing the object's feature inverse Jacobian, even at singular Jacobian postures [52]. Our proposed neural network-based approach does not require a priori knowledge of robot kinematics, hand–eye geometry, and camera models. To the best of our knowledge, this is the first visual mapping model to be used for the eye.

The output of the network proposed in other methods is mostly the relative pose of the camera or the absolute pose of the end-effector. Our network was trained to estimate the displacement Δr required to move from the current pose to the target pose, according to the current image feature s and the target image s^* . $\Delta r = \left(\Delta t_f, \Delta \eta_f \right)^T$ is the displacement of the pinpoint relative to the pinpoint reference system. The output of the network can be expressed in two forms: (1) $\Delta r = f(\Delta s, s)$ and (2) $\Delta r = f(s^*, s)$. We believe that the second approach was more favorable to calculate the relative displacement because the two inputs had the same order without the differential. A basic visual servo can be implemented with the following control laws:

$$\Delta r_{k+1} = f(s_k + \lambda(s^* - s_k), s_k) \quad (1)$$

where Δr_{k+1} is the displacement of the next pinpoint and λ is the gain, thus driving the end-effector of the robot ever closer to the target position. The inverse kinematic algorithm [53], which inverse solved the displacement of each joint of the robot from the robot end displacement, was provided by the official libfranka library.

As shown in Figure 2a, we used the following image features for robot guidance:

1. We used iris ellipse equation parameters. This was because the iris diameter varied very little from person to person compared with the pupil. We used least squares to fit the iris edge to obtain the general equation of the ellipse: $x^2 + c_1xy + c_2y^2 + c_3x + c_4y + c_5 = 0$, which has five independent parameters, c_1, c_2, c_3, c_4 , and c_5 [54]. The five parameters of the elliptic equation limited the five degrees of freedom (DOF) of the robot and ensured the exact distance from the final position to the iris edge.
2. We used the tilt angle of the line connecting the orientation indicator point to the center of the iris. The injection site needed to be on the inferior temporal side, so an orientation indicator point needed to be placed at an angle of 45° from the horizontal line. This feature was not accurate during eye rotations, but it was well tolerated in practical applications.

The image features captured by a single camera were a seven-dimensional vector, and the captures from the two cameras were combined into a 14-dimensional vector. The final robot pose is shown in Figure 2b.

The Lyapunov stability of neural network methods in visual servo is now proved [55–57]. To estimate the relative pose between the target image features and the current image features, we applied an SNN architecture. The SNN is one of the best choices for comparing two element vectors and outputting the similarity, where the two identical neural networks work in parallel and compare their outputs at the end [58]. The amplitude of the relative pose is positively correlated with the similarity between the current image features and the target image features, which is well suited for SNN. The network proposed by

F. Tokuda et al. is very inspiring and also applies an SNN architecture, but the difference is that we used RBFNN instead of convolutional neural networks (CNN) to process the features because our input was a feature vector rather than the image itself [59]. As shown in Figure 3, our network contained two parts: the high-level feature extraction part and the regression part. The high-level feature extraction part contained two parallel embedding architectures with shared weights and parameters, consisting of an RBF layer and a fully connected layer. The main advantages of RBF were a short training phase and a reduced sensitivity to the order of presentation of training data. The regression part consisted of a subtraction process and a fully connected layer. The subtraction process constrained the extracted features to zero when the two input image features were the same. This constraint alleviated the high nonlinearity between the image feature space and the end-effector space, which made it easy for the network to learn feature embeddings. The last layer of the network was the fully connected (FC) layer, where the output was the next motion relative to the current pose. Unlike the output of F. Tokuda et al., our output was relative to the injection needle itself rather than the absolute coordinate system, which was experimentally found to be effective to improve the track performance in the case of eye rotation. We also use quaternions instead of Euler angles to represent the rotation to avoid the problem of gimbal lock [60]. All activation functions in the network were Gaussian error linear unit (GELU) functions, and Dan Hendrycks et al. showed that GELU has better performance in regression tasks compared with rectified linear unit (ReLU) and exponential linear unit (ELU) [61].

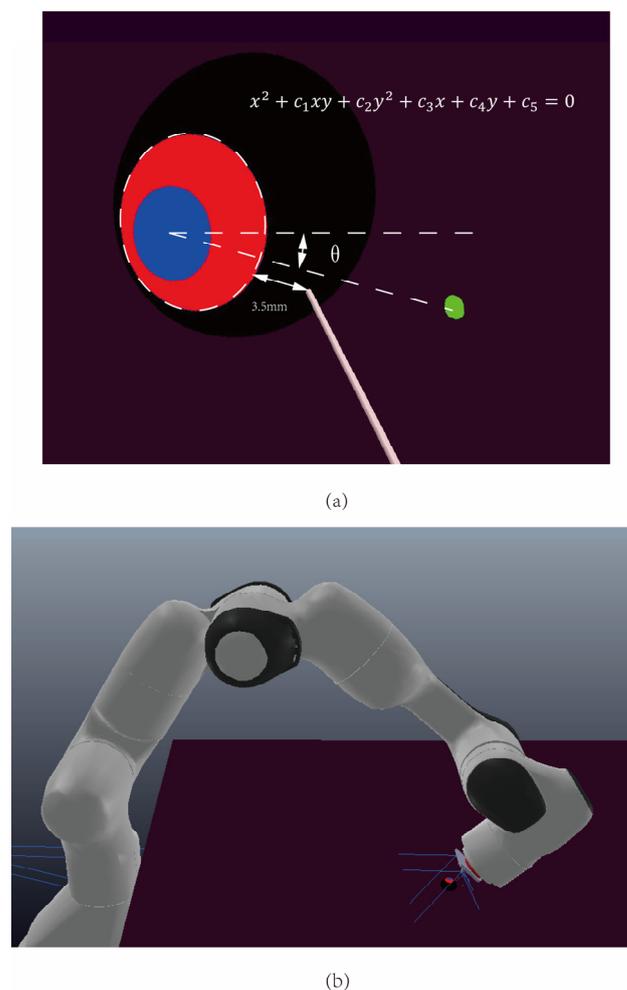


Figure 2. (a) Image features for robot guidance and injection site location. (b) The final robot pose.

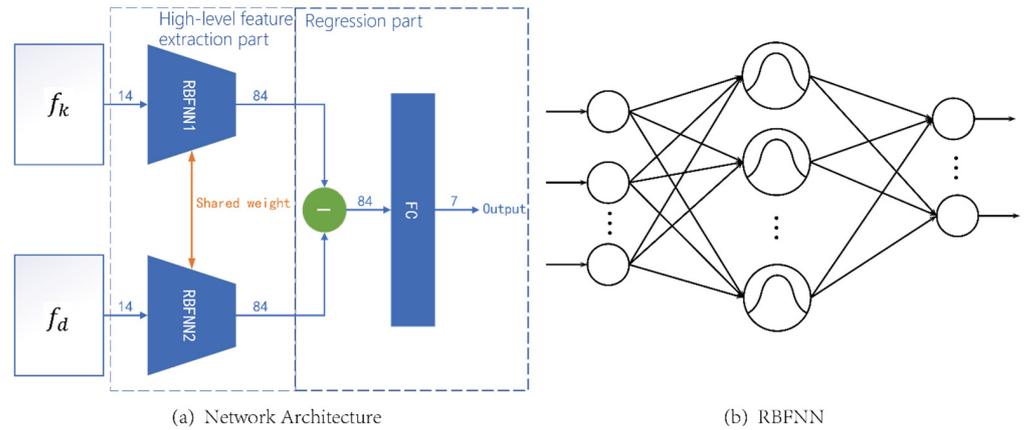


Figure 3. Visual mapping model architecture. (a) Overall model architecture. (b) Structure of RBFNN.

The data acquisition method was divided into two steps. In the first step, the space near the reference path was sampled. The reference path was interpolated to obtain the base point. The robot was then operated to move randomly around the base point, sampling a relatively small number of image features and pinpoint locations. In the second step, we selected two of the data collected in the first step and found the data used for training. We did not randomly select two data for calculation based on combinatorial theory as F. Tokuda et al. did. This was because the data set generated in this way was too large, and the large movements did not meet the requirements of practical applications. We required that the base points of the two sampled data must be adjacent to each other to limit the magnitude of the relative motion. In the first step, we captured the image feature s_i and the absolute position of the pinpoint r_i . In the second step, the training data $(s_i, s_j, \Delta r_{i,j})$, $(s_j, s_i, \Delta r_{j,i})$, and $(s_i, s_i, 0)$ were selected two at a time from the sample data, where $\Delta r_{i,j}$ denotes the relative displacement of r_j with respect to r_i .

The loss between the estimated vector and the ground truth vector was computed to regress the relative pose of the end-effector between the target and current images. As a reference, the loss function defined by F. Tokuda was

$$E = \alpha \left| \left| \Delta \tilde{t}_f - \Delta t_f \right| \right|_2 + \beta \left| \left| \Delta \tilde{\eta}_f - \Delta \eta_f \right| \right|_2 \quad (2)$$

where $\Delta \tilde{t}_f$, $\Delta \tilde{\eta}_f$, Δt_f , and $\Delta \eta_f$ are the ground truth of the relative translation, the ground truth of the relative orientation, the predicted relative translation, and the predicted relative orientation.

However, we defined the error of rotation as the angle between two quaternions instead of the mean square error of the Euler angles, so that the properties of the Lie group $SO(3)$ space could be fully utilized [62]. α and β are parameters used to adjust the training speed of the translation and rotation vector. In this paper, $\alpha = 1.0$ and $\beta = 1.0$ were used. This loss function was more favorable for motion samples with larger amplitudes, whereas in the final stage of intravitreal injection, there were a large number of motions with small amplitudes. Therefore, we defined the loss function as the relative error as follows:

$$E = \alpha \frac{\left| \left| \Delta \tilde{t}_f - \Delta t_f \right| \right|_2}{\left| \left| \Delta t_f \right| \right|_2} + \beta \frac{\left| \left| \Delta \tilde{\eta}_f - \Delta \eta_f \right| \right|_2}{\left| \left| \Delta \eta_f \right| \right|_2} \quad (3)$$

The network was trained by AdaBelief [63] using PyTorch library [64] with learning rates of 10^{-3} and 10^{-4} for 700 and 300 rounds, respectively.

2.3. Visual Servo Controller

The vision mapping model enabled the simplest visual servo controller, but there was still room for optimization. The controlled system and controller are shown in Figure 4, where $D(z)$ is the controller, $V(z)$ is the visual mapping model, and $I(z)$ is an integrating element. The proper design of $D(z)$ can improve the performance of the whole visual servo system.

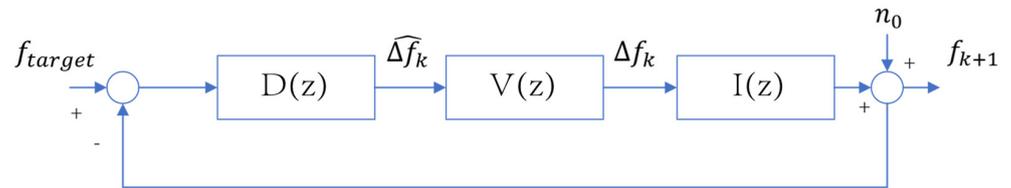


Figure 4. Visual servo system.

In the neighborhood of each node of the reference path, the system could be approximated as a linear time-invariant (LTI) system because the nodes were close to each other. With the input–output response of the neural network model, its approximate linear model $G(z) = \frac{b}{z-a}$ could be fitted using the least squares method, transformed into the differential equation as $y(k + 1) = ay(k) + bu(k)$. Although the uncalibrated IBVS did not require consideration of the camera and tool calibration, model fitting errors, system identification errors, robot repetitive motion errors, and image recognition errors could make the system model inaccurate and varied with the pose of the robot end-effector. An SMC was then designed to overcome these problems, which was robust to model uncertainties and disturbances in the environment. To overcome the chattering problem associated with the SMC, we used quasi-SMC [65] instead of the classical SMC, i.e., we used the saturation function instead of the sign function. The design of the sliding mode surface was $s(k) = y(k)$, the convergence law was the constant rate reaching law $\Delta s(k) = s(k + 1) - s(k) = -\eta \text{sgn}(s(k))$, and the control law could be calculated as $u(k) = \frac{1}{b}[(1 - a)y(k) - \eta \text{sgn}(s(k))]$, defining the Lyapunov function of the system as

$$\begin{aligned} V(k) &= \frac{1}{2}s(k)^2 > 0 (s(k) \neq 0), \\ \Delta V(k) &= V(k + 1) - V(k) = \frac{1}{2}[s(k + 1)^2 - s(k)^2] \\ &= -\frac{\eta \text{sgn}(s(k))}{2} [2s(k) - \eta \text{sgn}(s(k))] < 0 (|s(k)| > \frac{\eta}{2}) \end{aligned} \tag{4}$$

which meant that the system had Lyapunov stability. Then, the sign function was replaced with the saturated function

$$f(x) = \begin{cases} -1, & x \leq -\Delta \\ \frac{x}{\Delta}, & -\Delta < x < \Delta \\ 1, & x \geq \Delta \end{cases} \tag{5}$$

When $-\Delta < x < \Delta$, the controller was a linear controller, thus overcoming the chattering problem.

However, the optimal parameters tuned in the simulation may increase the risk of instability due to the presence of system uncertainty. Assuming that the system parameters had 10% uncertainty, we defined the problem with uncertain system parameters as a

min–max robust optimization problem, then $a \in [0.9\hat{a}, 1.1\hat{a}]$ and $b \in [0.9\hat{b}, 1.1\hat{b}]$, and the problem transformed to

$$\begin{aligned} \min_{\{\eta, \Delta\}} \max_{\{a, b\}} & \sum_{k=0}^{N-1} qy(k)^2 + ru(k)^2 + py(N)^2 \\ \text{s.t.} & a \in [0.9\hat{a}, 1.1\hat{a}] \\ & b \in [0.9\hat{b}, 1.1\hat{b}] \\ & \eta \in [0, 1] \\ & \Delta \in [0, 1] \end{aligned} \quad (6)$$

After determining the solution window N and the estimated values of the system parameters \hat{a} and \hat{b} , the values of η and Δ could be derived to obtain a robust sliding mode controller.

3. Experiments and Results

3.1. Simulation

To evaluate the performance of our proposed scheme, we first validated IBVS in a development simulation environment based on CoppeliaSim EDU V4.2 [66] and Python, in which we modeled visual features of the eye and simulated robot kinematics. There were 70,413 data sampled for model training. On the reference path, 20 values were inserted between every two neighboring nodes using the screw linear interpolation (ScLERP) [67] method, and then a uniformly distributed noise of different sizes was added.

To capture the response of the system in the face of deviation of image features, we selected two neighboring points from the sampled data as the initial and target points and ran 300 steps under the control of a proportional controller with a gain of 0.1 to obtain the input and output data of the system. For simplicity, we only analyzed the first image feature, and the other image features were similar. Based on the input and output data, the fitted discrete transfer function was $\frac{0.174452}{z-0.9310876}$, and the parameters of the robust sliding mode controller were $\eta = 0.0005$ and $\Delta = 0.01$ using the method described in the previous section. The general overshoot calculation formula was $(x_{max} - x_{\infty})/x_{\infty}$, and our calculation formula was $e_{min}/(f_{end} - f_{start}) = (f_{min} - f_{end})/(f_{end} - f_{start})$ because the feature 1 was reduced in this process, where e_{min} is the minimum value of feature error 1, f_{min} is the minimum value of feature 1, f_{start} is the feature 1 at the reference start pose, and f_{end} is the feature 1 at the reference end pose. The responses of the different methods to image feature deviations are shown in Figure 5.

We then used the proposed method to simulate the whole process. To verify the tracking performance of the algorithm, we also let the eye rotate at the 200th step (5° , 5° , 0°). The deviation of features from the target image is shown in Figure 6. All image features converged to near the ideal image feature but did not converge to 0. This was because the eye had rotated, and the target image was no longer accurate. The injection needle attitude also converged to near the ideal values, with a final deviation of 0.5 mm from the pinpoint position and an angle of 0.0438 rad for the injection needle attitude. The reference path, which was designed in Section 2.1, and the real path of the pinpoint are shown in Figure 7. Since the reference path was fixed relative to the eye, the reference path rotated accordingly after the eye rotated, which caused the jump in the reference path in Figure 7. We used the pytransform3d [68] library to calculate the reference path after rotation. Although the initial pose and the sudden eye rotation were located outside the task space, localization was accomplished through the generalization of the network. Proportional controller is commonly used in practice, and we used the proportional controller with gain $\lambda = 0.12$ as the baseline [69,70]. As a comparison, we also performed simulations using the baseline method, and a comparison of the two is shown in Table 1. The number of rising steps, overshoot, and standard deviation of the balance position in the table were derived using feature 1. The simulation results showed that the proposed method had a faster response,

less oscillation at equilibrium, and smaller final error, which could achieve the required accuracy with the actual surgery.

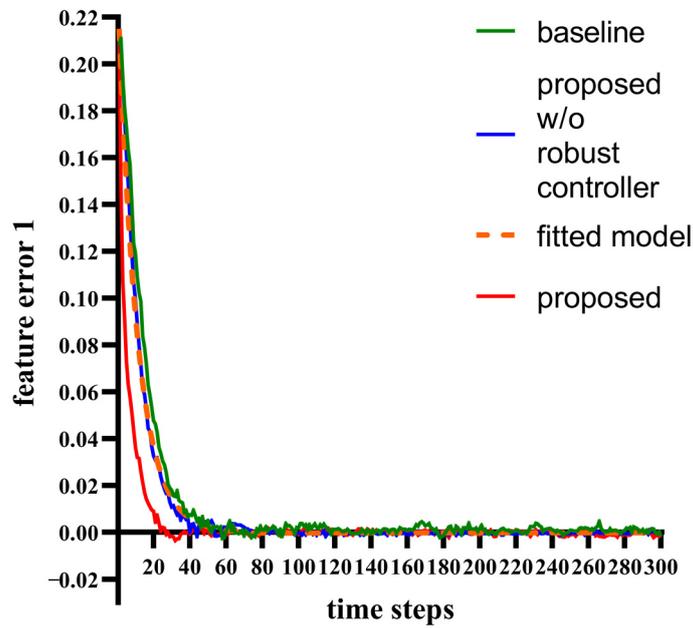


Figure 5. The responses of the different methods in the face of image feature deviations.

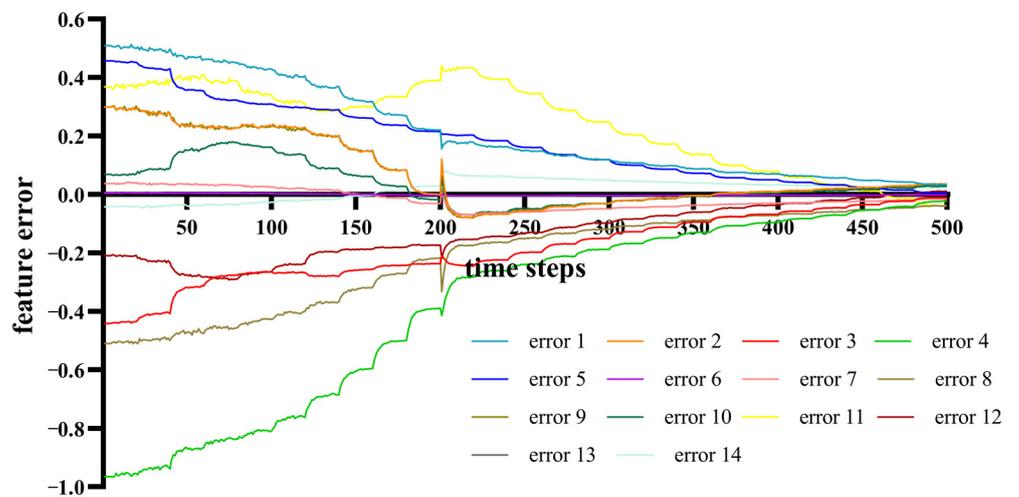


Figure 6. Image feature changes during the simulation process. Error 11 appeared to rise for a while because the needle tip needed to be gradually brought closer to the injection site from the inferior temporal direction, avoiding the bulging corneal. Other similar phenomena were also caused by the same reason.

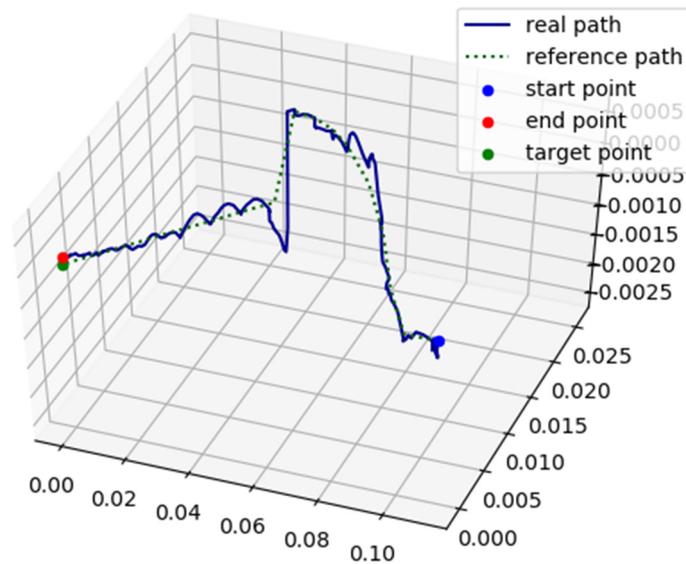


Figure 7. The path of the pinpoint during the simulation process.

Table 1. Comparison of the baseline method and the proposed method during the simulation process.

Item	Baseline	Proposed
Rise steps	28	14
Overshoot	1.23%	1.62%
Standard deviation of balance position	1.09×10^{-3}	9.48×10^{-4}
Simulation test final translation error (mm)	5.1	0.5
Simulation test final rotation error (rad)	0.0607	0.0438
Distance between injection site and iris edge (mm)	1.31	3.68
Angle of injection point deviation from horizontal line (rad)	0.0937	0.2391

3.2. Physical Model Experiments

The platform for the physical model experiments included a seven-degree-of-freedom robotic arm Franka Emika Panda, two symmetrically mounted cameras with USB video class (UVC) protocol, a coaxial ring light source, and an injector, as shown in Figure 8. The proposed methodology was implemented on a 2.9 GHz AMD computer running Linux (Ubuntu 20.04, Canonical) with real-time patches (RT-PREEMPT) [71]. During each cycle, the current robot pose was updated, and a target robot pose was computed and sent to the robot controller. Rigid body kinematics was implemented using the libfranka library that came with the robot system. Semantic segmentation of eye images is already a relatively mature technique and has been commonly used in, for example, the eye tracker and small incision lenticule extraction (SMILE) myopia surgery, so in the actual experiment, a simplified eye model (Figure 9b) was used instead of the complex actual surgical head model (Figure 9a) to execute the experiment. The simplified eye model replaced the real eye with a ball, the pupil with a solid blue circle in the center above the ball, and the iris with a red circle around it, and the rest was considered the sclera region, so that the pupil and iris regions could be accurately segmented using a simple method of color thresholding, allowing the study to focus on visual servo algorithms. The steering gear was connected to the eye model by an axis with an extension of the axis passing through the eye model sphere center. The steering gear controlled the rotation of the eye model around the axis to simulate the left–right rotation of the eye, with the rotation angle set from -4.4° to 4.4° . We conducted a robot-assisted intravitreal injection experiment to validate the performance of our system in a physically realistic environment. A total of 125,550 images were collected for model training from an initial position of the needle pinpoint 15 cm from the center of the eye to an end position where it finally touched the eyeball.

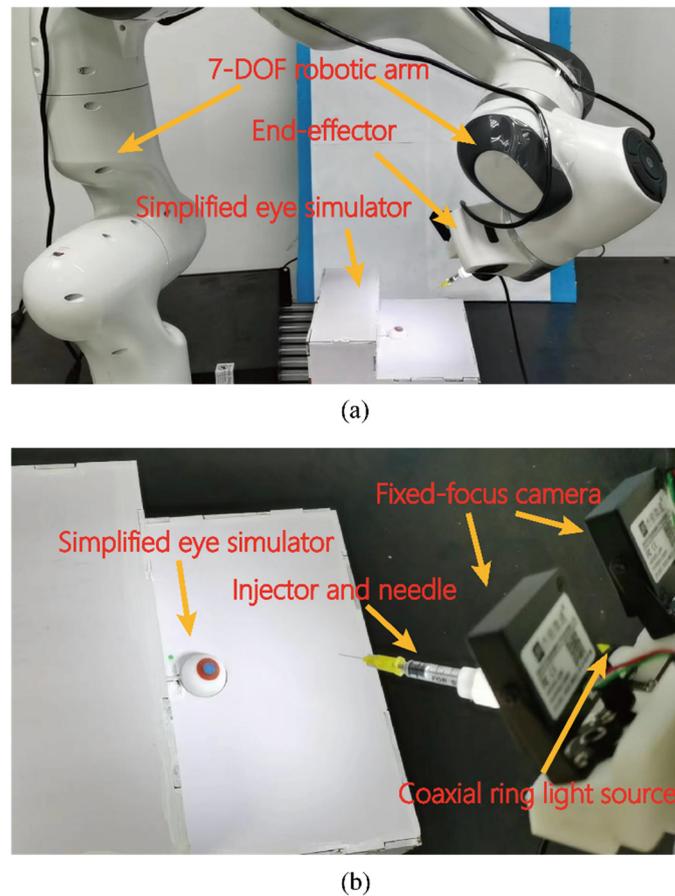


Figure 8. The model for physical experiments. (a) The injection system. (b) End-effector structure.

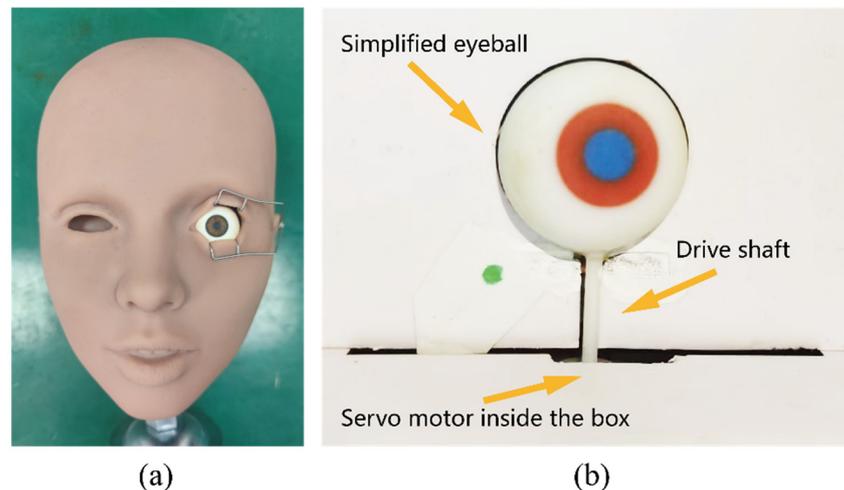


Figure 9. Simplified eye simulator. (a) Realistic eye simulator. (b) Simplified eye simulator.

The experimental results are shown in Figures 10 and 11, respectively. All image features converged to the vicinity of the desired image features, and the pose of the injection needle could reach the vicinity of the desired pose. In Figure 10, the simulation and physical model experiments used the same reference path design method, and their reference paths were not exactly the same due to differences in the positions of the eyeball model, robot, and camera. Figure 11 shows a partial zoomed-in view of the final approach to the target position to better represent the details. The reference path was specified manually during the design phase and later generated and executed by the machine in patients with the

same iris diameter, using the method in Section 2.3. The translation error was 0.4 mm, and the rotation error was 0.0032 rad, which was increased compared with the simulation, and we think it was caused by inaccurate camera and image feature extraction, but it still met the needs of the surgery. As a comparison, the positional accuracy of Hynes et al. is similar to ours, although the systems and methods of both are very different [28]. The camera acquired images at convergence as shown in Figure 12. After repeated tests, the maximal distance between the pinpoint and the iris edge was approximately 3.6 mm, which could achieve the required accuracy with the actual surgery.

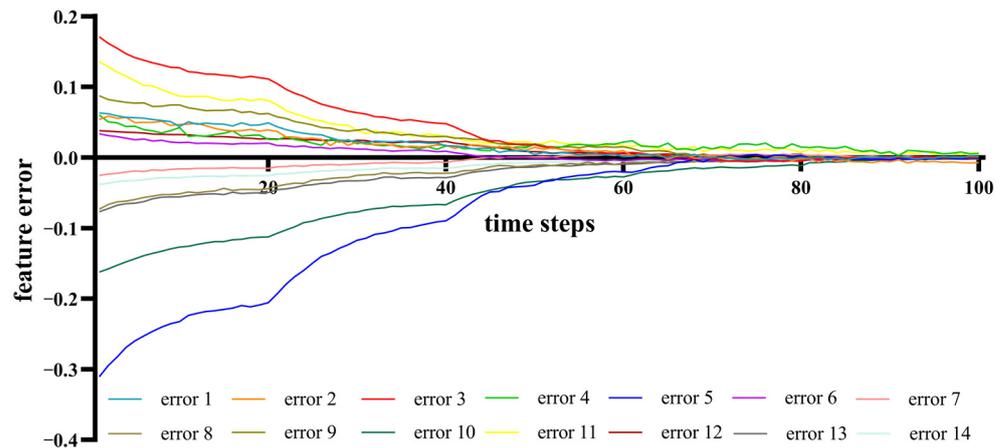


Figure 10. Image feature changes during the physical model experiments.

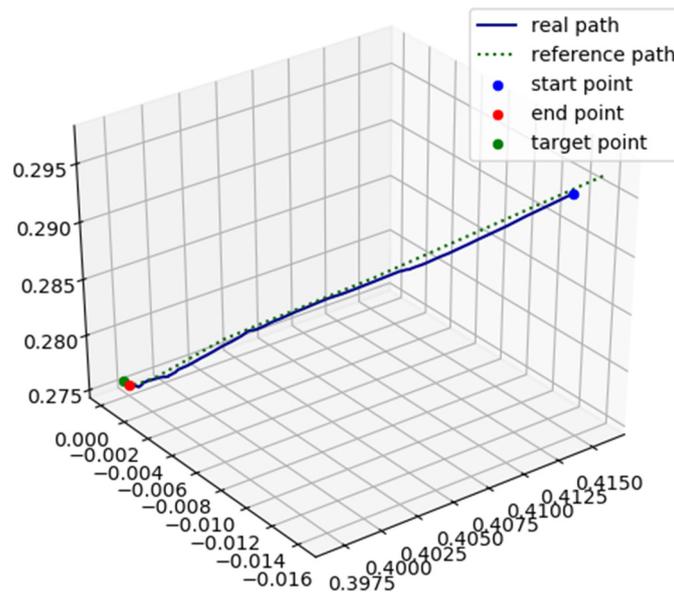


Figure 11. The path of the pinpoint during the physical model experiments. The reference path in the figure was designed manually using the method in Section 2.3 and interpolated using the ScLERP algorithm.

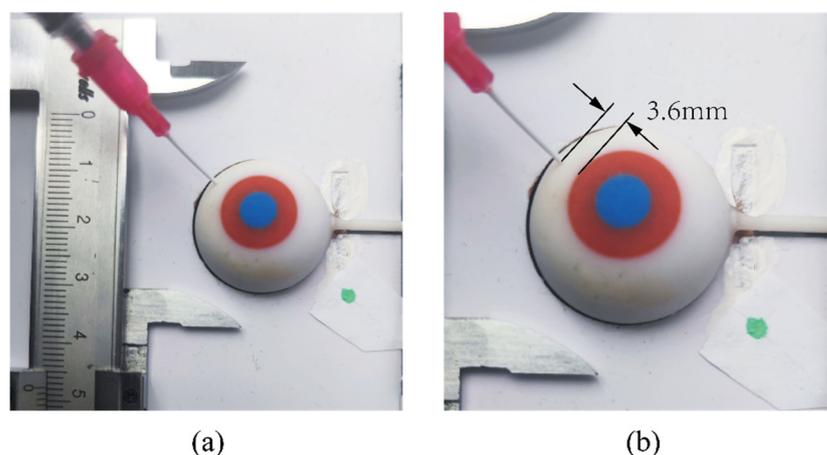


Figure 12. The final position and posture of the injection needle. (a) Measurement method for distance between the pinpoint and the iris edge. (b) Measurement results.

4. Discussion

4.1. Ablation Study

We enumerated the ablation studies in Table 2 to investigate the impact of the different components of the proposed approach. The ablation of the robust controller was studied by replacing it with a proportional controller with a gain of 0.1, the ablation of the GELU was studied by replacing it with a ReLU function, and the ablation of the loss function was studied by replacing it with a loss function of the original form proposed by F. Tokuda et al. The simulation results show that all components of the proposed method had significant positive effects.

Table 2. Ablation study.

	Proposed	w/o Robust Controller	w/o GELU	w/o Proposed Loss
Rise steps	14	25	5	6
Overshoot	1.62%	0.20%	2.04%	1.32%
Simulation test final translation error (mm)	0.5	0.7	1.6	0.8
Simulation test final rotation error (rad)	0.0438	0.0485	0.0407	0.0419
Distance between injection site and iris edge (mm), reference value: 3.5	3.68	3.55	4.03	3.45
Angle of injection point deviation from horizontal line (rad)	0.2391	0.241	0.195	0.240

4.2. Robustness to Noise

Due to the differences in equipment, environment, and lighting, there was noise pollution in the actual acquired images, which required the navigation system to have robustness to noise. To simulate this situation, we added different intensities of Gaussian noise to the image features. A comparison of the experimental results for different signal-to-noise ratio (SNR) cases with the noiseless case is shown in Table 3. For simplicity, robust controllers were not included in any of the methods compared. We also tried the simulation with an SNR = 40 dB, but the results were divergent. In the experimental results, the cases with SNRs of 100 dB, 80 dB, 60 dB, and 50 dB were almost indistinguishable from the case without adding noise, and the simulation results only deteriorated rapidly when SNR = 45 dB. The experiments illustrated that this algorithm had good robustness for environments containing noise.

Table 3. Robustness to noise.

SNR	∞	100 dB	80 dB	60 dB	50 dB	45 dB
Translation error (mm)	0.7	0.7	0.7	0.7	0.6	0.8
Rotation error (rad)	0.0485	0.0486	0.0486	0.0488	0.0489	0.0421
Distance between injection site and iris edge (mm)	3.55	3.55	3.55	3.53	3.54	3.98
Angle of injection point deviation from horizontal line (rad)	0.241	0.241	0.241	0.241	0.245	0.235

4.3. Adaptability to Fewer Samples

The time cost of training data acquisition was high, so it made sense to achieve comparable accuracies on fewer samples for the practical application of the model. With all other experimental conditions unchanged, we randomly selected 10%, 25%, 50%, and 75% samples from the original training set to form a new training set for retraining the models. The comparison of the simulation results with the original data set is shown in Table 4. For simplicity, robust controllers were not included in any of the methods compared. In the experimental results, there was almost no difference between 100%, 75%, and 50% of the cases, and 25% of the cases were still within the acceptable range, while the remainder could not meet the requirements. The experiments illustrated that the algorithm was also adaptable to a few samples within a certain range.

Table 4. Adaptability to fewer samples.

Sampling Rate	100%	75%	50%	25%	10%
Translation error (mm)	0.7	0.7	0.8	0.9	1.1
Rotation error (rad)	0.0485	0.0490	0.0481	0.0480	0.0437
Distance between injection site and iris edge (mm)	3.55	3.57	3.57	3.77	4.12
Angle of injection point deviation from horizontal line (rad)	0.241	0.228	0.247	0.243	0.188

5. Conclusions

In this work, a robust navigation system for autonomous intravitreal injection was proposed. Patient individual differences and the reflective and refractive properties of the eyeball make it impossible to use PBVS, which relies on accurate depth information. Moreover, the lack of point features in the eye that can be used for guidance also causes great difficulty for autonomous injections. To circumvent these difficulties, the system employed a combination of RBF-SNN and robust SMC for image-based visual servo. This enhanced autonomous strategy did not require a priori knowledge of robot kinematics, hand–eye geometry, and camera models. To verify this strategy, we compared the quality criteria of intravitreal injection—including the error of translation and rotation and the distance between the injection site and iris edge—of the developed system, mainstream intravitreal injection requirements, and similar visual servo tasks. Data from the simulation and the physical model experiments indicated that the developed system was capable of accurately guiding the end-effector of the robot to the intended injection location based on the limited information available to the eyeball. Although further technological advances are necessary for clinical applications, the principle and system may provide important information for the development of more advanced and automated robot systems for intravitreal injection in the future.

Author Contributions: Conceptualization, X.H. and H.L.; methodology, X.H. and H.L.; software, X.H. and H.L.; validation, X.H. and H.L.; formal analysis, X.H. and H.L.; investigation, X.H. and H.L.; resources, X.H. and H.L.; data curation, X.H. and H.L.; writing—original draft preparation, X.H. and H.L.; writing—review and editing, X.H., X.W., Y.F. and H.L.; visualization, X.H. and H.L.; supervision, X.H. and H.L.; project administration, X.W., Y.D., Y.F. and H.L.; funding acquisition, X.W., Y.D. and H.L. All authors have read and agreed to the published version of the manuscript.

Funding: This work was funded by Major Science and Technology Projects of Sichuan Province of China (2020ZDZX0023) and Sichuan Science and Technology Program (2022YFS0025 and 2022YFG0220).

Data Availability Statement: Not applicable.

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

References

- Chopra, R.; Preston, G.C.; Keenan, T.D.L.; Mulholland, P.; Patel, P.J.; Balaskas, K.; Hamilton, R.D.; Keane, P.A. Intravitreal injections: Past trends and future projections within a UK tertiary hospital. *Eye* **2021**, *36*, 1373–1378. [[CrossRef](#)] [[PubMed](#)]
- Avery, R.L.; Bakri, S.J.; Blumenkranz, M.S.; Brucker, A.J.; Cunningham, E.T., Jr.; D’amico, D.J.; Dugel, P.U.; Flynn, H.W., Jr.; Freund, K.B.; Haller, J.A.; et al. Intravitreal Injection Technique And Monitoring: Updated Guidelines of an Expert Panel. *Retina* **2014**, *34*, S1–S18. [[CrossRef](#)] [[PubMed](#)]
- Campbell, R.J.; Bronskill, S.E.; Bell, C.M.; Paterson, J.M.; Whitehead, M.; Gill, S.S. Rapid Expansion of Intravitreal Drug Injection Procedures, 2000 to 2008 A Population-Based Analysis. *Arch. Ophthalmol.* **2010**, *128*, 359–362. [[CrossRef](#)] [[PubMed](#)]
- Agarwal, A.; Nagpal, M. Intravitreal moxifloxacin injections in acute post-cataract surgery endophthalmitis: Efficacy and safety. *Indian J. Ophthalmol.* **2021**, *69*, 326–330. [[CrossRef](#)]
- Roblain, Q.; Louis, T.; Yip, C.; Baudin, L.; Struman, I.; Caolo, V.; Lambert, V.; Lecomte, J.; Noël, A.; Heymans, S. Intravitreal injection of anti-miRs against miR-142-3p reduces angiogenesis and microglia activation in a mouse model of laser-induced choroidal neovascularization. *Aging* **2021**, *13*, 12359–12377. [[CrossRef](#)]
- DeSouza, P.; Nidamarthi, D.; Moshiri, A.; Yiu, G.; Park, S.S.; Emami-Naeini, P. Effect of intravitreal steroid injection or implant on visual and imaging outcomes in patients with non-infectious uveitis. *Investig. Ophthalmol. Vis. Sci.* **2021**, *62*, 732.
- Januschowski, K.; Boden, K.T.; Szurman, P.; Stalmans, P.; Siegel, R.; Pérez Guerra, N.; Becker, S.L.; Rickmann, A.; Bisorca-Gassendorf, L. Effectiveness of immediate vitrectomy and intravitreal antibiotics for post-injection endophthalmitis. *Graefes Arch. Clin. Exp. Ophthalmol.* **2021**, *259*, 1609–1615. [[CrossRef](#)]
- Scott, I.U.; Luu, K.M.; Davis, J.L. Intravitreal antivirals in the management of patients with acquired immunodeficiency syndrome with progressive outer retinal necrosis. *Arch. Ophthalmol.* **2002**, *120*, 1219–1222.
- Zhuang, H.; Ding, X.; Zhang, T.; Chang, Q.; Xu, G. Vitrectomy combined with intravitreal antifungal therapy for posttraumatic fungal endophthalmitis in eastern China. *BMC Ophthalmol.* **2020**, *20*, 435. [[CrossRef](#)]
- Cox, J.; Elliott, D.; Sobrin, L. Inflammatory Complications of Intravitreal Anti-VEGF Injections. *J. Clin. Med.* **2021**, *10*, 981. [[CrossRef](#)]
- Heimann, H. Intravitreal Injections: Techniques and Sequelae. In *Medical Retina*; Holz, F.G., Spaide, R.F., Eds.; Springer: Berlin/Heidelberg, Germany, 2007; pp. 67–87.
- Jalil, A.; Chaudhry, N.L.; Gandhi, J.S.; Odat, T.M.; Yodaiken, M. Inadvertent injection of triamcinolone into the crystalline lens. *Eye* **2007**, *21*, 152–154. [[CrossRef](#)] [[PubMed](#)]
- Han, J.; Davids, J.; Ashrafian, H.; Darzi, A.; Elson, D.S.; Sodergren, M. A systematic review of robotic surgery: From supervised paradigms to fully autonomous robotic approaches. *Int. J. Med. Robot. Comput. Assist. Surg.* **2021**, *18*, e2358. [[CrossRef](#)] [[PubMed](#)]
- Kisinde, S.; Hu, X.B.; Hesselbacher, S.; Lieberman, I.H. The predictive accuracy of surgical planning using pre-op planning software and a robotic guidance system. *Eur. Spine J.* **2021**, *30*, 3676–3687. [[CrossRef](#)] [[PubMed](#)]
- Topsakal, V.; Matulic, M.; Assadi, M.Z.; Mertens, G.; Van Rompaey, V.; Van de Heyning, P. Comparison of the Surgical Techniques and Robotic Techniques for Cochlear Implantation in Terms of the Trajectories Toward the Inner Ear. *J. Int. Adv. Otol.* **2020**, *16*, 3–7. [[CrossRef](#)]
- Sauvée, M.; Poignet, P.; Dombre, E. Ultrasound image-based visual servoing of a surgical instrument through nonlinear model predictive control. *Int. J. Robot. Res.* **2008**, *27*, 25–40. [[CrossRef](#)]
- Li, Q.; Du, Z.; Yu, H. Grinding trajectory generator in robot-assisted laminectomy surgery. *Int. J. Comput. Assist. Radiol. Surg.* **2021**, *16*, 485–494. [[CrossRef](#)]
- Cornella, K.N.; Palafox, B.A.; Razavi, M.K.; Loh, C.T.; Markle, K.M.; Openshaw, L.E. SAVI SCOUT as a Novel Localization and Surgical Navigation System for More Accurate Localization and Resection of Pulmonary Nodules. *Surg. Innov.* **2019**, *26*, 469–472. [[CrossRef](#)]
- Braun, D.; Yang, S.; Martel, J.N.; Riviere, C.N.; Becker, B.C. EyeSLAM: Real-time simultaneous localization and mapping of retinal vessels during intraocular microsurgery. *Int. J. Med. Robot. Comput. Assist. Surg.* **2018**, *14*, e1848. [[CrossRef](#)]
- Yang, S.; Martel, J.N.; Lobes, L.A.; Riviere, C.N. Techniques for robot-aided intraocular surgery using monocular vision. *Int. J. Robot. Res.* **2018**, *37*, 931–952. [[CrossRef](#)]

21. Becker, B.C.; MacLachlan, R.A.; Lobes, L.A.; Hager, G.D.; Riviere, C.N. Vision-Based Control of a Handheld Surgical Micromanipulator With Virtual Fixtures. *IEEE Trans. Robot.* **2013**, *29*, 674–683. [[CrossRef](#)]
22. Hutchinson, S.; Hager, G.D.; Corke, P.I. A tutorial on visual servo control. *IEEE Trans. Robot. Autom.* **1996**, *12*, 651–670. [[CrossRef](#)]
23. Allen, M.; Westcoat, E.; Mears, L. Optimal Path Planning for Image Based Visual Servoing. *Procedia Manuf.* **2019**, *39*, 325–333. [[CrossRef](#)]
24. Dille, J.; Camara, M.; Omar, I.; Carter, A.; Pratt, P.; Vale, J.; Darzi, A.; Mayer, E.K. Evaluating the impact of image guidance in the surgical setting: A systematic review. *Surg. Endosc.* **2019**, *33*, 2785–2793. [[CrossRef](#)] [[PubMed](#)]
25. Anderson, P.; Wu, Q.; Teney, D.; Bruce, J.; Johnson, M.; Sünderhauf, N.; Reid, I.; Gould, S.; van den Hengel, A. Vision-and-Language Navigation: Interpreting visually-grounded navigation instructions in real environments. In Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Salt Lake City, UT, USA, 18–22 June 2018.
26. Mei, H.; Yang, X.; Wang, Y.; Liu, Y.; Lau, R. Don't Hit Me! Glass Detection in Real-World Scenes. In Proceedings of the 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Seattle, WA, USA, 14–19 June 2020.
27. Molnár, C.; Nagy, T.D.; Elek, R.N.; Haidegger, T. Visual servoing-based camera control for the da Vinci Surgical System. In Proceedings of the 2020 IEEE 18th International Symposium on Intelligent Systems and Informatics (SISY), Subotica, Serbia, 17–19 September 2020; pp. 107–112.
28. Hynes, P.; Dodds, G.I.; Wilkinson, A.J. Uncalibrated visual-servoing of a dual-arm robot for surgical tasks. In Proceedings of the 2005 International Symposium on Computational Intelligence in Robotics and Automation, Espoo, Finland, 27–30 June 2005; pp. 151–156.
29. Ullrich, F.; Michels, S.; Lehmann, D.; Roel, S.P.; Becker, M.; Bradley, J.N. Assistive Device for Efficient Intravitreal Injections. *Ophthalmic Surg. Lasers Imaging Retin.* **2016**, *47*, 752–762. [[CrossRef](#)]
30. Haidegger, T. Autonomy for Surgical Robots: Concepts and Paradigms. *IEEE Trans. Med. Robot. Bionics* **2019**, *1*, 65–76. [[CrossRef](#)]
31. Yang, G.-Z.; Cambias, J.; Cleary, K.; Daimler, E.; Drake, J.; Dupont, P.E.; Hata, N.; Kazanzides, P.; Martel, S.; Patel, R.V.; et al. Medical robotics—Regulatory, ethical, and legal considerations for increasing levels of autonomy. *Sci. Robot.* **2017**, *2*, eaam8638. [[CrossRef](#)]
32. Maillard, E.P.; Gueriot, D. Ieee, RBF neural network, basis functions and genetic algorithm. In Proceedings of the 1997 IEEE International Conference on Neural Networks, Houston, TX, USA, 12 June 1997; Volume 1–4.
33. Fedorenko, F.; Usilin, S. Real-time object-to-features vectorisation via Siamese neural networks. In Proceedings of Ninth International Conference on Machine Vision (ICMV 2016), Nice, France, 18–20 November 2016.
34. Hajiloo, A.; Keshmiri, M.; Xie, W.-F.; Wang, T.-T. Robust On-Line Model Predictive Control for a Constrained Image Based Visual Servoing. *IEEE Trans. Ind. Electron.* **2015**, *63*, 2242–2250. [[CrossRef](#)]
35. Zhang, J.; Liu, D. Online Estimation of Image Jacobian Matrix Based on Robust Information Filter. *J. Xi'an Univ. Technol.* **2011**, *27*, 133–138.
36. Corke, P. Vision-Based Control. In *Robotics, Vision and Control: Fundamental Algorithms in MATLAB®*; Corke, P., Ed.; Springer: Berlin/Heidelberg, Germany, 2011; pp. 455–479.
37. Fazekas, Z.; Lócsi, L.; Soumelidis, A.; Schipp, F.; Németh, Z. Rational Zernike Functions Capture the Rotations of the Eye-Ball. In *Progress in Industrial Mathematics at ECMI 2018*; Springer International Publishing: Cham, Switzerland, 2019; pp. 215–221. [[CrossRef](#)]
38. Poonguzhal, N.; Ezhilarasa, M. Identification Based on Iris Geometric Features. *J. Appl. Sci.* **2015**, *15*, 792–799. [[CrossRef](#)]
39. Masek, L. Recognition of Human Iris Patterns for Biometric Identification. Master's Thesis, University of Western Australia, Crawley, WA, Australia, 2003.
40. Pierce, J.E.; Clementz, B.A.; McDowell, J.E. Saccades: Fundamentals and Neural Mechanisms. In *Eye Movement Research: An Introduction to Its Scientific Foundations and Applications*; Klein, C., Ettinger, U., Eds.; Springer International Publishing: Cham, Switzerland, 2019; pp. 11–71.
41. Green-Simms, A.E.; Ekdawi, N.S.; Bakri, S.J. Survey of Intravitreal Injection Techniques Among Retinal Specialists in the United States. *Am. J. Ophthalmol.* **2011**, *151*, 329–332. [[CrossRef](#)]
42. Haeussler-Sinangin, Y.; Kohnen, T. Corneal Diameter. In *Encyclopedia of Ophthalmology*; Schmidt-Erfurth, U., Kohnen, T., Eds.; Springer: Berlin/Heidelberg, Germany, 2018; pp. 521–522.
43. Colan, J.; Nakanishi, J.; Aoyama, T.; Hasegawa, Y. Optimization-Based Constrained Trajectory Generation for Robot-Assisted Stitching in Endonasal Surgery. *Robotics* **2021**, *10*, 27. [[CrossRef](#)]
44. Zouaoui, R.; Mekki, H. 2D visual servoing of wheeled mobile robot by neural networks. In Proceedings of the 2013 International Conference on Individual and Collective Behaviors in Robotics (ICBR), Sousse, Tunisia, 15–17 December 2013; pp. 130–133.
45. Wang, F.; Sun, F.; Zhang, J.; Lin, B.; Li, X. Unscented Particle Filter for Online Total Image Jacobian Matrix Estimation in Robot Visual Servoing. *IEEE Access* **2019**, *7*, 92020–92029. [[CrossRef](#)]
46. Salehian, M.; RayatDoost, S.; Taghirad, H.D. Robust unscented Kalman filter for visual servoing system. In Proceedings of the 2nd International Conference on Control, Instrumentation and Automation, Shiraz, Iran, 27–29 December 2011. [[CrossRef](#)]
47. Ren, X.; Li, H.; Li, Y. Online Image Jacobian Identification Using Optimal Adaptive Robust Kalman Filter for Uncalibrated Visual Servoing. In Proceedings of the 2017 2ND Asia-Pacific Conference on Intelligent Robot Systems (ACIRS), Wuhan, China, 16–18 June 2017.

48. Zhao, Q.; Zhang, L.; Chen, Y. Online estimation technique for Jacobian matrix in robot visual servo systems. In Proceedings of the 2008 3rd IEEE Conference on Industrial Electronics and Applications, Singapore, 3–5 June 2008; pp. 1270–1275.
49. Zhang, Y.-B. Unscented Kalman filter for on-line estimation of Jacobian matrix. *J. Comput. Appl.* **2011**, *31*, 1699–1702. [[CrossRef](#)]
50. Qian, J.; Su, J. Online estimation of image Jacobian matrix by Kalman-Bucy filter for uncalibrated stereo vision feedback. In Proceedings of the 2002 IEEE International Conference on Robotics and Automation (Cat. No. 02CH37292), Washington, DC, USA, 11–15 May 2002; Volume 1, pp. 562–567.
51. Gu, J.; Wang, H.; Pan, Y.; Wu, Q. Neural network based visual servo control for CNC load/unload manipulator. *Optik* **2015**, *126*, 4489–4492. [[CrossRef](#)]
52. Matter, E. Epipolar-kinematics relations estimation neural approximation for robotics closed loop visual servo system. In Proceedings of the 2010 2nd International Conference on Computer and Automation Engineering (ICCAE 2010), Singapore, 26–28 February 2010; Volume 5, pp. 441–445. [[CrossRef](#)]
53. Nakamura, Y.; Hanafusa, H. Inverse Kinematic Solutions With Singularity Robustness for Robot Manipulator Control. *J. Dyn. Syst. Meas. Control* **1986**, *108*, 163–171. [[CrossRef](#)]
54. Halír, R.; Flusser, J. Numerically stable direct least squares fitting of ellipses. In Proceedings of the 6th International Conference in Central Europe on Computer Graphics and Visualization, WSCG, Bory, Czech Republic, 9–13 February 1998; Volume 98, pp. 125–132.
55. Wang, F.; Liu, Z.; Chen, C.; Zhang, Y. Adaptive neural network-based visual servoing control for manipulator with unknown output nonlinearities. *Inf. Sci.* **2018**, *451–452*, 16–33. [[CrossRef](#)]
56. Loreto, G.; Yu, W.; Garrido, R. Stable visual servoing with neural network compensation. In Proceedings of the 2001 IEEE International Symposium on Intelligent Control (ISIC'01), Mexico City, Mexico, 5–7 September 2001; pp. 183–188.
57. Qiu, Z.; Wu, Z. Adaptive neural network control for image-based visual servoing of robot manipulators. *IET Control Theory Appl.* **2022**, *16*, 443–453. [[CrossRef](#)]
58. Chicco, D. Siamese Neural Networks: An Overview. In *Artificial Neural Networks. Methods in Molecular Biology*; Cartwright, H., Ed.; Springer: New York, NY, USA, 2020; Volume 2190, pp. 73–94.
59. Tokuda, F.; Arai, S.; Kosuge, K. Convolutional Neural Network-Based Visual Servoing for Eye-to-Hand Manipulator. *IEEE Access* **2021**, *9*, 91820–91835. [[CrossRef](#)]
60. Hemingway, E.G.; O'Reilly, O.M. Perspectives on Euler angle singularities, gimbal lock, and the orthogonality of applied forces and applied moments. *Multibody Syst. Dyn.* **2018**, *44*, 31–56. [[CrossRef](#)]
61. Hendrycks, D.; Gimpel, K. Gaussian error linear units (gelus). *arXiv* **2016**, arXiv:1606.08415.
62. Park, F.C.; Bobrow, J.E.; Ploen, S.R. A lie group formulation of robot dynamics. *Int. J. Robot. Res.* **1995**, *14*, 609–618. [[CrossRef](#)]
63. Zhuang, J.; Tang, T.; Ding, Y.; Tatikonda, S.C.; Dvornek, N.; Papademetris, X.; Duncan, J. Adabelief optimizer: Adapting stepsizes by the belief in observed gradients. *Adv. Neural Inf. Process. Syst.* **2020**, *33*, 18795–18806.
64. Paszke, A.; Gross, S.; Massa, F.; Lerer, A.; Bradbury, J.; Chanan, G.; Killeen, T.; Lin, Z.; Gimelshein, N.; Antiga, L.; et al. Pytorch: An imperative style, high-performance deep learning library. *Adv. Neural Inf. Process. Syst.* **2019**, *32*, 8026–8037.
65. Bartoszewicz, A. Discrete-time quasi-sliding-mode control strategies. *IEEE Trans. Ind. Electron.* **1998**, *45*, 633–637. [[CrossRef](#)]
66. Rohmer, E.; Singh, S.P.N.; Freese, M. V-REP: A versatile and scalable robot simulation framework. In Proceedings of the 2013 IEEE/RSJ International Conference on Intelligent Robots and Systems, Tokyo, Japan, 3–7 November 2013; pp. 1321–1326.
67. Sarker, A.; Sinha, A.; Chakraborty, N. On Screw Linear Interpolation for Point-to-Point Path Planning. In Proceedings of the 2020 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), Las Vegas, NV, USA, 25–29 October 2020; pp. 9480–9487.
68. Fabisch, A. pytransform3d: 3D Transformations for Python. *J. Open Source Softw.* **2019**, *4*, 1159. [[CrossRef](#)]
69. Maxim, A.; Lazar, C.; Burlacu, A.; Copot, C. Robotic visual servoing system based on SIFT features. In Proceedings of the 2012 16th International Conference on System Theory, Control and Computing (ICSTCC 2012), Sinaia, Romania, 12–14 October 2012; pp. 1–6.
70. Assa, A.; Janabi-Sharifi, F. Two DOF controller for decoupled image-based visual servoing. In Proceedings of the 2014 IEEE 27th Canadian Conference on Electrical and Computer Engineering (CCECE), Toronto, ON, Canada, 4–7 May 2014; pp. 1–6.
71. Reghenzani, F.; Massari, G.; Fornaciari, W. The Real-Time Linux Kernel: A Survey on PREEMPT_RT. *ACM Comput. Surv.* **2019**, *52*, 18. [[CrossRef](#)]