6DOF Needle Pose Estimation for Robot-Assisted Vitreoretinal Surgery

MINGCHUAN ZHOU[1], **XING HAO**[1], **ABOUZAR ESLAMI**[2], **KAI HUANG**[3,7], **CAIXIA CAI**[4], **CHRIS P. LOHMANN**[5], **NASSIR NAVAB**[6], **ALOIS KNOLL**[1], **AND M. ALI NASSERI**[5]

[1]Chair of Robotics, Artificial Intelligence and Real-Time Systems, Technical University of Munich, 85748 Munich, Germany
[2]Carl Zeiss Meditec AG, 81379 Munich, Germany
[3]School of Data and Computer Science, Sun Yat-sen University, Guangzhou 510006, China
[4]Institute for Infocomm Research(I2R), A*STAR, Singapore 138632
[5]Klinik und Poliklinik für Augenheilkunde, Klinikum rechts der Isar der, Technical University of Munich, 81675 Munich, Germany
[6]Computer Aided Medical Procedures and Augmented Reality, Technical University of Munich, 85748 Munich, Germany
[7]State Key Laboratory of Ophthalmology, Zhongshan Ophthalmic Center, Sun Yat-sen University, Guangzhou 510060, China

Corresponding author: Caixia Cai (ccxtum@gmail.com)

ABSTRACT Vitreoretinal (VR) surgery is typical microsurgery with delicate and complex surgical procedures. The vision-based navigation for robot-assisted VR surgery has not been fully exploited because of the challenges that arise from illumination, high precision, and safety assessments. This paper presents a novel method to estimate the 6DOF needle pose specifically for the application of robotic intraocular needle navigation using optical coherence tomography (OCT) volumes. The key ingredients of the proposed method are 1) 3D needle point cloud segmentation in OCT volume and 2) needle point cloud 6DOF pose estimation using a modified iterative closest point (ICP) algorithm. To address the former, a voting mechanism with geometric features of the needle is utilized to robustly segment the needle in OCT volume. Afterward, the CAD model of the needle point cloud is matched with the segmented needle point cloud to estimate the 6DOF needle pose with a proposed shift-rotate ICP (SR-ICP). This method is evaluated by the existing ophthalmic robot on ex-vivo pig eyes. The quantitative and qualitative results are evaluated and presented for the proposed method.

I. INTRODUCTION

The present situation of eye pathologies, which contribute to more than 280 million visual impairments [1], raises an increasing demand for the ophthalmic surgery. Vitreoretinal (VR) disease is the disease located at the posterior segment of the eye that is related to the vitreous body and the retina. The VR diseases include age-related macular degeneration, epiretinal membranes, diabetic eye disease, retinal detachments, etc., which make up 10% visual impairments. VR surgery is a typical ophthalmic surgery consisting of complex manual tasks, shown as in Fig. 1. Three tools: surgical tool, irrigation cannula, and light source [2] are inserted inside the eye through the trocars, which are placed on the sclera. The irrigation cannula is used for liquid injection to maintain appropriate intraocular pressure. The light

source is used to illuminate the intended area on the retina, allowing the planar view of the area obtained and analyzed by surgeons through the microscope. During the surgery, the insufficient intraocular illumination and the en-face view from microscopy limit the visual feedback for the surgeon.

Besides the inadequate visual feedback, the ophthalmologist also suffer from hand tremors, which are normally around $100~\mu m$ and also may increase with the aging. However the accuracy of VR surgery varies depending on the specific operation. For the internal limiting membrane (ILM) peeling, which removes a very thin and transparent acellular membrane on the surface of the retina, the surgeon can do it quite well manually, so this means that the $182~\mu m$ accuracy (the hand tremor RMS amplitude [3]) is enough. For the sub-retinal injection, the average thickness of retina is around $200~\mu m$, therefore, $20~\mu m$ would be an acceptable position accuracy. For the retinal vein cannulation (RVC), the ideal position accuracy would be $20~\mu m$, since the diameter of

---

The associate editor coordinating the review of this manuscript and approving it for publication was Qingsong Ai.
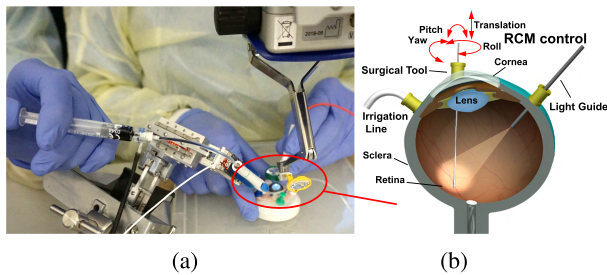
FIGURE 1. (a) The robot-assisted retinal surgery setup on the ex-vivo pig eye using iRAM!S robot [4]. The robot hold the instrument instead of the surgeon's hand. (b) The cross-section view of surgical setup. The movement of needle is restricted by RCM control where two translation movements are constrained.

branch retinal veins is typically less than 200 μm. Thus most of robot for VR surgery are designed to have the position accuracy of 10 μm to meet these surgical requirements [2].

The robot-assisted surgery (RAS) is a promising technology to help the surgeon to improve the treatment outcome, reduce the learning curve and extend the ability of surgeon in micro-operation. A series of ophthalmic robots have been developed to assist the surgeon to cope with their physical condition and extend their service life. In September 2016, surgeons at Oxford's John Radcliffe Hospital performed the world's first robotic Internal Limiting Membrane (ILM) peeling, using a Robotic Retinal Dissection Device (R2D2) [5], [6], which proved the feasibility of robot-assisted VR surgery. Among various ophthalmic surgery routines, the beveled needle is used to deliver the drug to the particular part of eye, e.g. the operation of subretinal injection and RVC. In robot-assisted ophthalmic surgery, the accurate 6DOF pose information of the needle is essential for image guided operation. However, it is inaccurate to estimate 6DOF pose of the needle tip part via forward kinematics due to the very thin, long and relatively flexible needle body [7].

Many researches have addressed on instrument pose estimation in 2D microscope images [8]. However, 2D images from single microscope are insufficient to estimate the 6DOF movements during the operation. Traditional navigation solutions such as optical tracking or electromagnetic tracking are not applicable as they usually have an accuracy in the range of 200 to 1400 μm. To overcome the above mentioned challenges, Probst *et al.* [9] first proposed to use a stereo-microscope to detect and localize the instrument in robot-assisted VR surgery, which can achieve a precision of 100 μm. However, VR surgery presents unique challenges, such as strong illumination changes, blur and even higher accuracy requirements, which limits the stereo-microscope method in some scenarios, e.g. subretinal injection and RVC. Optical Coherence Tomography (OCT), which is originally used for diagnosis of ophthalmic diseases because of its suitable resolution, has been developed to present real-time image interactions between the surgical tool and intraocular tissue. Microscope-mounted intraoperative OCT solution developed by Carl Zeiss Meditec (*RESCAN* 700), firstly

described in clinical use in 2014 [10], can share the same optical path with the microscope and give real-time cross section information of the target scan area, which is an ideal imaging modality for ophthalmic surgery. Taking advantage of OCT image modality, we have the chance to obtain the 3D scan of target area intraoperatively.

Estimating 6DOF pose of an object from incomplete point cloud has drawn much attention in computer vision with significant applications. Kehl *et al.* [11] introduced a light-weight 3D tracking with 6DOF pose estimation. However, their method cannot be directly applied in our study since the iterative closest point (ICP) with 6DOF parameters heavily relies on the initial guess from the object viewpoint features, e.g. clustered viewpoint feature histogram (CVFH), and the geometrical feature of the needle is cylindrical and has a bevel shape at the needle tip without strong features. This will lead to a local optima result and may not be suitable for safety-critical surgical applications.

In this paper, we introduce a modified iterative closest point to estimate the 6DOF pose of the needle directly from the OCT volume data. The main premise for the proposed method is that the actual dimensions of needle are within the range that obeys the standard for manufacturing medical devices (ISO 9626:2016), typical ±6.4 μm in diameter and $\pm1°$ in bevel angle. The method consists of two main parts. The first part is a robust needle segmentation method introduced to get the 3D needle point cloud in OCT volume. Due to the infrared light source of OCT and geometrical feature of the needle, the segmentation result will be robust to illumination variation and speck reflection. The second part is a shift-rotate ICP (SR-ICP) to estimate the 6DOF pose of the segmented needle point cloud. Using the geometrical features of the needle, the 6DOF pose is reduced to a 2DOF optimization problem, which can dramatically decrease the chance of getting local optima. Furthermore, different from the typical methods which use object viewpoint features to start the initial guess, we propose to align the CAD model tip to the visual needle tip in the OCT volume. This initial guess is very close to the global optimum. To validate the proposed method, we compare the result with brutal grid search (GS) and standard ICP. The experiment performed on the ex-vivo pig eyes demonstrated that the proposed method is qualified in 6DOF needle pose estimation for the VR surgery application. Especially, the position accuracy can be controlled within 10 μm with 95% confidence, which meets the most of the surgical requirements.

The rest of this paper is organized as follows: Section II briefly presents related work. The proposed method is introduced in Section III. Then, Section IV gives the experimental evaluations to prove the effectiveness of our proposed method. Section V concludes this paper with discussion and future work.

II. RELATED WORK

Many studies have been carried out with significant progress in the needle tracking through microscopic images [12]–[14].
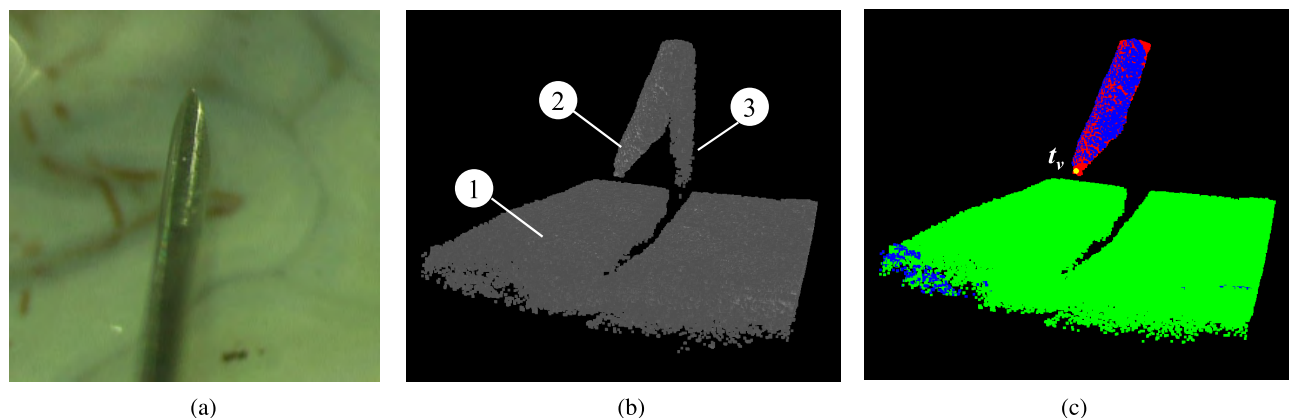
FIGURE 2. (a) The needle placed above the retina in microscope image. (b) The 3D OCT volume with needle and retina inside (① denotes retina, ② denotes needle, and ③ denotes needle reflection). (c) 2D detected ellipses are colored in blue and used for voting of needle segmentation. The red cluster is the voted needle cluster which is the ② in the OCT volume. The red point show up only when it is not overlapped by the blue point. The green cluster is the retina. Some blue points on the retina cluster means that there are votes in wrong cluster, however, the needle cluster will get the most of the votes. t_v denotes the visual needle tip.

These work obtained satisfactory results using either color-based or geometry-based features. However, due to the limitation of single 2D microscopic images, these methods can not provide enough information to estimate 6DOF needle pose.

Other 3D medical imaging technologies, including computed tomography (CT) scans, fluoroscopy, magnetic resonance (MR) and ultrasound technology are already applied in the cardiac, brain, and thoracic surgeries, not only for diagnostic procedures but also as a real time surgical guidance [15]–[17]. However, these imaging technologies can hardly achieve the ideal resolution for application of ophthalmic surgery. For MRI-guided interventions with resolution in millimeters in breast and prostate biopsies, 18 gauge needle with the diameter of 1.27 *mm* is used, while for ophthalmic surgery, 30 gauge needle, which has the diameter of 0.31 *mm*, requires resolution of submillimeter [18].

Some researchers used OCT for estimation of distance between the surgical tool and eye tissue. Song *et al.* [19] developed a robotic-surgical tool with an OCT probe integrated to estimate the one-dimensional distance between tool tip and tissue. Yu *et al.* [20], and Liu *et al.* [21] applied the OCT probe to assist robotic ophthalmic surgery. However, they focused on integrating the OCT into the surgical system, rather than estimating the instrument pose. Weiss *et al.* [22] developed an algorithm to track the needle 5DOF pose without needle rotation information. Gessert *et al.* [23] proposed a 3D convolutional neural network (CNN) to segment and estimate the pose of a small marker geometry from OCT volumes directly with a mean error of 14.89 ± 9.3 μm. They used a special marker instead of the needle with more geometrical features. The benefit of deep learning is that it can archive the end-to-end pose estimation while reducing the progress of geometrical modeling. However, the drawback is that a big amount of data is required to get a well-trained network,

meanwhile, the different type of needles, e.g. with different diameter, contributes to an even more data set preparation and network tuning overhead.

In contrast to deep learning approach, which is based on the geometric features of beveled needle and its imaging characteristic in OCT cube, in this paper, a modified ICP algorithm is proposed and evaluated to obtain the 6DOF pose of needle for ophthalmic surgery.

III. METHOD
The method section contains two parts. The first part is to segment the needle point cloud from the scan area. In the second part, we propose the segmented needle point cloud for needle pose estimation.

A. NEEDLE SEGMENTATION
The OCT 3D volume is obtained by setting the OCT engine into C-scan mode. One C-scan is constituted by multi B-scan gray images that are two dimensional, cross-sectional images of the scan area, as shown in Fig. 5. Due to the fact that one needle is sufficient to perform most the movement of surgery and multi needles simultaneous operation are avoided to reduce the incision ports on the sclera, there is only one needle at most in the scan view. As soon as the operation area is targeted by surgeon, the scan area could be fixed and cover the operation area. The needle will be controlled moving in this area. Afterwards, the captured original B-scan gray image is transformed into a binary image by the adaptive thresholding method [24]. We eliminate the noise inside the binary image by applying a median filter and a Gaussian filter. A voting mechanism is used to specify whether a point in the B-scan image belongs to the needle body or not. Since the needle body part in the B-scan image is a half ellipse, which can be considered as a strong feature, the ellipse fitting is applied to the topmost contours of each B-scan [24]. In order to filter any fitted ellipse E_i other than the needle body,
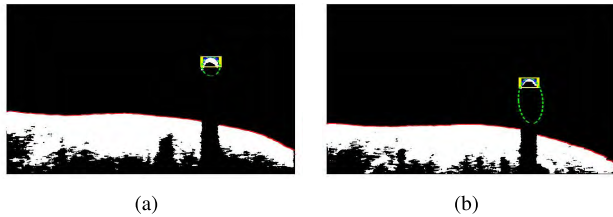
FIGURE 3. (a) The correct fitting of ellipse. (b) The fitting of ellipse with large deviation. The fitting of ellipse (green dash line) and bounding box (yellow solid line). The correct fitting ellipse is tangent with the bounding box.

we constrain the ellipse's minor axis m_e to a value less than m_t, where m_e is related to the diameter of the needle. The contour with fitted ellipse E_i can also be overlapped by a bounding box $B = (Bx_i, Bz_i, w_i, h_i)$, as shown in Fig. 3, where Bz_i and By_i are the left corner of the bounding box and w_i and h_i are the width and height of the bounding box, respectively. The information of bounding box location can be used to remove the needle reflection part in the original OCT volume. The needle reflection part is generated because needle reaches out of OCT imaging range, while blocking the imaging path as shown in Fig. 2(b). The E_i will also be filtered if the upper edge of corresponding bounding box B_w reaches close to the image range. The E_i is represented as,

$$E_i = \begin{cases} \varnothing, & \text{if } m_e \geq m_t \text{ or } \overline{Z} - Bz_i \leq \mu \\ E_i, & \text{if } m_e < m_t \text{ or } \overline{Z} - Bz_i > \mu \end{cases} \quad (1)$$

where $\overline{Z}$ is the OCT imaging range in Z direction, μ is the tolerance between the upper edge of corresponding bounding box and image range in Z direction.

The OCT volume is represented as a point cloud structure P (see Fig. 2(b)) with (x_i, y_i, z_i, b_i), where (x_i, y_i, z_i) is the position of point from P, and b_i is the boolean value to identify whether the point belongs to the needle or not. Afterwards, we differentiate between objects (needle and retina) in the point cloud with Euclidean cluster extraction [25]. In this way we can obtain a set of Euclidean point clusters. The cluster that has the greatest number of points with $b_i = 1$ (i.e. most voting, blue color in Fig. 2(c)) is treated as the needle. To ensure that the visual needle tip is located in the first B-scan of the segmented needle, the B-scan direction can be adjusted manually or automatically to match with the needle insertion direction [24]. The yellow point (shown in Fig. 2(c)), which represents the needle visual tip t_v, is the centroid of the needle cluster in this slice.

B. NEEDLE 6DOF POSE ESTIMATION

The part of needle tip can be treated as a rigid body and the needle pose in OCT can be shown as in Fig. 5(a). Then, The 6DOF of the needle can be defined by the needle tip point position, the needle center axis direction n and the rotation angle φ. In order to localize the space line L, which is the center axis of the needle, the center of the ellipse e_i in each B-scan needs to be calculated. The straightforward way is to fit the needle pixel with a ellipse equation. However,
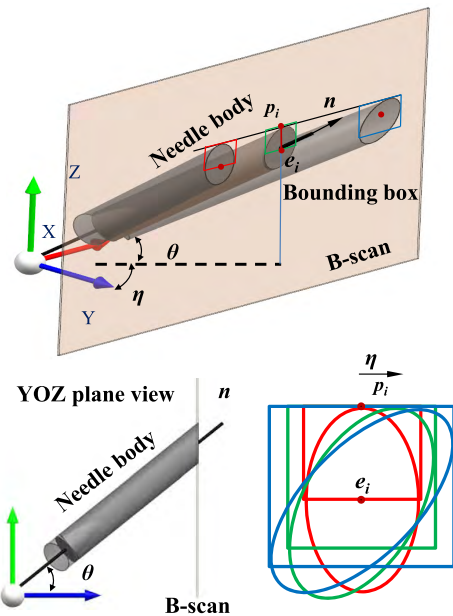
**FIGURE 4.** The geometrical relationship of the ellipse and the bounding box.

this half ellipse contains noise, which could easily lead to a large deviation from the ground truth, see Fig. 3. Therefore, we investigate the geometrical relationship of the ellipse and the bounding box, which is robust and has less computation overhead. As shown in Fig. 4, η is the angle between n and YOZ, and θ is the angle between n and XOY. From the YOZ plane view, we can find that with η changing, the distance between p_i and e_i remains the same, where e_i is the center of ellipse and p_i is the mid-point for the top border of bounding box. The distance of $p_i e_i$ can be obtained by the known needle diameter d and θ. Then we can calculate the e_i as follows,

$$e_i = (Bx_i, By_i, Bz_i - \frac{d}{2\cos\theta}) \quad (2)$$

To analyze the needle pose in 3D OCT volume, we can refer to Fig. 5(a). Considering that the target needle point cloud has a fixed axis L, which we could calculate from e_i by RANSAC line fitting, the source needle point cloud can be slid along and rotated around L. Therefore, the needle pose can be determined by L, δ and φ, where δ is the shift distance along this axis, and φ is the rotation angle around this axis.

In order to make the needles in OCT and the CAD model as the same length while keep most of the point information (bevel part), which helps to determine the 6DOF pose, we crop the OCT point cloud to have a consistent length with the CAD model by using split plane Γ, shown in Fig. 5(a). The CAD needle point cloud and OCT needle point cloud are respectively denoted as $C = \{c_i\}_{i=1}^{N_c} \in \mathbb{R}^3$ and $O = \{o_i\}_{i=1}^{N_o} \in \mathbb{R}^3$. The cropped OCT needle point cloud $O' = \{o'_i\}_{i=1}^{N_{o'}} \in \mathbb{R}^3 \subseteq O$ is cropped by Γ, which can be calculated as:

$$\Gamma : n_x(x - x_v - ln_x) + n_y(y - y_v - ln_y)$$
$$+ n_z(z - z_v - ln_z) = 0 \quad (3)$$
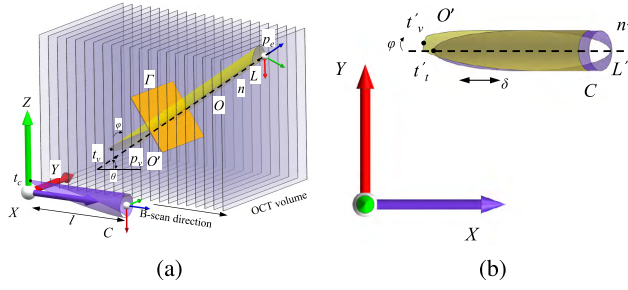
FIGURE 5. (a) The illustration of needle pose in OCT volume. (b) The initial guess of SR-ICP is shifted and rotated value of *C* to match t'_v and t'_t on the projection.

where $n = (n_x, n_y, n_z)$ denotes the unit direction vector for space line L, $t_v = (x_v, y_v, z_v)$ denotes the visual needle tip in the OCT needle point cloud and l denotes the length of needle for matching. l is selected by minimum value between the length of bevel part of needle in CAD model and the length of needle in OCT volume, which can be calculated as,

$$l = \min(l_c, l_o) \begin{cases} l_c = d/\cos(\beta) \\ l_o = \|p_v p_e\| - d\tan\theta/2 \end{cases} \quad (4)$$

where p_v is the projection point of t_v on the center axis line L, and p_e is the intersection point of line L and the last B-scan plane. p_v is calculated by,

$$p_v = p_a + \frac{p_a t_v \cdot p_a p_b}{p_a p_b \cdot p_a p_b} \times p_a p_b \quad (5)$$

where p_a and p_b are any two different points on the line L. t_v is obtained from needle segmentation.

As previously mentioned, we obtain the adjusted target needle point cloud O' from OCT images and the source point cloud C from the CAD model, as well as the fixed needle center axis L. In the following, we introduce the SR-ICP with 2DOF (shift along and rotate around L). We achieve the 2DOF by constraining the center axis of C by the center axis of O'. The proposed matching algorithm is consecutively minimizing the following functions,

$$C(i) = \underset{j \in \{1,\dots,N_c\}}{\operatorname{argmin}} \|(A_{k-1}c_i + t_{k-1}) - o'_i\|_2^2,$$
$$\text{for all } i \in \{1,\dots,N_{o'}\} \quad (6)$$

$$(A_k, t_k) = \underset{A,t}{\operatorname{argmin}} \frac{1}{N_{o'}} \sum_{i=1}^{N_{o'}} \|(Ac_i + t) - c_i\|_2^2 \quad (7)$$

$$s.t. \begin{cases} A = (\cos\varphi)I + (\sin\varphi)[n]_\times \\ \quad\quad +(1-\cos\varphi)(n \otimes n), \\ t = p_v - \delta n \end{cases} \quad (8)$$

Equation 6 finds the correspondence $C(i)$ ($i = 1,\dots,N$ where N is the number of correspondence) between CAD and OCT point cloud. This equation 6 minimizes the Euclidean distance between the cropped needle points in OCT volume and the transformed points of the needle CAD model under

constrains of Equation 7. A and t are the affine and translational matrices. A_k and t_k are the desired affine and translational matrices at iteration k. Equation 8 is the constraint for angle φ rotating around L and shifting distance δ along L. The rotation is clockwise and the shift is starting from the visual tip, p_v. I is the identity matrix. $[n]_\times$ is the cross product matrix of n, which can be calculated as,

$$[n]_\times = \begin{bmatrix} 0 & -n_z & -n_y \\ n_z & 0 & -n_x \\ -n_y & n_x & 0 \end{bmatrix} \quad (9)$$

$\otimes$ is the tensor product, where can be calculated as,

$$n \otimes n = \begin{bmatrix} n_x^2 & n_x n_y & n_x n_z \\ n_x n_y & n_y^2 & n_y n_z \\ n_x n_z & n_y n_z & n_z^2 \end{bmatrix} \quad (10)$$

The Equation. 7 is typical hyperparameter optimization problem that can be solved by gradient-based and grid search methods. The grid search method is exhaustive searching through a manually specified subset of hyperparameter space, and it can avoid local minima, but at the cost of high computation overhead. Gradient-based method, by contrast, is much less in computation overhead but can lead to local minima.

The proper search space of hyperparameter and initial guess are critical for the performance of both methods. As shown in Fig. 5 (b), L', n', t'_t, and t'_v are the projection of L, n, t_t, and t_v on plane XOY, respectively. t_t is the needle CAD model tip after transformation from the original position. Due to the fact that the OCT scan has a certain resolution and the optical coherence feature would miss the needle tip point reflection in the OCT images, the visual needle tip t_v is not the actual position of real needle tip but the distance between t_v and the real needle tip is small. The projection of actual needle tip on XOY is also close to t'_v. Therefore, we could use t'_v as the initial guess, and φ and δ could be adjusted to make the two points t'_v and t'_t overlapping with each other. The outputs of rotation angle and shift distance value are denoted as φ_0 and δ_0, which are the initial search values for SR-ICP. The search space can be box restricted with reference to the initial guess.

$$(\delta_0, \varphi_0) = \underset{\delta,\varphi}{\operatorname{argmin}} \|(At_t + t)_{XOY} - t'_v\|_2^2 \quad (11)$$

where $(At_t + t)_{XOY}$ donates the projection of $(At_t + t)$ on XOY plane. The Levenberg-Marquardt algorithm is introduced to solve Equation 11. The gradient-based method, which is combined with Levenberg-Marquardt algorithm (LM based SR-ICP) and grid search (GS based SR-ICP) method with several iterations, are programmed to solve Equation 7 to obtain the optimized A and t. The search area is constrained in a box area with $\delta \in [\delta_0 + \triangle\delta, \delta_0 - \triangle\delta]$, $\delta \in [\varphi_0 + \triangle\varphi, \varphi_0 - \triangle\varphi]$, where $\triangle\varphi$ and $\triangle\delta$ are the regulation parameter for adjusting the searching range.

With the desired output affine and translation value A_k and t_k, the estimated needle tip position can be calculated as,

$$t_t = A_k t_c + t_k \quad (12)$$
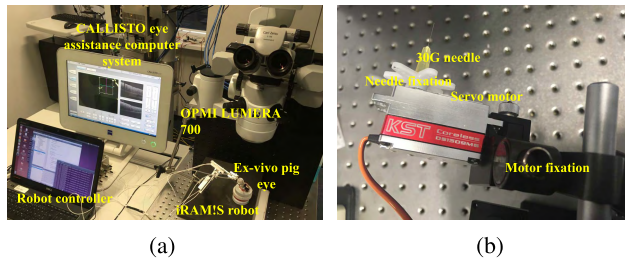
FIGURE 6. (a) The experiment setup with iRAM!S eye surgical robot. (b) The needle rotation experiment setup with a servo motor.

TABLE 1. Descriptive statistics for needle tip position error (in micron).

	AE	ME	RSME	0.95CI
$\bar{e}_G{}^1$	-3.1	18.7	6.7	[-14.6, 9.2]
$\bar{e}_L{}^1$	0.9	29.9	10.1	[-24.4, 21.6]
$\bar{e}_V{}^1$	2.2	34.8	10.9	[-16.1, 25.0]
$\bar{e}_G{}^3$	9.5	3.7	4.1	[-7.8, 1.5]
$\bar{e}_L{}^3$	0.8	15.2	5.0	[-9.6, 9.8]
$\bar{e}_V{}^3$	1.9	17.2	4.7	[-7.3, 10.9]
$\bar{e}_G{}^5$	-2.8	6.8	3.1	[-6.0, 0.3]
$\bar{e}_L{}^5$	1.9	9.4	3.4	[-3.9, 7.7]
$\bar{e}_V{}^5$	234.1	230.0	4.1	[-3.7, 9.6]

where t_c is the tip position for needle CAD model in the original position, which can be calculated by,

$$t_c = (d/2, 0, 0) \qquad (13)$$

IV. EXPERIMENTS AND RESULTS

The experimental setup is depicted in Fig. 6. The OCT engine is set to the maximum speed available with 27000 A-scans per second in cube resolution of $128 \times 512 \times 1024$ in the corresponding scan range of 3 mm×3 mm×2 mm. The movement of needle can be controlled by the surgical robot named iRAM!S [26] with accuracy of 1 μm with piezo motors (SmarACT GmbH, Germany), as shown in Fig. 6(a). Since there is no roll control of needle in the robot, we use a servo motor (Security GmbH, Germany) with 1° accuracy to estimate the performance of different rotation of needle, see Fig. 6(b). Due to the fact that one volume capture takes 500 ms, all the movements are proceeded by the robot until the volume capture is finished. The framework is implemented on the *CALLISTO* eye assistance computer system with Intel Core i7 CPU of 200 ms processing time for needle segmentation to obtain the needle point cloud with parallel programming in C++ with OpenCV 3.4 and Point Cloud Library (PCL) 1.8. The experiments contains two parts, (1) the performance analysis of SR-ICP (GS and LM based) and normal ICP (standard algorithm from PCL 1.8), and (2) the accuracy validation of needle movement and needle rotation for SR-ICP.

Fig. 7 shows needle point cloud matching results for the LM based SR-ICP. As we can see in Fig. 7(c), SR-ICP resulted in a better matching (white) compared to the normal ICP method (purple). The normal ICP trends to mismatch with the point cloud without the constraint of the needle center axis. Fig. 8 shows the performance of three methods with different point cloud sample size. As soon as the larger sample size is selected, the sparser point cloud is obtained. The sum of distance is the all over distance for the correspondence with the optimized A_k and $\tilde{t}_k$, which can be calculated as,

$$\sum_{i=1}^{N_{o'}} \|(A_k c_i + t_k) - c_i\|_2^2 \qquad (14)$$

We can see that the ML based SR-ICP and GS based SR-ICP perform is better than ICP in matching with smaller

sum of distances for corresponding points (see Fig. 8(a)). In Fig. 8(b), the distance between estimated and visual needle tip does not have significant difference, which means that the estimated needle tip would be similar for all three methods. When the sample size increases to 60 μm, the ICP trends to be far away from the visual tip than the ICP, so this indicates that the SR-ICP has a better performance when the point sample size is large.

Fig. 8(c) shows that LM based SR-ICP has the best computation performance. When the sample size is 10 μm, the computation time is 335.4 ms for the LM based SR-ICP, which is 4.9 and 18.6 times faster than the ICP and GS based SR-ICP, respectively.

In the second part of the experiment, we demonstrate the effectiveness of the proposed method. The point cloud sample size is set as 10 μm considering the algorithm processing speed. As the experiment setup in Fig. 6, the robot is controlled to move the needle in X, Y, and Z direction with 20 and 40 μm and each movement is repeated 25 times. The robot is also controlled to yaw and pitch around the needle tip as the RCM control point with virtual fixture control [27]. The yaw and rotation angle is set from $-45°$ to $45°$ with 15 ° as the step. The servo motor is controlled to rotate the needle from $-45°$ to $45°$ in 15° increments and each angle is collected 10 times. As we can see from Fig 9, the mean error of needle tip distance estimation is 3.1 μm (maximum 18.7 μm) for LM based SR-ICP, 2.4 μm for GS based SR-ICP (maximum 29.9 μm), and 2.6 μm for visual tip (maximum 34.8 μm). We can find that the mean error accuracy does not have significant difference, while the SR-ICP has the lower maximum value compared to visual tip method. The main reason is that the visual tip is directly obtained from the images which varies more depending on the resolution of the OCT scan in each direction. The standard deviation is 6.0 μm for GS based SR-ICP, 10.1 μm for LM based SR-ICP, and 10.7 μm for visual tip, which demonstrates that the GS based SR-ICP have the most steady output results. This verifies that the GS based SR-ICP is slow but more robust for single evaluation. In order to further analysis the performance of these method, we put all the movement error data in processing method with different evaluation metrics and list them in Table 1.

The evaluation metrics includes average error of estimated value to the ground truth (AE), the error of maximum
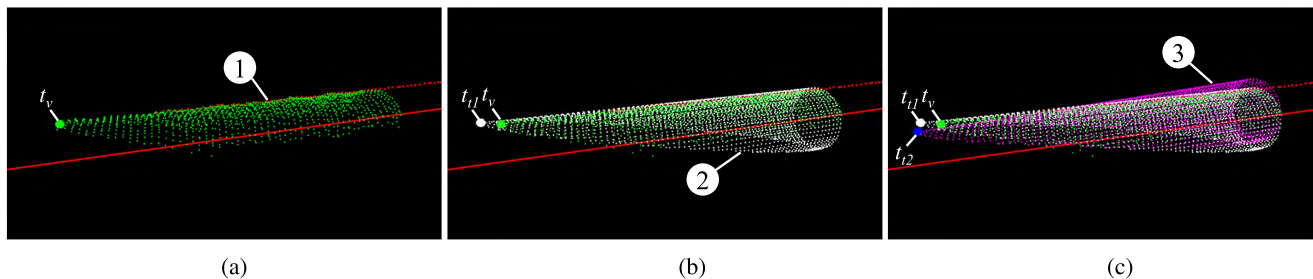
FIGURE 7. (a) The cropped OCT needle point cloud (① denotes needle point cloud with green color, the red points are the middle point for the top border of bounding box, and the red line is the fitted axis line *L*). (b) The LM based SR-ICP matching result (② denotes the transformed CAD needle point cloud using SR-ICP in white color). (c) The comparison of LM based SR-ICP matching and ICP result (③ denotes the transformed CAD needle point cloud using ICP in purple color). t_v denotes visual needle tip with white dot, t_{t1} and t_{t2} denotes the estimated needle tip via SR-ICP with green dot and ICP with blue dot, respectively.
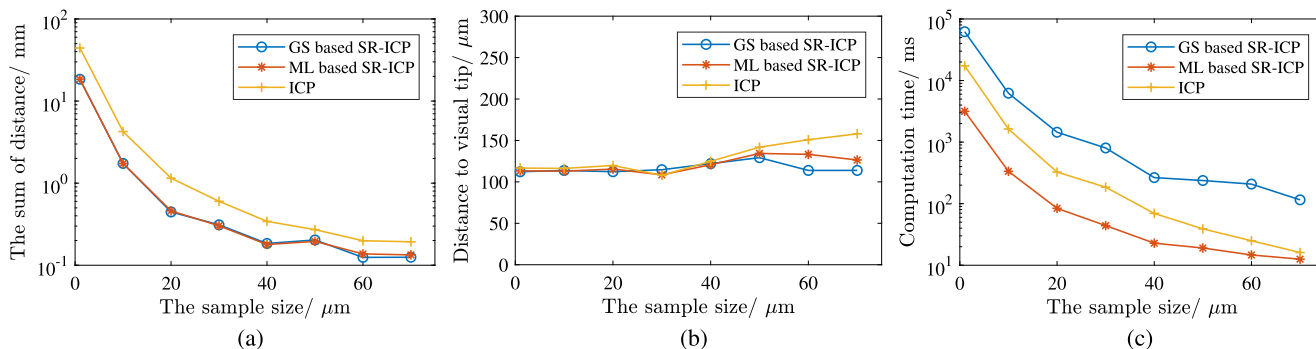
FIGURE 8. (a) Sum of distances between corresponding points calculated by three methods with change of sample size. (b) The distance between estimated and visual needle tip calculated by three methods with change of sample size. (c) The computation time by three methods with change of sample size.
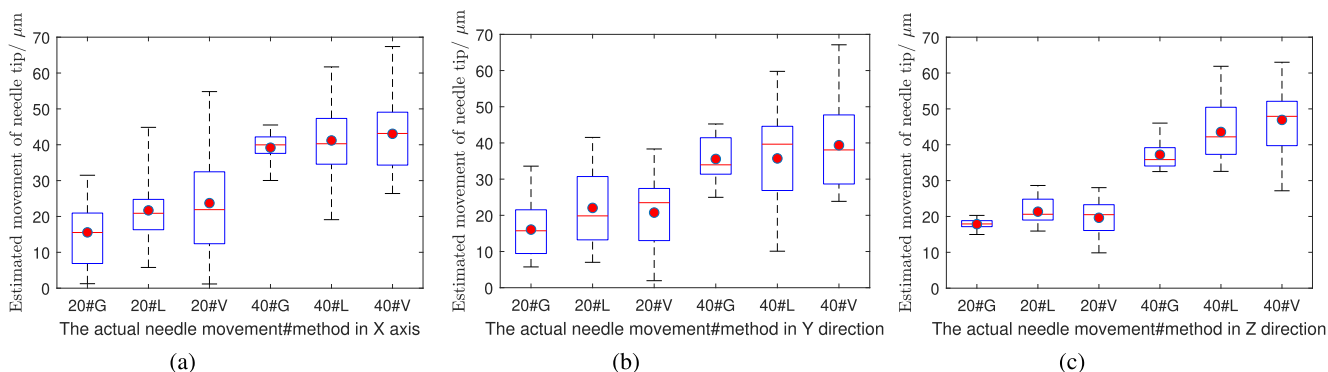

FIGURE 9. The actual movement for the needle tip and the estimated movement calculated by the Grid based SR-ICP, LM based SR-ICP, and visual tip with 20 and 40 μm in (a) X direction, (b) Y direction, and (c) Z direction. The whiskers show the minimum and maximum recorded change of the distance while the first and third quartile show the start and the end of the box. Band, red dot, and cross represent median, mean, and outliers of the recorded changes respectively.

estimated value and actual value (ME), the root mean square error (RSME),and the 95% of estimated confidence interval (0.95CI). $\bar{e}_G{}^n$, $\bar{e}_L{}^n$ and $\bar{e}_V{}^n$ denote average of randomly selected n estimated error values from 150 data samples using LM based SR-ICP, GS based SR-ICP, and visual tip, respectively. The 0.95CI for $\bar{e}_G{}^n$, $\bar{e}_L{}^n$ and $\bar{e}_V{}^n$ can be significantly narrowed down with the increasing of n. The advantage of the GS based SR-ICP method is that the processing speed is faster with the benefit of averaging more data in the same time

duration. Table 1 shows that with more data samples averaging, the accuracy of the needle position can be increased. When 3 data samples are averaged, 10 μm needle tip position accuracy can be achieved with confidence of 95% for GS based SR-ICP and LM based SR-ICP.

Fig. 10 shows the performance of needle rotation. The needle roll angle around L can be calculated by the methods of GS based SR-ICP and LM based SR-ICP. The yaw and pitch angle is calculated by the equation of line L, thus
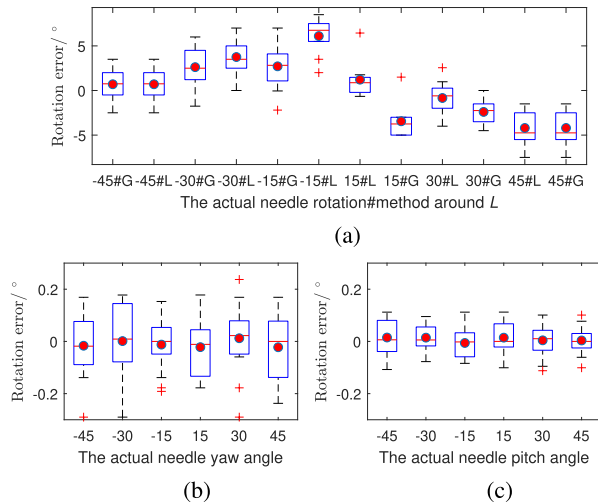
FIGURE 10. (a) Sum of distances between corresponding points calculated by three methods with different sample size. (b) The distance between estimated and visual needle tip calculated by three methods with change of sample size. (c) The computation time by three methods with change of sample size.

the result of yaw and pitch angle are the same for the GS based SR-ICP and LM based SR-ICP. The mean error of the estimated needle roll angle is $0.367°$ (maximum $7.5°$) and $0.0835°$ (maximum $8.5°$) for the GS based SR-ICP and LM based SR-ICP, respectively. The performance of roll angle estimation between two methods is similar with the consideration of the servo motor error is $1°$. The mean error is $0.010°$ (maximum $0.290°$) for yaw angle, and $0.008°$ (maximum $0.112°$) for pitch angle. The above two parts of experiment demonstrate that the SR-ICP has a better estimation result compared to the normal ICP algorithm. There is no significant difference between the GS based SR-ICP, LM based SR-ICP, and visual tip regarding the performance for relative position accuracy of the needle tip. However, visual tip method only provide needle tip information without needle rotation information. LM based SR-ICP have the advantage of similar optima output ability with GS based SR-ICP but much less computational overhead.

V. CONCLUSION

This paper presents a novel approach to estimate the 6DOF needle pose directly from the OCT volume. A SR-ICP algrithm is proposed to estimated the incompleted needle point cloud after the segmentation of needle from OCT volume. Two solving strategies are used to solve the optimum problem that is GS based SR-ICP and LM based SR-ICP. Both of the methods have the ability of estimating the needle tip position in accuracy of 10 μm with confidence of 95%, which meets the positioning accuracy requirement for most surgical applications in VR surgery. The mean error of roll angle, yaw angle, and pitch angle are $0.0835°$, $0.010°$, and $0.008°$, respectively. Our future work will focus on path planning and trajectory design in a way that the drug can be delivered directly and precisely to the designed area with 6DOF needle navigation.

REFERENCES

[1] *Universal Eye Health: A Global Action Plan 2014–2019. Prevention of Blindness and Visual Impairment*, World Health Org., Geneva, Switzerland, 2013.

[2] T. Nakano, N. Sugita, T. Ueta, Y. Tamaki, and M. Mitsuishi, "A parallel robot to assist vitreoretinal surgery," *Int. J. Comput. Assist. Radiol. Surg.*, vol. 4, no. 6, pp. 517–526, 2009.

[3] C. N. Riviere and P. S. Jensen, "A study of instrument motion in retinal microsurgery," in *Proc. 22nd Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. (IEMBS)*, vol. 1, Jul. 2000, pp. 59–60.

[4] M. A. Nasseri et al., "Kinematics and dynamics analysis of a hybrid parallel-serial micromanipulator designed for biomedical applications," in *Proc. IEEE/ASME Int. Conf. Adv. Intell. Mechatron. (AIM)*, Jul. 2013, pp. 293–299.

[5] B. Gonenc, A. Chamani, J. Handa, P. Gehlbach, R. H. Taylor, and I. Iordachita, "3-DOF force-sensing motorized micro-forceps for robot-assisted vitreoretinal surgery," *IEEE Sensors J.*, vol. 17, no. 11, pp. 3526–3541, Jun. 2017.

[6] T. L. Edwards et al., "First-in-human study of the safety and viability of intraocular robotic surgery," *Nature Biomed. Eng.*, vol. 2, pp. 649–656, Jun. 2018.

[7] N. Rieke et al., "Real-time localization of articulated surgical instruments in retinal microsurgery," *Med. Image Anal.*, vol. 34, pp. 82–100, Dec. 2016.

[8] I. Laina et al., "Concurrent segmentation and localization for tracking of surgical instruments," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent.* Quebec, QC, Canada: Springer, 2017, pp. 664–672.

[9] T. Probst, K.-K. Maninis, A. Chhatkuli, M. Ourak, E. V. Poorten, and L. Van Gool, "Automatic tool landmark detection for stereo vision in robot-assisted retinal surgery," *IEEE Robot. Autom. Lett.*, vol. 3, no. 1, pp. 612–619, Jan. 2018.

[10] J. P. Ehlers, P. K. Kaiser, and S. K. Srivastava, "Intraoperative optical coherence tomography using the RESCAN 700: Preliminary results from the DISCOVER study," *Brit. J. Ophthalmol.*, vol. 98, no. 10, pp. 1329–1332, 2014.

[11] W. Kehl, F. Tombari, S. Ilic, and N. Navab, "Real-time 3D model tracking in color and depth on a single CPU core," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jul. 2017, pp. 745–753.

[12] Y. Li, C. Chen, X. Huang, and J. Huang, "Instrument tracking via online learning in retinal microsurgery," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent.* Boston, MA, USA: Springer, 2014, pp. 464–471.

[13] N. Rieke et al., "Surgical tool tracking and pose estimation in retinal microsurgery," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent.* Munich, Germany: Springer, 2015, pp. 266–273.

[14] R. Sznitman, K. Ali, R. Richa, R. H. Taylor, G. D. Hager, and P. Fua, "Data-driven visual tracking in retinal microsurgery," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent.* Nice, France: Springer, 2012, pp. 568–575.

[15] N. McDannold, G. Clement, P. Black, F. Jolesz, and K. Hynynen, "Transcranial MRI-guided focused ultrasound surgery of brain tumors: Initial findings in patients," in *Proc. Int. Soc. Mag. Reson. Med.*, vol. 17, Stockholm, Sweden, 2009, p. 447.

[16] E. R. McVeigh et al., "Real-time interactive MRI-guided cardiac surgery: Aortic valve replacement using a direct apical approach," *Magn. Reson. Med.*, vol. 56, no. 5, pp. 958–964, Nov. 2006.

[17] G. J. Vrooijink, M. Abayazid, and S. Misra, "Real-time three-dimensional flexible needle tracking using two-dimensional ultrasound," in *Proc. IEEE Int. Conf. Robot. Automat.*, May 2013, pp. 1688–1693.

[18] T. T. Lam, P. Miller, S. Howard, and T. M. Nork, "Validation of a rabbit model of choroidal neovascularization induced by a subretinal injection of FGF-LPS," *Investigative Ophthalmol. Vis. Sci.*, vol. 55, no. 13, p. 1204, 2014.

[19] C. Song, P. L. Gehlbach, and J. U. Kang, "Active tremor cancellation by a 'smart' handheld vitreoretinal microsurgical tool using swept source optical coherence tomography," *Opt. Express*, vol. 20, no. 21, pp. 23414–23421, 2012.

[20] H. Yu, J.-H. Shen, K. M. Joos, and N. Simaan, "Design, calibration and preliminary testing of a robotic telemanipulator for OCT guided retinal surgery," in *Proc. IEEE Int. Conf. Robot. Automat.*, May 2013, pp. 225–231.

[21] X. Liu, M. Balicki, R. H. Taylor, and J. U. Kang, "Towards automatic calibration of Fourier-Domain OCT for robot-assisted vitreoretinal surgery," *Opt. Express*, vol. 18, no. 23, pp. 24331–24343, 2010.

[22] J. Weiss, N. Rieke, M. A. Nasseri, M. Maier, A. Eslami, and N. Navab, "Fast 5DOF needle tracking in iOCT," *Int. J. Comput. Assist. Radiol. Surg.*, vol. 13, no. 6, pp. 787–796, 2018.

[23] N. Gessert and M. Schlüter, and A. Schlaefer, "A deep learning approach for pose estimation from volumetric OCT data," *Med. Image Anal.*, vol. 46, pp. 162–179, May 2018.

[24] H. Roodaki, K. Filippatos, A. Eslami, and N. Navab, "Introducing augmented reality to optical coherence tomography in ophthalmic microsurgery," in *Proc. IEEE Int. Symp. Mixed Augmented Reality*, Sep./Oct. 2015, pp. 1–6.

[25] M. Zhou *et al.*, "Towards robotic eye surgery: Marker-free, online hand-eye calibration using optical coherence tomography images," *IEEE Robot. Autom. Lett.*, vol. 3, no. 4, pp. 3944–3951, Oct. 2018.

[26] M. A. Nasseri, M. Maier, and C. P. Lohmann, "A targeted drug delivery platform for assisting retinal surgeons for treating age-related macular degeneration (AMD)," in *Proc. 39th Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. (EMBC)*, Jul. 2017, pp. 4333–4338.

[27] M. A. Nasseri *et al.*, "Virtual fixture control of a hybrid parallel-serial robot for assisting ophthalmic surgery: An experimental study," in *Proc. 5th IEEE RAS/EMBS Int. Conf. Biomed. Robot. Biomech.*, Aug. 2014, pp. 732–738.

MINGCHUAN ZHOU received the B.Eng. degree in agricultural mechanization engineering from Northeast Agriculture University, in 2012, and the M.Eng. degree from Zhejiang University, China, in 2015. He is currently pursuing the Ph.D. degree with the Chair of Robotics, Artificial Intelligence and Real-time Systems, Department of Computer Science, Technical University of Munich. His research interests include the autonomous systems, deep learning, medical robotics, image processing, and agricultural robotics. He was a recipient for the Finalist of Best Paper Awards IEEE-ROBIO 2017.

XING HAO received the B.S. degree in mechanical engineering from the Hefei University of Technology, Anhui, China, in 2014. He is currently pursing master's degree in mechanical engineering with the Technical University of Munich. His research interest includes computer vision and robot.

ABOUZAR ESLAMI received the Ph.D. degree in electrical engineering from the Sharif University of Technology, Tehran, Iran, 2010. From 2011 to 2013, he was a Postdoctoral Researcher with the Chair for Computer Aided Medical Procedures and Augmented Reality, Technical University of Munich, Munich, Germany. He was a Project Manager of interventional imaging with Carl Zeiss Meditec AG, Germany, where he is currently the Head of the Translational Research Laboratory. His research interests include registration/visualization, segmentation, reconstruction, medical imaging, and machine learning for medical applications.

KAI HUANG received the B.Sc. degree from Fudan University, China, in 1999, the M.Sc. degree from the University of Leiden, The Netherlands, in 2005, and the Ph.D. degree from ETH Zürich, Switzerland, in 2010. He was a Research Group Leader with Fortiss GmbH, Munich, Germany, in 2011, and also a Senior Researcher with the Computer Science Department, Technical University of Munich, Germany, from 2012 to 2015. He joined Sun Yat-sen University as a Professor, in 2015. He was appointed as the Director of the Institute of Cyber Physical Systems, School of Data and Computer Science, in 2016. His research interests include techniques for the analysis, design, and optimization of embedded systems, particularly in the automotive domain. He was a recipient of the Best Paper Awards ESTIMedia 2013, SAMOS 2009, and General Chairs Recognition Award for Interactive Papers in CDC 2009.

CAIXIA CAI received the master's degree in informatics from China Agricultural University, in 2010, and the Ph.D. degree in computer science from Technical University of Munich, in 2017. She is currently with the Agency for Science, Technology and Research (A*STAR), Singapore, as a Scientist. Her research interests include robotics manipulation, vision-based control, machine learning for robotics, and human–robot interaction.

CHRIS P. LOHMANN studied medicine with Giessen University (Justus-Liebig-Universität) and LMU Munich. He received the Ph.D. degrees in Gieen, in 1991, and in London, in 1995, and the postdoctoral teaching qualification (habilitation) degree in Regensburg, in 2000. From 2004 to 2005, he served as an Acting Head of the Department of Ophthalmology, Klinikum rechts der Isar, TUM's University Hospital. Since 2002, he has also been an Adjunct Professor with the Eye Research Institute, University of Ottawa, Canada. Since 2005, he has been the Director of the Department of Ophthalmology, TUM, where he held a W3 professorship position. His research interests include corneal diseases and refractive surgery.

NASSIR NAVAB studied mathematics and physics, computer engineering and systems controll. He received the Ph.D. degree from INRIA, Paris XI. He then did two years of Postdoctoral Research at the MIT Media Laboratory, CA, USA. Prior to becoming a Full Professor at TUM, in 2003, he was a Distinguished Member of the Technical Staff at Siemens Corporate Research (SCR), Princeton, USA. In 2006, he became a Board Member of MICCAI, the organizer of the world's leading conference on medical image computing and computer assisted intervention. He has authored 100s of scientific publications and has filed more than 60 international patents. His research interests include computer-aided medical procedures, augmented reality, developing technologies to improve the quality of medical intervention, and bridges the gap between medicine and computer science. He is on the editorial board of many international journals, including the IEEE TMI, *MedIA*, and *Medical Physics*.

ALOIS KNOLL received the Diploma (M.Sc.) degree in electrical/communications engineering from the University of Stuttgart, Germany, in 1985, and the Ph.D. degree *(summa cum laude)* in computer science from the Technical University of Berlin, Germany, in 1988. He served on the Faculty of the Computer Science Department, TU Berlin, until 1993, when he qualified for teaching computer science at a university (habilitation). He was a Full Professor and the Director of the Computer Engineering Research Group, University of Bielefeld, until 1991, when he assumed the professorship of Real-Time Systems and Robotics, TUM. From 2001 to 2005, he was the Director of the Educational Robotics Department and a member of the Management Committee of the Fraunhofer Institute for Autonomous Intelligent Systems, Bonn. Since 2009, he has been the Director of Fortiss, an affiliated Institute of TUM, and the Director of the Graduate School of Information Science in Health (GSISH). His research interests include cognitive, medical and sensor-based robotics, multi-agent systems, data fusion, adaptive systems, multimedia information retrieval, model-driven development of embedded systems with applications to automotive software and electric transportation, and simulation systems for robotics and traffic.

M. ALI NASSERI received the Ph.D. degree from the Technical University of Munich, Germany, in 2015, and the Ph.D. degree in biomedical engineering with a joint degree from the Medical School, Mechanical Engineering Faculty and Informatics Faculty, TUM. He is currently the Head of the Medical Technology Laboratory, Ophthalmology Department, Klinikum rechts der Isar, which is the university hospital of the Technical University of Munich. He has coauthored more than 50 scientific publications, patents, and book chapters. His interests include surgical robotics, medical imaging, machine learning, and translation of these technologies to the healthcare systems. He is an Active Member of ARVO and DOG.

• • •