Toward Image-Guided Automated Suture Grasping Under Complex Environments: A Learning-Enabled and Optimization-Based Holistic Framework

Bo Lu, Bin Li, Wei Chen, Yueming Jin, *Member, IEEE*, Zixu Zhao, Qi Dou, *Member, IEEE*, Pheng-Ann Heng, *Senior Member, IEEE*, and Yunhui Liu, *Fellow, IEEE*

Abstract—To realize a higher-level autonomy of surgical knot tying in minimally invasive surgery (MIS), automated suture grasping, which bridges the suture stitching and looping procedures, is an important yet challenging task needs to be achieved. This paper presents a holistic framework with image-guided and automation techniques to robotize this operation even under complex environments. The whole task is initialized by suture segmentation, in which we propose a novel semi-supervised learning architecture featured with a suture-aware loss to pertinently learn its slender information using both annotated and unannotated data. With successful segmentation in stereo-camera, we develop a Sampling-based Sliding Pairing (SSP) algorithm to online optimize the suture's 3D shape. By jointly studying the robotic configuration and the suture's spatial characteristics, a target function is introduced to find the optimal grasping pose of the surgical tool with Remote Center of Motion (RCM) constraints. To compensate for inherent errors and practical uncertainties, a unified grasping strategy with a novel vision-based mechanism is introduced to autonomously accomplish this grasping task. Our framework is extensively evaluated from learning-based segmentation, 3D reconstruction, and image-guided grasping on the da Vinci Research Kit (dVRK) platform, where we achieve high performances and successful rates in perceptions and robotic manipulations. These results prove the feasibility of our approach in automating the suture grasping task, and this work fills the gap between automated surgical stitching and looping, stepping towards a higher-level of task autonomy in surgical knot tying.

Note to Practitioners—This paper aims to automate the suture grasping task in surgical knot tying by leveraging stereo visual guidance. To effectively robotize this procedure, it requires multidisciplinary knowledge to achieve suture segmentation, 3D shape reconstruction, and reliable automated grasping, while there are no existing works tackling this procedure especially using robots with RCM kinematics constraints and under complex environments. In this article, we propose a learning-driven method along with a 3D shape optimizer, which can conduct the suture segmentation and output its accurate spatial coordinates, serving as guidance for automated grasping operation. Apart from this, we introduce a unified function to optimize the grasping pose, and a vision-based grasping strategy is also proposed to intelligently complete this task. The experiments extensively validate the feasibility of our framework for automated suture grasp, and its successful completion can serve as a basis for the following looping manipulation, hence filling a step gap in robot-assisted knot tying. This framework can be also encapsulated into the medical robotic system, and by simply indicating (e.g. mouse click) the rough position of the suture's tip in one camera frame, the overall framework can be initialized and further accomplish the suture grasping task, which further prompts a full autonomy of surgical knot tying in the near future.

I. INTRODUCTION

WITH enhanced capability in precision, dexterity, and maneuverability, robot-assisted surgery (RAS) nowadays is becoming a popular paradigm in the operating theater [1], [2], helping surgeons in gynecological [3], otorhinolaryngological [4], and ophthalmological [5] surgeries.

Surgical autonomy, acting as a fast-growing sector [6] in RAS, integrates state-of-the-art techniques in the field of medical robotics to enable the robotic system with capabilities of conducting preoperative planning by multi-modal data fusion [7], carrying out intraoperative sensor-based interventions [8], and implementing learning-driven operations [9]. The prosperous developments in surgical automation at present are not only encouraging the integration of artificial intelligence [10], [11] and multiple sensors-based control schemes [12], [13] with medical robotic systems, but also enhancing the efficiency and altering operational modes of numerous kinds of surgical sub-tasks in clinical room.

Surgical suturing and knot tying is among one of these sub-tasks which is tedious yet important [14], [15], and it
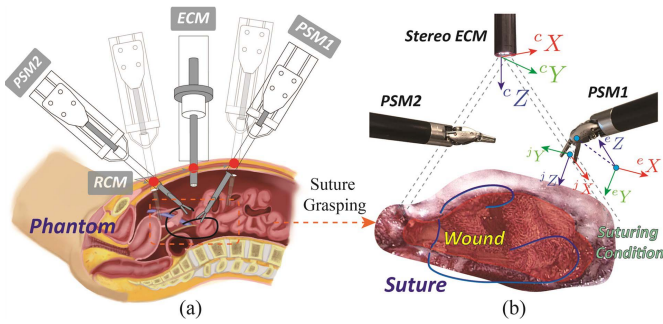
Fig. 1. Image-guided automated suture grasping on the dVRK platform. (a) The illustration of suture grasping using surgical robots with RCM constraints. (b) Typical example of the grasping scenario and the end-effector's pose.

exists from in-vitro cuticle wound closure [13] to in-vivo transanal total mesorectal excision (taTME) [16], and from macro-scale suturing in laparoscopic hysterectomy [17] to micro-scale endoscopic vascular anastomosis [18]. To reduce fatigues from surgeons when conducting this manipulation, researchers put their attentions on developments of automation skills in the aspects of the task planning [19], [20], soft tissue manipulation [21], [22], hardware design [23], [24], learning-based robotic manipulations [25], [26], and etc. However, some of these on-the-shelf approaches only focused on generations of system level human-machine collaborative modes [23], [26] for surgical knot tying, in which the robot needs continuous supervision and control from human. In addition, other works focused on automating the sub-tasks of suture stitching [20], [24], [25] and its looping operations [13], [19]. As claimed in [27], the suture needs to be cut after the stitching manipulation, especially when adopting the interrupted suturing technique. Consequently, an automated suture grasping between the aforementioned two sub-tasks (stitching and looping) is urgent to be achieved.

To resolve this problem, in this paper, we proposed a holistic framework to pave the way towards this automated suture thread grasping. Based on our prior works [27], [28], this approach can achieve a task-level autonomy [6] with the integration of multidisciplinary knowledge, and it can be seamlessly deployed on the da Vinci Research Kit (dVRK) platform, emulating the state-of-the-art paradigm of surgical knot tying as is depicted in Fig. 1. This paper is a comprehensively improved version of our previous work [29], and to the best of our knowledge, this is the first work that paves the way towards robotizing the suture grasping operation, setting a solid foundation for surgical knot tying automation.

Our main contributions can be summarized as follows:

1). Considering complex and varying environments in surgical knot tying, we develop an evolved suture segmentation scheme using a semi-supervised learning architecture, in which a small number of labeled data and a large portion of unlabeled images containing sutures under various conditions are utilized, so as to generate a more generic model and save the massive labeling time. Besides, a suture-aware loss which can learn a suture's curvilinear and slender structure is designed to improve the overall segmentation performance.

2). The pixel-wise topological sequence of a suture is completely figured out in the camera frame. To precisely yield its 3D shape, a sampling-based sliding pairing (SSP)

scheme is jointly exploited to align the suture's featureless key points in the stereo Endoscopic Camera Manipulator (ECM) for triangulation and 3D shape optimization.

3). By analyzing the characteristics of the suture's 3D shape, we propose universal criteria and pose optimization functions, calculating the grasping position and orientation of a RCM constrained Patient Side Manipulator (PSM), and the final deployment position of the suture after its successful grasping is also formulated.

4). To compensate for practical errors induced from the hand-to-eye calibration [30], inherent mechanical deviations, 3D computational errors, and etc., we generate a novel strategy by conducting the grasping trial with four designated actions. In contrast to the human-in-the-loop system, an autonomous vision-based mechanism is also proposed to identify the grasping status of the suture, thereby further liberating the manual supervision from the operational loop.

The rest of this article is organized as follows. In Section II, we review relevant research topics towards the realization of an automated suture grasping, and then the learning-enabled and optimization-based framework is systematically introduced in Section II. The performances of our approach, ranging from visual perception to robotic manipulations, are thoroughly validated, and some future works are also discussed in Section IV. We finally raise our conclusion and acknowledge in Section V and V.

II. RELATED WORKS

Regarding our proposed framework, relevant literature can be systematically reviewed from three-folds, including the curvilinear object segmentation, 3D shape reconstruction, and image-guided automated grasping operation.

For the segmentation task, there are some existing works implemented in relevant domains, such as catheter detection [31]–[33], vascular and retinal [34]–[37] segmentation. In [31], Hernandez-Vela *et al.* proposed a fully automatic approach which integrates the local and contextual information to detect tubular structures, e.g. the catheter and the vessel's centerline, and Hoffmann *et al.* [32] introduced a graph-search method with manual initialization for the electro-physiology catheter detection. Besides, Lee *et al.* [33] used deep-learning method and developed a cascading segmentation system for catheter's localization. In addition, Li *et al.* [36] remolded the vessel segmentation task as a cross-modality data transfer from retinal image to vessel structure using a supervised learning approach, and Yan *et al.* [37] optimized their loss function to well balance the thick and thin structures among retinal. These methods, however, segment curvilinear objects using radiology images [31]–[33], and our ECM system may contain different environmental noises when conducting image-guided manipulations. When adopting transitional method or image operators [34], [35], the outcome may contain massive noises. Besides, the learning-based methods typically rely on vast labeling data [36], [37], and it will be time-consuming to manually output masks in new images. Our semi-supervised learning method tackles these existing problems and provides a more feasible way for the suture's perception in knot tying.

As for the 3D reconstruction, the depth camera is a commonly used tool, such as for the indoor environment or object

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

LU *et al.*: TOWARD IMAGE-GUIDED AUTOMATED SUTURE GRASPING UNDER COMPLEX ENVIRONMENTS 3
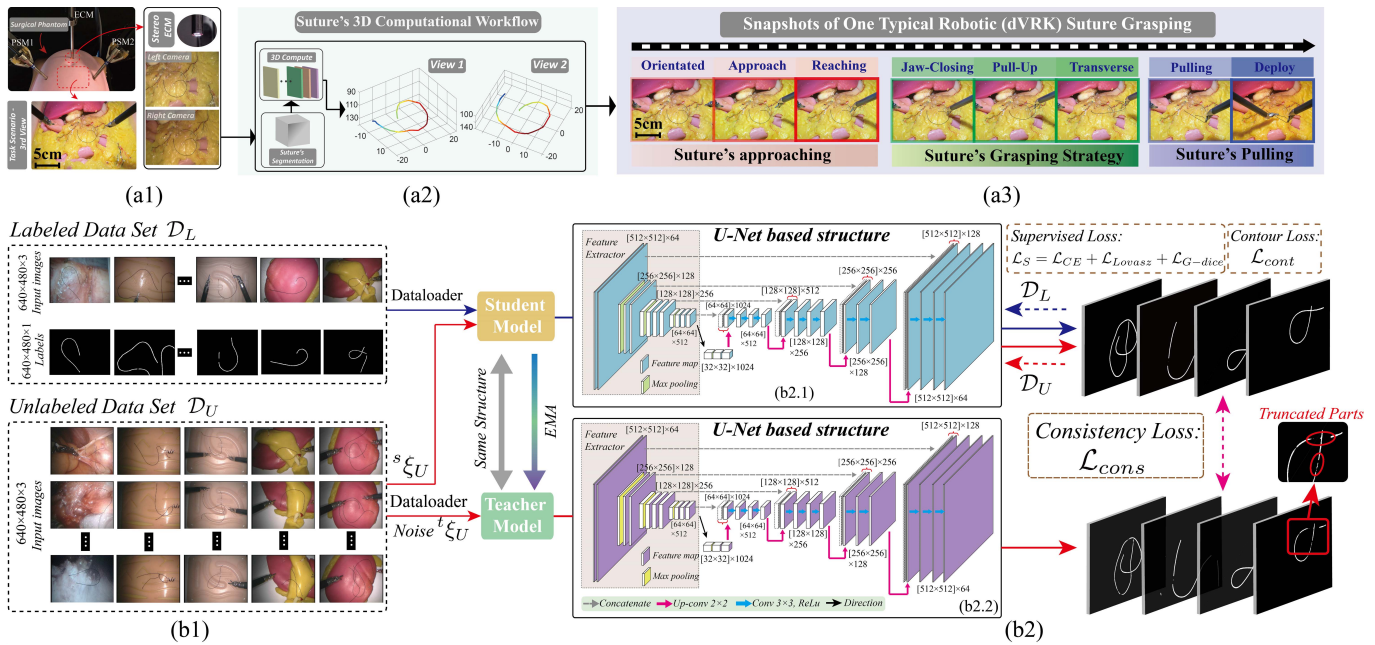

Fig. 2. (a). The workflow of our learning-enabled and optimization-based suture grasping scheme. From the stereo images-based suture's segmentation and its 3D reconstruction, to the robotic execution using PSM to pick up the suture. (b). The semi-supervised learning architecture for suture segmentation. We combine an enhanced supervised loss, the contour loss, and the consistency loss to train a model that can segment the suture under ECM imaging system.

reconstruction [38]–[40] and the 3D curvilinear object (rope) computation [41]. But the inherent limitation is the minimum sensing depth of the camera, which cannot provide accurate feedback within limited space in MIS. Stereo-camera is another option in the intro-operative imaging system. However, matching featureless key points of the suture precisely in two frames is a critical issue before its 3D reconstruction. In [32], the 3D shape of the curved-like catheter was calculated by triangulating the segmented 2D curves in two images directly. Besides, Jackson *et al.* [42] used a non-uniform rational B-spline (NURBS) curve-based method to track the suture and aligned stereo feature points by minimizing a matching energy. Similarly, Padoy and Hager [43] combined discrete Markov random field and NURBS to formulate their 3D computation scheme. Our method, however, gets rid of the limitation of the curve model so as to realize a pixel-level understanding of a suture's sequence, and the optimized key points matching scheme ensures a complete and accurate 3D model output.

On the topic of automated grasping, it historically involves various works inspired by model-based methods [44], [45], sensor-guided manipulations [46], [47], and learning-driven approaches [48], [49]. However, these existing works mainly focused on the automated grasping of domestic objects using macro-scale robotic grippers, and the demand of the grasping accuracy will dramatically increase when it comes to the autonomy in laparoscopic surgery. Different to [46], [49] in which uncertainties can be induced from the inference procedure or reinforcement learning, one of our focuses is to explicitly exploit the pose of our instrument towards the grasping of the designated suture. Besides, some grasping tasks in MIS, e.g. automated needle's grasping, are conducted

in a structured environment [50], or rely on the improvement of the hand-to-eye calibration [30], [51], [52]. Our method does not rely on additional marker information and can also tolerate practical errors in such RCM constrained operations.

III. METHODOLOGY

A. Overview of Framework

We develop a holistic framework using visual guidance to autonomously execute the suture grasping manipulation in various scenarios. As depicted in Fig. 2(a), the operator only needs to click around the suture's cutting tip in one image to initialize the task.

By leveraging a learning-based method, our framework can segment the entire suture under complex and unstructured environments, so as to facilitate its pixel level sequence inference in the stereo-camera. Utilizing this result, an optimal 3D shape of the suture can be obtained. With this visual perception outcome, a grasping strategy that can tolerate inherent errors is forwarded. We also design a vision-based approach to monitor the suture grasping status, so that manual supervision can be further released from the loop.

B. Semi-Supervised Learning for Suture Segmentation

Acting as the initial sensory guidance, an accurate segmentation of the suture plays a key role in our framework. When there exist environmental noises, traditional image processing methods such as color filtering, edge operator, or Hessian matrix-based approach can hardly output satisfactory results. With prompt developments of sophisticated architectures, deep learning-based approach is a potential solution to resolve this dilemma. To generate a reliable model for the suture's segmentation task, we consider the following principles:

Principle 1: Since sutures' annotations are tedious and time-consuming, operators or surgeons are unwilling to spend their energies on generations of various mask labels especially when the dataset goes large. To maintain the usability of our segmentation model, it should be capable of rapidly learning the suture's curvilinear structure and texture feature with limited labeled data.

Principle 2: In practice, there exist larger numbers of similar legacy images that were recorded from different surgical demonstrations or on-site manipulations. To make full use of them, the proposed architecture should also be able to grab our interested information in an unsupervised manner, and in return, cultivate a more robust and reliable model that can adapt to more scenarios and output satisfactory results even under complex environments.

As a result of these concerns, we build up our semi-supervised curvilinear suture segmentation model by integrating the U-Net architecture [53] to the mean teacher structure [54] as is shown in Fig. 2(b). U-Net exhibits superior performances in biomedical image processing, which can provide an exceptional feature extraction base for our semi-supervised learning architecture. To embed the mean teacher model, our structure is divided into the supervised and unsupervised parts, namely the student and teacher models respectively, both using U-Net as backbones.

In our task, the training dataset is formulated with N_L annotated and N_U unlabeled images. We respectively denote the labeled and unlabeled images as $\mathcal{D}_L = \{(\mathcal{IN}_i, \mathcal{LA}_i)\}_{i=1}^{N_L}$ and $\mathcal{D}_U = \{\mathcal{IN}_i\}_{i=N_L+1}^{N_L+N_U}$, where $\mathcal{IN}_i \in \mathbb{R}^{H_{\mathcal{I}} \times W_{\mathcal{I}} \times 3}$ is the input image and $\mathcal{LA}_i \in \{0,1\}^{H_{\mathcal{I}} \times W_{\mathcal{I}}}$ denotes normalized annotations of sutures. To compensate for data noises and maintain the overall stability, we purposely induce random perturbations $^{s}\xi_U$ and $^{t}\xi_U$ when importing unlabeled images to the student and teacher models. The objective of our semi-supervised learning task can be formulized as the minimization of the following suture-aware loss function:

$$\min\{\sum_{i=1}^{N_L} \mathcal{L}_S(\mathbf{W}(\mathcal{IN}_i), \mathcal{LA}_i) + \lambda_1 \sum_{i=1}^{N_L} \mathcal{L}_{cont}(\mathbf{W}(\mathcal{IN}_i),$$

$$\mathcal{LA}_i) + \lambda_2 \sum_{i=N_L+1}^{N_L+N_U} \mathcal{L}_{cons}(\mathbf{W}(\mathcal{IN}_i, {}^{s}\xi_U), \mathbf{W}(\mathcal{IN}_i, {}^{t}\xi_U))\} \quad (1)$$

Within Eq. (1), $\mathbf{W}(\bullet)$ represents the segmentation workflow, and this loss has three main terms explained as follows:

Term 1: $\mathcal{L}_S$ denotes a hybrid supervised loss computed using $\mathcal{D}_L$ through the student model. To highlight the slender information of the suture in the supervised learning branch, this hybrid loss $\mathcal{L}_S$ is formulated as:

$$\mathcal{L}_S = w_1 \mathcal{L}_{CE} + w_2 \mathcal{L}_{G-Dice} + (1 - w_1 - w_2) \cdot \mathcal{L}_{Lovasz} \quad (2)$$

where $\mathcal{L}_{CE}$ is the commonly used cross entropy loss, and $\mathcal{L}_{G-Dice}$ depicts the generalized dice loss [55] which is featured of trading off the foreground and background areas by assigning balancing weights when training the network. The Lovász loss [56] is also combined with the $\mathcal{L}_S$, so as to improve the performance of Intersection of Union (IoU) for the slender suture segmentation. In our work, we tried different

values of weight ratios w_1 and w_2, aiming to achieve the best model that can be applied to suture thread segmentation. Finally, we optimize these two values as $w_1 = 0.2$ and $w_2 = 0.4$, accordingly.

Term 2: $\mathcal{L}_{cont}$ is a novel contour loss that is adopted to ensure the completeness of the suture segmentation by particularly emphasizing its continuous peripheral contour. This term is derived as:

$$\mathcal{L}_{cont} = - \sum_{R_I=1}^{H_I} \sum_{C_I=1}^{W_I} \{\frac{\partial(\mathcal{LA}_i)}{\partial u_{im}} \cdot \log[\frac{\partial(\mathbf{W}(\mathcal{IN}_i))}{\partial u_{im}}]$$
$$+ \frac{\partial(\mathcal{LA}_i)}{\partial v_{im}} \cdot \log[\frac{\partial(\mathbf{W}(\mathcal{IN}_i))}{\partial v_{im}}]\}|_{\{R_I, C_I\}} \quad (3)$$

where u_{im}, v_{im} denote the horizontal and vertical directions, R_I, C_I are the row and column numbers in the image. By calculating derivatives along the suture's boundary, it penalizes the false segmentation along the contour, hence improving the overall prediction accuracy. Since the labeling and segmentation values are between 0 and 1, it ensures $\mathcal{L}_{cont}$ can be positive by introducing the logarithm operation in Eq. (3).

Term 3: $\mathcal{L}_{cons}$ is the consistency loss for measuring prediction differences between the teacher and student models under random perturbations, and we use the exponential moving average (EMA) with a standard decay parameter to update the teacher model's weights in the training process. With the effect of $\mathcal{L}_{cons}$, the total suture-aware loss, in return, can enhance the quality of the student model, and the entire semi-supervised architecture can be gradually modified. Besides, λ_1 and λ_2 in Eq. (1) are the ramp-up parameters that adjust the weights between the $\mathcal{L}_{cont}/\mathcal{L}_{cons}$ and $\mathcal{L}_S$.

C. Perception and Optimization of 3D Suture

Let $^{l}\mathcal{X}_{\widehat{\mathbf{S}}_l}$ and $^{r}\mathcal{X}_{\widehat{\mathbf{S}}_r}$ represent the predicted 2D masks of the suture in the left and right cameras. Using the composite cost function based searching strategy introduced in our previous paper [29], the pixel-level sequence $^{l}\mathcal{X}_{\mathbf{S}_l} \in {}^{l}\mathcal{X}_{\widehat{\mathbf{S}}_l}$, $^{r}\mathcal{X}_{\mathbf{S}_l} \in {}^{r}\mathcal{X}_{\widehat{\mathbf{S}}_r}$ can be obtained and denoted as:

$$\begin{cases} ^{l}\mathcal{X}_{1\sim\mathbf{S}_l} = [^{l}\mathcal{X}_1, {}^{l}\mathcal{X}_2, \ldots, {}^{l}\mathcal{X}_\phi, \ldots, {}^{l}\mathcal{X}_{\mathbf{S}_l}] \\ ^{r}\mathcal{X}_{1\sim\mathbf{S}_r} = [^{r}\mathcal{X}_1, {}^{r}\mathcal{X}_2, \ldots, {}^{r}\mathcal{X}_\phi, \ldots, {}^{l}\mathcal{X}_{\mathbf{S}_r}] \end{cases} \quad (4)$$

where $\mathbf{S}_l$ and $\mathbf{S}_r$ are the total numbers of valid pixels along the suture in two frames. As the workflow shown in Fig. 2, the 3D shape of the suture can be afterwards computed with the results in Eq. (4). Based on the assumption of parallel optical axes of the stereo ECM, we should ideally obtain: $\mathbf{S}_l = \mathbf{S}_r$. Thus, by allocating pixels with the same sequential number, dense 3D point clouds representing the suture can be theoretically constructed. However, due to the limited imaging resolution, as well as inherent errors from the camera calibration, suture segmentation, and sequence inference, it may result in 3D results with unexpected deviations or a grotesque shape if merely matching 2D points with the same sequential number.

To confront this problem, we proposed a sampling-based sliding pairing (SSP) approach, in which multiple candidates for each key point can be constructed in the 3D space. As depicted in Fig. 3, let use the stereo pair computation of $^{l}\mathcal{X}_\phi$ for a detailed explanation. In the SSP method, the sequential

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

LU *et al.*: TOWARD IMAGE-GUIDED AUTOMATED SUTURE GRASPING UNDER COMPLEX ENVIRONMENTS 5
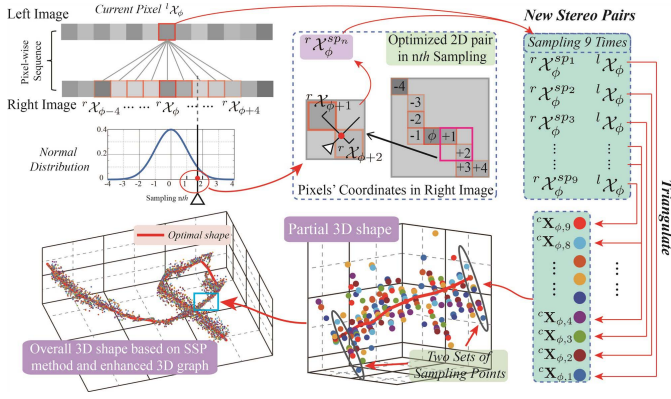

Fig. 3. Gaussian sampling-based sliding pairing (SSP) approach for the 3D shape computation of the suture with the 3D graph enhancement.

number of the pairing point with respect to $^l\mathcal{X}_\phi$ follows a general normal distribution rule: $sp_{i_s} \sim \mathcal{N}(\phi, 1)$, in which the sampling probability $g(\bullet)$ is expressed as:

$$g(sp_{i_s}|(\phi, 1)) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}(sp_{i_s} - \phi)^2\right) \quad (5)$$

where $\phi \in \mathbb{N}$ is the current key point's sequential number among $[1, \min\{\mathbf{S}_l, \mathbf{S}_r\}]$, and $sp_{i_s} \in \mathbb{R}^+$ represents the number obtained at $n\,th$ sampling in between $[1, \min\{\mathbf{S}_l, \mathbf{S}_r\}]$. Eq. (5) enables pixels closer to $^r\mathcal{X}_\phi$ with a higher probability when sampling the stereo pair, while remote ones can be rarely selected. Besides, we use linear interpolation to attain the pixel coordinates of the candidate point when the sampling number is not integral, which can compensate for the inaccuracy introduced from the limited image resolution. In our case, the sampling time is set as 9 to generate multiple stereo pairs for each key point, and the corresponding 3D candidates can be hence calculated based on the triangulation, as is denoted by: $\{^r\mathcal{X}_\phi^{sp_{i_s}}, {}^l\mathcal{X}_\phi\}_{i_s=[1,9]} \rightarrow {}^c\mathbf{X}_{\phi, i_s}$. In the following contents, we adopt $^\bullet\mathbf{X}_\circ$ to express the 3D Coordinates of Point $\circ$ in $\bullet$ Frame. Here, $^c\mathbf{X}_{\phi, i_s}$ depicts the 3D coordinates of $i_s\,th$ candidate of $\phi\,th$ key point in the camera frame. Repeating this step, multiple candidates can be generated in 3D space.

To reconstruct the suture, we adopt Dijkstra's shortest path theory [57] with our enhance 3D graph in order to obtain an optimal shape based on all these 3D candidates. In this sector, our 3D graph, acting as a critical component, is established to characterize the spatial structure of the suture, adapting to complex topologies and ensuring the generality of our method. To obtain this graph, we first construct multiple paths by connecting any two candidates with the following equation:

$$\forall\{^c\mathbf{X}_{\phi, i_s}, {}^c\mathbf{X}_{\psi, j_s}\}; \{i_s, j_s\} \in [1, 9], \{\phi, \psi\} \in [1, \min\{\mathbf{S}_l, \mathbf{S}_r\}]:$$

$$\mathbf{D}_{\phi \leftrightarrow \psi}(i_s, j_s) = \begin{cases} |^c\mathbf{X}_{\phi, i_s} - {}^c\mathbf{X}_{\psi, j_s}|, & |^c\mathbf{X}_{\phi, i_s} - {}^c\mathbf{X}_{\psi, j_s}| \leqslant \text{TS} \\ \infty, & |^c\mathbf{X}_{\phi, i_s} - {}^c\mathbf{X}_{\psi, j_s}| > \text{TS} \end{cases}$$

where $\mathbf{D}_{\phi \leftrightarrow \psi}(i_s, j_s)$ denote the length of the valid path between $\phi\,th$ and $\psi\,th$ key points formed by their $i_s\,th$ and $j_s\,th$ sampling candidates. It should be noticed, $\phi \neq \psi$ and all valid paths are limited within a predefined physical length TS. Thus,

the physical length graph $\mathbf{G}_{len}$ can be written as:

$$\mathbf{G}_{len}(9(\phi - 1) + i_s, 9(\psi - 1) + j_s) = \begin{cases} \mathbf{D}_{\phi \leftrightarrow \psi}(i_s, j_s), & \phi \neq \psi \\ \infty, & \phi = \psi \end{cases}$$

$$(6)$$

Based on $\mathbf{G}_{len}$, a shortest path which has a global minimal length between the designated starting and ending points can be figured out. However, when computing a suture with self-intersections or cross over with other sutures, it may potentially output a 3D shape with sharp turning at any crossing point, without reaching other necessary components. To characterize the entire 3D information, we hence developed another penalty graph $\mathbf{G}_{pen}$ to virtually increase length of paths that are connected by two candidates with a large difference in between their sequential number:

$$\mathbf{G}_{pen}(9(\phi - 1) + i_s, 9(\psi - 1) + j_s) = 1 + \exp\left(\frac{|\phi - \psi|}{\text{scale}} - 2\right)$$

$$(7)$$

When searching for valid candidates, we prefer to add more probabilities on points which are closer to the central space. Hence, we additionally define and compute a clustering degree parameter for each candidate with the following steps:

1. Set up a virtual sphere which is centered by point $^c\mathbf{X}_{\phi, i_s}$ with a radius Υ, which we set as 1.5mm in our case;

2. Count the total number $N_{clu}(\phi, i_s)$ which measures how many other candidates exit within this virtual sphere;

3. Compute the clustering weight for each candidate using $w_{clu}(\phi, i_s) = s_{clu}/(-N_{clu}(\phi, i_s))$ with setting up the scale factor as $s_{clu} = 10$ in our case;

4. Generate the clustering weighting graph $\mathbf{G}_{clu}$ for each path by multiplying the clustering degree parameters of its two end points, which is expressed as:

$$\mathbf{G}_{clu}(9(\phi - 1) + i_s, 9(\psi - 1) + j_s) = w_{clu}(\phi, i_s) \cdot w_{clu}(\psi, j_s)$$

$$(8)$$

Taking Eq. (6) and (7) into account, the optimized graph $\mathbf{G}_O$ with path enhancements can fully characterize the 3D suture, and it can be calculated using the Hadamard product:

$$\mathbf{G}_O = \mathbf{G}_{len} \odot \mathbf{G}_{pen} \odot \mathbf{G}_{clu} \quad (9)$$

As shown in Fig. 3, with concentrated 3D points generated by our SSP method, the overall suture shape $^c\mathbf{X}_{suture}$ with self-intersections can be smoothly obtained using the shortest path theory with our enhanced spatial graph. Acting as a visual guidance, the enhancement in 3D suture shape accuracy can significantly benefit the following grasping manipulation. The general steps of our 3D shape optimization is listed in Alg. 1.

D. Suture Grasping With Pose Optimization

In this subsection, we first define: $\mathcal{F}_\bullet$: notation of $\bullet$ Frame; $^\bullet\mathbf{X}_\circ$: 3D coordinates of Point $\circ$ in $\bullet$ Frame; $^\bullet\mathbf{R}_\circ$: Rotation matrix from $\circ$ Frame to $\bullet$ Frame; $^\bullet\mathbf{P}_\circ$: Translation vector from $\circ$ Point to $\bullet$ Point; $^\bullet\mathbf{T}_\circ$: Transformation matrix from $\circ$ Frame to $\bullet$ Frame; $^\bullet X, {}^\bullet Y, {}^\bullet Z$: X, Y, Z axis of $\bullet$ Frame; $\text{Rot}(\bullet, \circ)$: Rotate about $\bullet$ Axis for $\circ$ angle(s).

Algorithm 1 Suture 3D Shape Optimization With SSP

Data: Segmentation Results $^{l}\mathcal{X}_{1\sim S_l}$ and $^{r}\mathcal{X}_{1\sim S_r}$

1 Load `triangulate, graph` MATLAB functions;

2 **if** $|S_l - S_r| \leqslant 20$ **then**

3   **for** $\phi \in [1, \min\{S_l, S_r\}]$ **do**

4   **for** $i_s \in [1, 9]$ **do**

5       $^{l}\mathcal{X}_\phi \xrightarrow{Eq.(5)}$ Probabilistic Sampling: sp_{i_s};

6   `triangulate`$(^{l}\mathcal{X}_\phi, ^{r}\mathcal{X}_\phi^{sp_{i_s}}) \rightarrow {}^{c}\mathbf{X}_{\phi,i_s}$;

7   **end**

8   **end**

9   Construct $\mathbf{D}_{\phi \leftrightarrow \psi}(i_s, j_s) \xrightarrow{Eq.(6),(7) \ and \ (8)}$ compute $\mathbf{G}_{len}, \mathbf{G}_{pen},$ and $\mathbf{G}_{clu} \xrightarrow{Eq.(9)}$ calculate $\mathbf{G}_O$;

10   Compute spatial graph $\rightarrow$ `graph`$(\mathbf{G}_O) \xrightarrow[End \ Node]{Begin \ Node}$ Output 3-D Suture $^{c}\mathbf{X}_{suture}$;

11 **else**

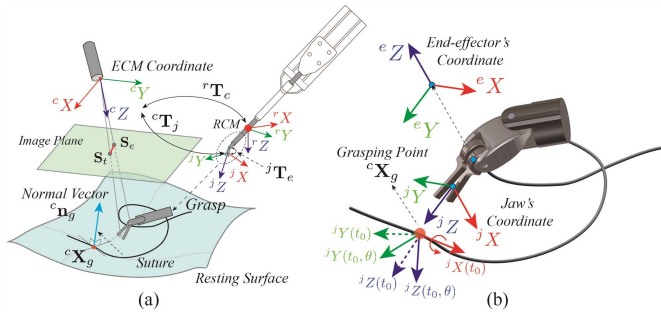12   Print "Error in Segmentation!"

13 **end**

Fig. 4. Suture grasping using the dVRK system. (a). The illustration of the ECM-PSM robotic system with a computed 3D suture and its resting surface; (b). The orientation of $\mathcal{F}_e$ and $\mathcal{F}_j$ when grasping the suture.

As depicted in Fig. 4, the origin of the end-effector's frame $\mathcal{F}_e$ is located at the distal joint of PSM, and ^{e}Z represents the rotational axis. With ^{e}Y pointing to its tip along the longitudinal direction, ^{e}X can be settled using the right-hand rule. To grasp the suture using PSM, we also defined a jaw frame $\mathcal{F}_j$, whose origin is located at the half length of the jaw. As defined in Fig. 4 (b), its ^{j}X is perpendicular to the jaw's opening and closing plane, while ^{j}Z and ^{j}Y are accordingly along ^{e}Y and $-^{e}X$. In contrary to the dynamic $\mathcal{F}_e$ and $\mathcal{F}_j$, the base frame $\mathcal{F}_r$ of PSM is mounted at RCM point, which is passive and stable during the overall manipulation.

On the dVRK platform, PSM is operated by controlling the motion of $\mathcal{F}_e$ with respect to $\mathcal{F}_r$. Hence, to obtain a generic suture grasping method, certain criteria regarding the grasping pose of $\mathcal{F}_e$ relative to $\mathcal{F}_r$ should be investigated, so as to obtain the corresponding $^{r}\mathbf{T}_e$. According to the configuration of our system, this transformation relationship can be formulated using the following chain rule:

$$^{r}\mathbf{T}_e = {}^{r}\mathbf{T}_c \cdot {}^{c}\mathbf{T}_j \cdot {}^{j}\mathbf{T}_e \tag{10}$$

By leveraging Tsai's hand-to-eye calibration method [58], $^{r}\mathbf{T}_c$ can be obtained offline after setting the passive joints of

PSM and ECM. According to the definitions of $\mathcal{F}_e$ and $\mathcal{F}_j$ on the dVRK system, $^{j}\mathbf{T}_e$ can be formed using:

$$^{j}\mathbf{T}_e = \begin{bmatrix} \text{Rot}(^{j}Z, -\pi/2) \cdot \text{Rot}(^{j}X, \pi/2) & ^{j}\mathbf{P}_e \\ \mathbf{0} & 1 \end{bmatrix} \tag{11}$$

Since ^{j}Z overlaps the jaw's longitudinal direction, with l_{jaw} denotes its length, $^{j}\mathbf{P}_e$ can be expressed as: $[0, 0, -l_{jaw}/2]$. Towards automating this task, the critical issue by now is to figure out $^{c}\mathbf{T}_j$ when conducting the grasping operation. To tackle this problem, three criteria regarding the translation and rotation relationships from $\mathcal{F}_j$ to $\mathcal{F}_c$ are forwarded:

Criterion 1: The grasping point $^{c}\mathbf{X}_g$ should overlap the origin of $\mathcal{F}_j$ when PSM's jaw successfully grasps the suture;

Criterion 2: ^{j}X should point along the spatial tangent direction of the suture when conducting the grasping operation;

Criterion 3: To avoid secondary injuries, ^{j}Z should be aligned orthogonal to the suture's resting surface as much as possible.

In **Criterion 1**, it directly gives the translation vector as $^{c}\mathbf{P}_j = {}^{c}\mathbf{X}_g$. To satisfy **Criterion 2**, the tangent direction at $^{c}\mathbf{X}_g$ is computed by taking all its nearby vectors formed by successive 3D key points into account, hence depicting ^{j}X as:

$$^{j}X(t_0) = \frac{^{j}\hat{X}}{\| ^{j}\hat{X} \|}, \ s.t. \ ^{j}\hat{X} = \sum_{i=-n_t}^{n_t} f_p(i) \cdot \frac{^{c}\mathbf{X}_{g+i} - {}^{c}\mathbf{X}_{g+i-1}}{\| ^{c}\mathbf{X}_{g+i} - {}^{c}\mathbf{X}_{g+i-1} \|} \tag{12}$$

where $^{j}X \in \mathbb{R}^3$ is under ECM (camera) frame $\mathcal{F}_c$, $g \in \mathbb{R}$ and t_0 denote the grasping point number and the initial time. Because a dense 3D vertex graph of the suture is constructed, we can assign a large number n_t to approximate the direction of ^{j}X. Besides, $f_p(i)$, which obeys the normal distribution: $f_p(i) = \frac{1}{\sqrt{2\pi}} \exp(-\frac{i^2}{2})$, is adopted as the weighting parameter in Eq.(12). This method considers the suture's local shape around $^{c}\mathbf{X}_g$ and implies that the vector with a remoter index has a fewer impact on the determination of ^{j}X, and thus can compensate for fortuitous errors among those nearby key points without the loss of computational generality.

Before aligning ^{j}Z using **Criterion 3**, we randomly initialize an $^{j}Z(t_0)$ which only satisfies $^{j}Z(t_0) \perp {}^{j}X(t_0)$, while $^{j}Y(t_0)$ can be calculated by $^{j}Z(t_0) \times {}^{j}X(t_0)$. Therefore, an initial $^{c}\mathbf{R}_j(t_0) \in \mathbb{R}^{3\times3}$ can be obtain:

$$^{c}\mathbf{R}_j(t_0) = [^{j}X(t_0), {}^{j}Y(t_0), {}^{j}Z(t_0)] \tag{13}$$

The further adjustment of $^{j}Z(t_0)$ according to **Criterion 3** can be regarded as a rotation around $^{j}X(t_0)$ for θ angle as is shown in Fig. 4(b), and the grasping orientation of $\mathcal{F}_j$ with respect to $\mathcal{F}_c$ can be adjusted as:

$$^{c}\mathbf{R}_j(t_0, \theta) = {}^{c}\mathbf{R}_j(t_0) \cdot \text{Rot}(^{j}X(t_0), \theta)$$
$$= \begin{bmatrix} r_{11} & r_{12} & r_{13} \\ r_{21} & r_{22} & r_{23} \\ r_{31} & r_{32} & r_{33} \end{bmatrix} \cdot \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos(\theta) & -\sin(\theta) \\ 0 & \sin(\theta) & \cos(\theta) \end{bmatrix} \tag{14}$$

Extracted from Eq. (14), $^{j}Z(t_0, \theta)$ is hence written as:

$$^{j}Z(t_0, \theta) = \begin{bmatrix} -r_{12}\sin(\theta) + r_{13}\cos(\theta) \\ -r_{22}\sin(\theta) + r_{23}\cos(\theta) \\ -r_{32}\sin(\theta) + r_{33}\cos(\theta) \end{bmatrix} \tag{15}$$

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

LU *et al.*: TOWARD IMAGE-GUIDED AUTOMATED SUTURE GRASPING UNDER COMPLEX ENVIRONMENTS 7

To satisfy **Criterion 3**, $^jZ(t_0, \theta)$ should align close to the normal vector $^c\mathbf{n}_g$ of the suture's resting surface as is depicted in Fig. 4(a). The resting surface of the suture can be approximated using a polynomial surface fitting by adopting all computed 3D vertexes of the suture [59], and $^c\mathbf{n}_g$ is hence achieved by the bi-cubic interpolation at $^c\mathbf{X}_g$ point, which is:

$$^c\mathbf{n}_g = [n_1, n_2, n_3]^\top \tag{16}$$

To minimize the angle between $^jZ(t_0, \theta)$ and $^c\mathbf{n}_g$, a target function $f(\theta)$ which measures the variation between each other is formulated as follows:

$$f(\theta) = {}^c\mathbf{n}_g^\top \cdot {}^jZ(t_0, \theta) = \hat{A} \cdot \sin\theta + \hat{B} \cdot \cos\theta$$
$$= \sqrt{\hat{A}^2 + \hat{B}^2} \cdot \sin(\theta + \Phi) \tag{17}$$

where $\hat{A} = (-r_{12}n_1 - r_{22}n_2 - r_{32}n_3)$, $\hat{B} = (r_{13}n_1 + r_{23}n_2 + r_{33}n_3)$, and $\Phi = \arctan(\hat{B}/\hat{A})$. To fulfill the objective listed in **Criterion 3**, we need to maximize $f(\theta)$ by figuring out an optimal rotation angle θ around $^jX(t_0)$, which can be thereby computed as:

$$\theta = \frac{\pi}{2} - \arctan(\hat{B}/\hat{A}) \tag{18}$$

where $\theta \in [0, \pi]$. Substituting all results obtained according to our proposed criteria into Eq.(10), the desired grasping pose of PSM's end-effector is hence give by:

$$^r\mathbf{T}_e = \underbrace{{}^r\mathbf{T}_c}_{Hand2Eye} \cdot \underbrace{\begin{bmatrix} {}^c\mathbf{R}_j(t_0, \theta), & {}^c\mathbf{X}_g \\ \mathbf{0} & 1 \end{bmatrix}}_{{}^c\mathbf{T}_j} \cdot \underbrace{{}^j\mathbf{T}_e}_{Eq.(11)} \tag{19}$$

where $^r\mathbf{T}_e = [\begin{smallmatrix} {}^r\mathbf{R}_e(t_0, \theta), & {}^r\mathbf{P}_e(t_0, \theta) \\ \mathbf{0}_{1\times3}, & 1 \end{smallmatrix}]$, in which $^r\mathbf{R}_e(t_0, \theta)$ and $^r\mathbf{P}_e(t_0, \theta)$ accordingly denote the orientation and position that controls the grasping pose of PSM.

E. Grasping Strategy and Planning

Considering practical errors from hand-to-eye calibration and inherent mechanical offsets, we may fail to grasp the suture if barely drive PSM to $^r\mathbf{T}_e$ since there may exist offset between the true grasping point and the computed one as is depicted in Fig. 5(a). Besides, due to the calibration error, we cannot use PSM's kinematic image re-projection to guide the motion. Although the vision-servoing control can potentially solve this problem, it requires a robust and dynamic tracking of tools with various poses under different scenarios, which is another challenging topic and also computational costly.

Facing these problems, a grasping strategy which integrates a vision-based mechanism is developed to compensate for the aforementioned errors. Before the grasping, we first drive PSM to a modified target position $^r\overline{\mathbf{P}}_e(t_0, \theta)$ expressed as: $^r\mathbf{P}_e(t_0, \theta) - l_{res} \cdot {}^r\mathbf{r}_{e,2}(t_0, \theta)$, in which the end-effector is retracted along its grasping direction by l_{res} distance. Here, $^r\mathbf{r}_{e,i}(t_0, \theta)$ is the i th column of $^r\mathbf{R}_e(t_0, \theta)$. After reaching $^r\overline{\mathbf{P}}_e(t_0, \theta)$, PSM will go towards $^r\mathbf{P}_e(t_0, \theta)$ with a step increment $5mm$ along ^eY direction, in order to make up for errors in depth. Besides, we also plan 5 grasping positions with position deviations $l_i \cdot {}^r\mathbf{r}_{e,1}(t_0, \theta)$ as depicted in Fig. 5(b). Here, l_i is
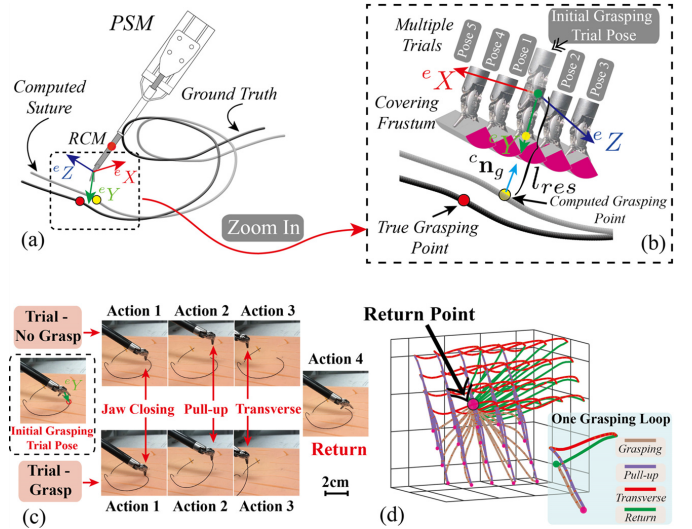

Fig. 5.   (a). An image-guided grasping with errors between the computed result and ground truth; (b). Multiple grasping trials at one grasping depth; (c). Four main actions in a grasping loop; (d). Motion trajectories of the end-effector plotted by the recorded kinematics in one typical grasping task.

set as: $\cdot[0, 1, 2, -1, -2]^\top mm$. It is worthy noticing that our method compensates for errors in ^eX and ^eY, whereas assumed the suture can be grasped with deviations along its longitudinal direction ^eZ. Using this mechanism, the end-effector can cover a comparatively large grasping space, while reducing the execution time by simplifying the traversal searching complexity from 3D space to 2D $^eX - {}^eY$ plane.

As shown in Fig. 5(c), we also designed four typical actions to form a universal loop when executing each grasping trial:

ACTION 1 - JAW-CLOSING: In the n_{inc} th increment, the 3D position now is $^r\mathbf{P}_e(t_0, \theta) - (l_{res} - 5mm \cdot n_{inc})^r\mathbf{r}_{e,2}(t_0, \theta)$. By adding the deviation $l_i \cdot {}^r\mathbf{r}_{e,1}(t_0, \theta)$, PSM goes to the desired point and closes the jaw;

ACTION 2 - PULL-UP: After Action 1, PSM pulls up by: $-20mm \cdot {}^r\mathbf{r}_{e,2}(t_0, \theta)$ in order to pick up the suture;

ACTION 3 - TRANSVERSE: Then, we embed a lateral motion by adding an additional term: $-30mm \cdot {}^r\mathbf{r}_{e,2}(t_0, \theta)$. In this action, we enable a vision-based mechanism to autonomously determine whether the suture is grasped or not. Using the 3D information, we can obtain the grasping point by reserving a $20mm \sim 30mm$ long suture to the cutting tip. Based on the segmentation results, we only monitor the 2D pixels from the cutting tip to the grasping point in the left camera. As shown in Fig. 6(a1-a4), $^l\mathcal{X}_{1\sim g}^{t_1} = [^l\mathcal{X}_1^{t_1} \sim {}^l\mathcal{X}_g^{t_1}]$ are taken as references. With the extraction of $^l\mathcal{X}_{1\sim g}^{t_i}$ in the following grasping trial, we compare the vector similarity between $^l\mathcal{X}_{1\sim g}^{t_1}$ and $^l\mathcal{X}_{1\sim g}^{t_i}$ by measuring the Frobenius norm:

$$\text{SIMI}_{t_1 \leftrightarrow t_i} = 1 - 2 \cdot ||^l\mathcal{X}_{1\sim g}^{t_1} - {}^l\mathcal{X}_{1\sim g}^{t_i}|| / (||^l\mathcal{X}_{1\sim g}^{t_1}|| + ||^l\mathcal{X}_{1\sim g}^{t_i}||) \tag{20}$$

If $\text{SIMI}_{t_1 \leftrightarrow t_i}$ is larger than 80%, it indicates the suture not being grasped, and Action 4 should be followed up.

ACTION 4 - RETURN: Without successful grasping, PSM returns to $^r\overline{\mathbf{P}}_e(t_0, \theta)$ so as to finish one operational loop, and the next grasping trial will be followed up until Action 3

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

8

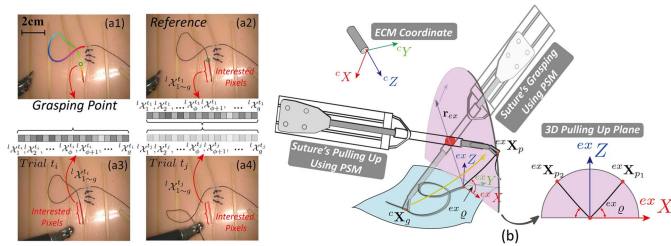IEEE TRANSACTIONS ON AUTOMATION SCIENCE AND ENGINEERING

Fig. 6. (a1-a4). A vision-based mechanism to monitor the suture grasping status; (b). After the suture grasping, it is pulled up to a semi-straight status to prepare for the following automated looping operation.

outputs the signal that the suture has been grasped. In Fig. 5(d), we intuitively plot the end-effector's trajectories in multiple trials, and different colors represent different typical actions. Using this strategy, the suture's status in the grasping procedure can be successfully estimated.

After grasping the suture, we need to further pull the suture up and keep it in a semi-straight state, in order to prepare for the following looping manipulation [13], [19]. To unify this step and adapt to different shapes, we establish an auxiliary exit frame $^{ex}\mathcal{F}$ in which the final deployment coordinates of the suture can be properly obtained. As shown in Fig. 6(b), the origin of $^{ex}\mathcal{F}$ is at $^c\mathbf{X}_{end}$, with ^{ex}Z overlaps the normal vector $^c\mathbf{n}_{end}$ and ^{ex}Y points to the suture's tangent direction. It hence outputs the relation $^c\mathbf{R}_{ex} = [^{ex}X, {}^{ex}Y, {}^{ex}Z]$. PSM finally pulls the suture to $^{ex}\mathbf{X}_p$, which is calculated as:

$$^{ex}\mathbf{X}_{p_i} = [(-1)^i \cdot \mathbf{r}_{ex} \cos({}^{ex}\varrho), \mathbf{r}_{ex} \sin({}^{ex}\varrho), 0]^\top, i \in \{1, 2\} \quad (21)$$

where $\mathbf{r}_{ex}$ is the length between $^c\mathbf{X}_g$ and $^c\mathbf{X}_{end}$ obtained using the suture's 3D information, $^{ex}\varrho$ denotes the pull-up angle between the suture and ^{ex}X-^{ex}Y plane which is set as $75°$ in order to be close to the following looping task setup [13]. With $^c\mathbf{T}_{ex} = [\begin{smallmatrix} ^c\mathbf{R}_{ex}, & ^c\mathbf{X}_{end} \\ \mathbf{0}_{1\times3}, & 1 \end{smallmatrix}]$, we can achieve $[^r\mathbf{X}_{p_i}, 1]^\top = {}^r\mathbf{T}_c {}^c\mathbf{T}_{ex} \cdot [^{ex}\mathbf{X}_{p_i}, 1]^\top$. By reading the current end-effector's position $^r\mathbf{X}_e$, we can alternate in between $^r\mathbf{X}_{p_1}$ and $^r\mathbf{X}_{p_2}$ and select the one which is closer to $^r\mathbf{X}_e$ as the final suture's deployment position, so as to finish the overall framework. The pseudo code of our grasping strategy is listed in Alg. 2.

IV. EXPERIMENTAL RESULTS

We conduct extensive experiments in the aspects of the suture segmentation capability, 3D shape reconstruction accuracy, and image-guided grasping outcome, in order to comprehensively evaluate the performance of each critical function block in our framework. As shown in Fig. 7, our experiments are performed on the dVRK platform which is equipped with a stereo ECM imaging system, providing the 640 by 480 resolution visual feedback. We adopt the large needle driver as the manipulation tool, and the robotic system is connected to a PC, whose processer and graphics are intel@ i7-10700 with 2.9GHz×16 and TITAN RTX, respectively.

A. Data Preparation and Evaluation of Suture Segmentation

To train a model for an online suture segmentation without losing the generality, we first prepare the training and testing

---

Algorithm 2 Automated Suture Grasping Strategy

Data: 3D suture and grasping point $^c\mathbf{X}_{suture}$, $^c\mathbf{X}_g$

1 Load `fit`, `surfnorm` MATLAB functions;
2 `fit`$(^c\mathbf{X}_{suture}) \rightarrow$ `surfnorm`$(^c\mathbf{X}_g) \rightarrow {}^c\mathbf{n}_g \rightarrow {}^c\mathbf{T}_j$;
3 Compute desired grasping pose: $^c\mathbf{X}_g \xrightarrow{Eq.(19)} {}^r\mathbf{T}_e$;
4 Modify initial grasping point as: $^r\overline{\mathbf{P}}_e(t_0, \theta)$;
5 Initialize $Tr = 1$, move to: $(^r\mathbf{R}_e(t_0, \theta), {}^r\overline{\mathbf{P}}_e(t_0, \theta))$;
6 **while** $Tr < Design\ Number$ **do**
7   Move to: $(^r\mathbf{R}_e(t_0, \theta), {}^r\overline{\mathbf{P}}_e(t_0, \theta) + 5Tr \cdot {}^r\mathbf{r}_{e,2}(t_0, \theta))$;
8   **for** $i \in 5\ designed\ Positions$ **do**
9   Jaw Opening $\rightarrow$ Move to: $(Position\ i)$;
10   **ACTION** $1 \rightarrow 2 \rightarrow 3 \xrightarrow{Eq.(20)}$ SIMI;
11   **if** SIMI $\leqslant 80\%$ **then**
12   Break and End(**while**);
13   **else**
14   Continue $\rightarrow i = +$
15   **end**
16   **ACTION** 4 (Move to: $(^r\mathbf{R}_e(t_0, \theta), {}^r\overline{\mathbf{P}}_e(t_0, \theta))$);
17   **end**
18 **end**
19 **if** SIMI $\leqslant 80\%$ **then**
20   Move to: $(^r\mathbf{R}_e(t_0, \theta), {}^r\mathbf{X}_{p_i}) \rightarrow$ Print: "Finished"
21 **else**
22   Print: "Grasping Fail"
23 **end**
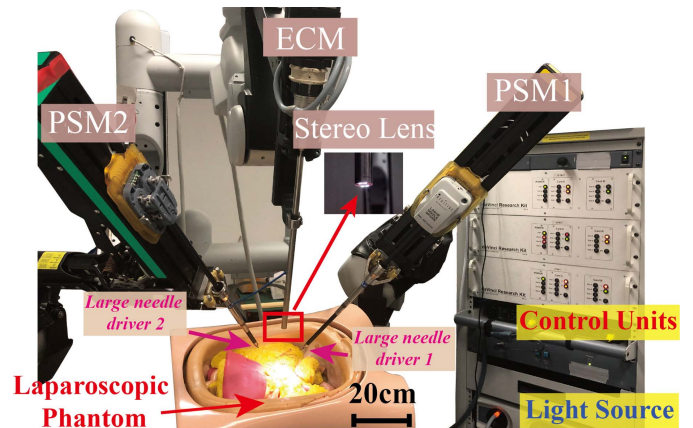
Fig. 7. Experimental setups of the suture grasping using the dVRK platform.

data with multiple backgrounds, i.e. different suturing phantoms, artificial organs and color-printed surgical scenes. Due to inherent features of U-Net based architecture backbone, we can rapidly learn the suture's information by using a few annotated images. In our work, 60 labeled images ($\mathcal{D}_L$) with 480 unlabeled ones ($\mathcal{D}_U$) are utilized as the training set, and another 20 annotated images are leveraged as the testing set for our semi-supervised training workflow, which saves massive data preparation time for practitioners. We use the cross validation for comprehensive evaluation and build up Split 1-4 as categorized in Table. I to accordingly train and test the performance of our model. Apart from this, we also include 2 extra splits in which the testing set contains some

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

LU *et al.*: TOWARD IMAGE-GUIDED AUTOMATED SUTURE GRASPING UNDER COMPLEX ENVIRONMENTS 9

TABLE I

COMPARISONS OF SUTURE SEGMENTATION UNDER VARIOUS CONDITIONS USING THE TRADITIONAL APPROACH, THE U-NET ARCHITECTURE, AND OUR PROPOSED LEARNING-BASED WORKFLOW

| Split Num	Traditional Frangi [28]			U-Net+Cross Entropy Loss [29]			Semi-Supervised+Suture-aware Loss(Ours)		
	IoU/Std	*Dice/Std*	*Recall/Std*	*IoU/Std*	*Dice/Std*	*Recall/Std*	*IoU/Std*	*Dice/Std*	*Recall/Std*
Split 1	0.229/0.083	0.365/0.109	0.860/0.084	0.832/0.062	0.907/0.038	0.916/0.058	0.862/0.041	0.925/0.024	0.939/0.025
Split 2	0.205/0.066	0.335/0.089	0.851/0.094	0.842/0.042	0.914/0.025	0.931/0.041	0.863/0.038	0.926/0.022	0.945/0.043
Split 3	0.220/0.086	0.354/0.113	0.817/0.091	0.833/0.040	0.908/0.024	0.890/0.049	0.869/0.023	0.930/0.013	0.922/0.029
Split 4	0.197/0.075	0.324/0.101	0.837/0.098	0.830/0.077	0.905/0.051	0.893/0.094	0.869/0.026	0.930/0.015	0.932/0.033
Split 5	0.219/0.068	0.354/0.091	0.857/0.103	0.781/0.072	0.875/0.049	0.912/0.098	0.815/0.055	0.897/0.034	0.957/0.034
Split 6	0.235/0.103	0.370/0.135	0.765/0.075	0.733/0.118	0.840/0.083	0.770/0.143	0.803/0.072	0.889/0.046	0.865/0.095
Average	0.218/0.080	0.360/0.106	0.831/0.091	0.808/0.069	0.891/0.045	0.885/0.081	0.847/0.043	0.916/0.026	0.927/0.043
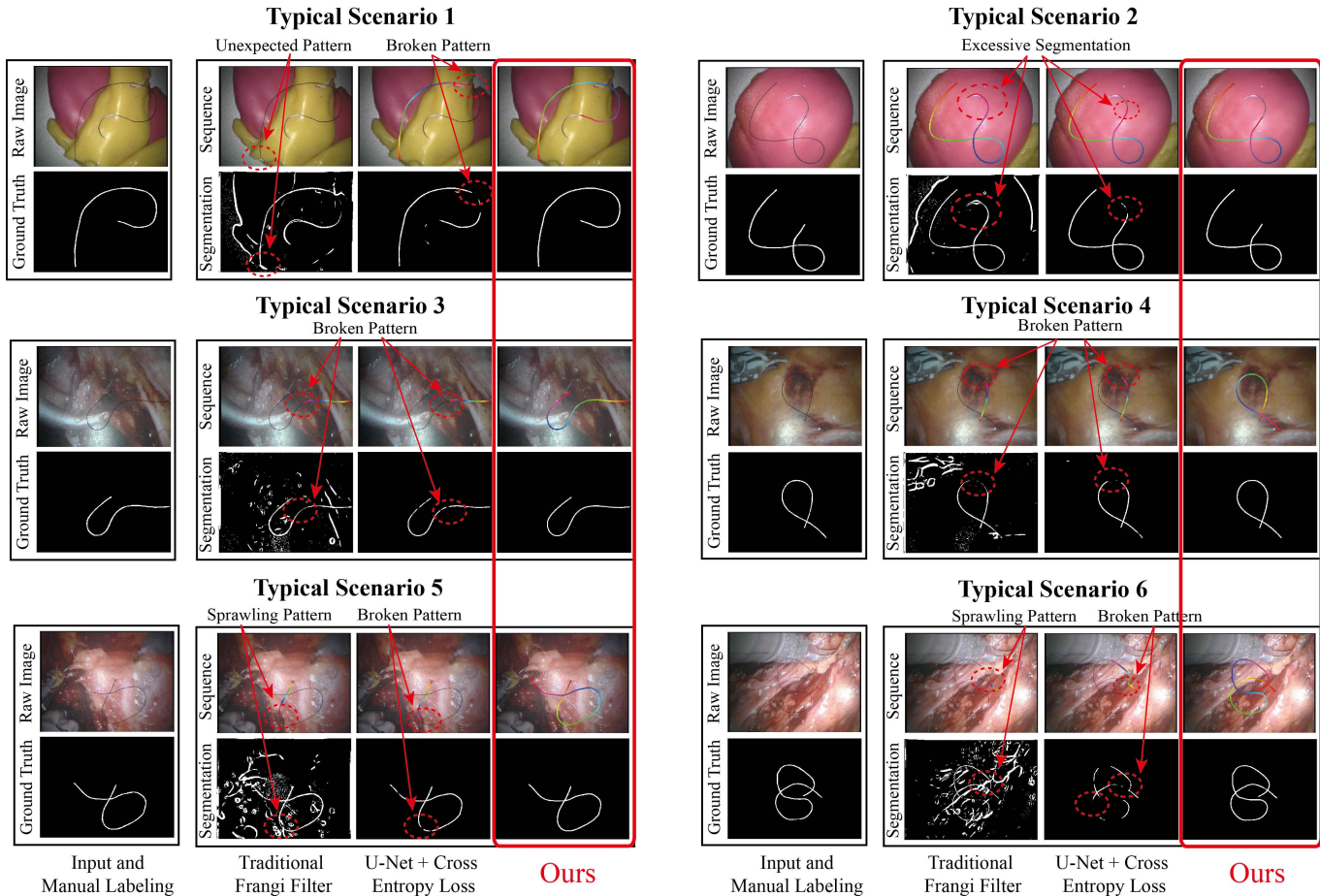
Fig. 8. Typical results of suture segmentation using traditional Frangi filter, the U-Net architecture, and our proposed semi-supervised learning method.

similar but unseen backgrounds in the training stage, so as to evaluate the adaptation capability of our model.

To distinguish our proposed learning method from the traditional Frangi filter [28] and classical U-Net architecture using the cross entropy loss, we demonstrate both the qualitative and quantitative results in Fig. 8 and Table. I. The performance is assessed by comparing Intersection of Union (IoU), Dice Similarity, and Recall Coefficient between the segmentation results and ground truths. As is noticed in the table, the Frangi filter, which is widely adopted for filtering out curvilinear objects, although obtains relatively close numbers comparing to two learning-based methods over Recall Coefficient, its lowest scores in IoU and Dice indicate a majority of false positive segmentation, which is correspondingly revealed by the qualitative results in Fig. 8. Especially under the complex environments as shown in *Typical Scenario 5* and *6*, the sprawling filtering results cannot be used for the following suture pixel-wise sequence inference, hence terminating the workflow in halfway.

For the classical U-Net Architecture [29], it obtains relatively high scores in IoU (80%), Dice (89.1%) and Recall (88.5%), rejecting the majority of the background noises as shown in Fig. 8. Since the suture is a slender object, our proposed approach, as highlighted in the table, outperforms

TABLE II

SUTURE SEGMENTATION PERFORMANCE USING DIFFERENT RATIOS OF LABELED AND UNLABELED IMAGES. LABELED IMAGES: 30

Split 6 - Ratio	1:1	1:5	1:16
IoU/Std	0.711/0.160	0.752/0.102	0.799/0.081
Dice/Std	0.819/0.128	0.854/0.071	0.886/0.054
Recall/Std	0.778/0.195	0.815/0.132	0.862/0.101

around 4%, 2.5% and 4% in average regarding three metrics, which is not a remarkable jump in numbers when comparing to this traditional method. However, it is worthy noticing that such improvements lie in inpainting the broken segmentation patterns under complex backgrounds, e.g. with similar appearance or background reflection, and distinguishing our desired suture from other disturbed objects, e.g. silver needle. The superior performance may result from our elaborated suture-aware loss function and the learning mechanism, in which our model highlights the suture's boundary completeness and learns more knowledge from large amounts of unannotated data. As can been seen from the results in Split 5 and 6, our proposed can also maintain relatively high performance with some unseen backgrounds in the testing set, and especially in Split 6, our method outperforms 7% and 9% in IoU and Recall compared to [29].

To evaluate the impact of the ratio between labeled-unlabeled images for our semi-supervised learning-based segmentation, we re-organize Split 6 by reducing the labeled images to 30, and the validation results under different ratios (1:1, 1:5, and 1:16, with totally 480 unlabeled images) are listed in Table. II. By feeding more unlabeled data into the training workflow, we can notice the overall segmentation performances can be gradually improved. By comparing with U-Net, as noticed, the performance is initially inferior when both fewer supervised and unsupervised data (30 labeled and 30 unlabeled images) is utilized, yet it can exceed U-Net when the ratio is 1:5, which proves the effectiveness of model enhancement by adopting more unlabeled images. But it should also notice that even when the ratio reaches 1:16, the performance cannot outperform the model trained using two times supervised data (60 images), which consequently validates the importance of labeled data in training.

B. Evaluation of 3D Reconstruction With SSP Optimization

Based on the learning-enabled segmentation technique, the suture's pixel-wise sequence can be figured out in both ECM frames using the method proposed in our previous paper [29]. In this part, we will further verify the performance of our SSP method for the suture's 3D reconstruction. To formulate a generic testing experiment, we use the suturing phantom, color-printed surgical scenes, laparoscopic torso phantom and porcine tissues to pursue environmental diversities, with sutures featured with two different lengths - 110mm and 190mm. For the 110mm long one, we only orientate it to curved and one self-intersection shapes, whilst for 190mm length, sutures are additionally orientated with two self-intersections.

We qualitatively assess the testing results as shown in Fig. 9, and then quantitatively compare the computed length to the ground truth. Besides, we intensively compare our optimized method to the approach only uses stereo key points matching for triangulation without SSP for enhancement, and the overall results are listed in Table. III. For each suture length under each background, 7 tests are conducted to verify and compare the reconstruction results. For the ease of reading, we also highlight the largest errors in each group and plot the error charts in Fig. 10 for intuitive comparisons.

As shown in the table, our optimized method can achieve an average absolute error under 2mm regardless of the suture's length and shape, nevertheless this error remarkably increases above 6mm without using our SSP optimization. Although the non-SSP method can achieve comparable precision in some cases, it takes credit to accurate alignments of stereo key points. Under complex shapes and orientations, e.g. self-intersections, small deviations in segmentation and sequence inference may result in incremental errors in stereo pairings, thus bring larger errors, e.g. there exist cases whose errors in the overall 3D length are more than 10mm. Referring to Fig. 10, the bar charts representing their errors are unstable. If a large computational deviation happens around the desired grasping point, it may potentially result in a grasping failure when guiding the automated manipulation with this inaccurate 3D output. As for our proposed optimization method, it can minimize the online computational errors and maintain a certain degree of curvature continuity. As can be seen in Fig. 9, satisfactory 3D shape can be obtained even when sutures possess intricate structures, hence it can serve as a reliable sensory feedback to automate the following grasping task.

C. Experiments on Automated Grasping

With validations from 2D to 3D perceptions, in this part, we carry out extensive experiments by implementing the overall framework to the automated suture grasping task. To achieve a verification generality, we use 3 different suturing phantoms, 1 laparoscopic torso phantom, and the porcine tissue as experimental objects with sutures stitched on them. 10 experiments are conducted under each background.

To quantitatively assess our grasping strategy, we list results regarding 4 proposed evaluation criteria (*E1-E4*) in Table. IV. To demonstrate the grasping procedure, we use one case under each background and select typical frames from our recorded videos to show some key steps. Besides, we record the kinematics of the end-effector and plot its trajectory using the gradient curves in Fig. 11(a), as well as highlighting the workspace occupied by PSM during the overall manipulation with the magenta color. In addition, errors between the true grasping point and the computed position are also illustrated in Fig. 11(b1)-(b5). As noticed, larger errors may happen in the suture's depth estimation, which results from the short baseline of the laparoscopic stereo lens. There exist 26% experiments in which grasping errors exceed 10mm as revealed by *E3*, and as the data summarized in Fig. 11(c1), our grasping strategy can compensate for various local deviations in different directions, which consequently proves the ability of our method in confronting these practical errors in suture grasping.

Apart from these, we analyze X/Y/Z/Total grasping errors with respect to the grasping time as plotted in Fig. 11(c2),

TABLE III

WE USE SUTURES WITH DIFFERENT LENGTHS AND MULTIPLE TYPOLOGIES TO ANALYZE OUR ENHANCED 3D COMPUTATIONAL SCHEME. BG 1, 2, 3, AND 4 ACCORDINGLY DENOTE THE *suturing phantom, color-printed surgical scenes, laparoscopic phantom, and porcine tissue*. UNIT: MM

| Items		Suture - Ground Truth - 110mm								Suture - Ground Truth - 190mm							
Num		Curved Suture				Intersected Suture			|Error|	Curved Suture			1 Intersection		2 Intersections		|Error|
		Exp1	Exp2	Exp3	Exp4	Exp5	Exp6	Exp7		Exp1	Exp2	Exp3	Exp4	Exp5	Exp6	Exp7	
BG/1	w	112.40	111.64	109.12	112.89	108.23	109.51	112.69	1.82	188.17	191.12	187.08	191.65	188.43	194.34	190.03	1.92
	w/o	117.90	113.28	106.22	109.49	107.96	112.60	115.76	3.70	188.67	194.27	192.08	194.92	183.30	200.79	197.88	5.42
BG/2	w	110.26	109.88	109.08	108.89	110.99	110.97	108.75	0.80	190.18	191.28	190.16	189.95	190.81	190.20	191.42	0.59
	w/o	108.13	113.11	115.31	116.66	122.44	116.19	106.71	5.55	192.81	198.02	197.48	194.71	197.58	193.27	199.97	6.26
BG/3	w	112.27	108.97	110.57	111.37	111.58	110.43	109.19	1.15	191.36	190.86	188.92	186.55	190.47	188.91	188.93	1.34
	w/o	105.04	108.20	113.19	106.97	107.26	113.16	105.12	3.39	201.09	189.71	186.38	192.38	198.96	199.03	193.68	5.58
BG/4	w	111.09	112.36	112.66	110.57	109.45	109.11	110.57	1.24	192.28	189.16	190.91	190.64	192.32	188.22	191.51	1.47
	w/o	114.03	115.95	114.82	106.03	103.94	113.29	112.68	4.40	198.05	199.14	196.45	195.02	186.06	193.21	198.44	6.32
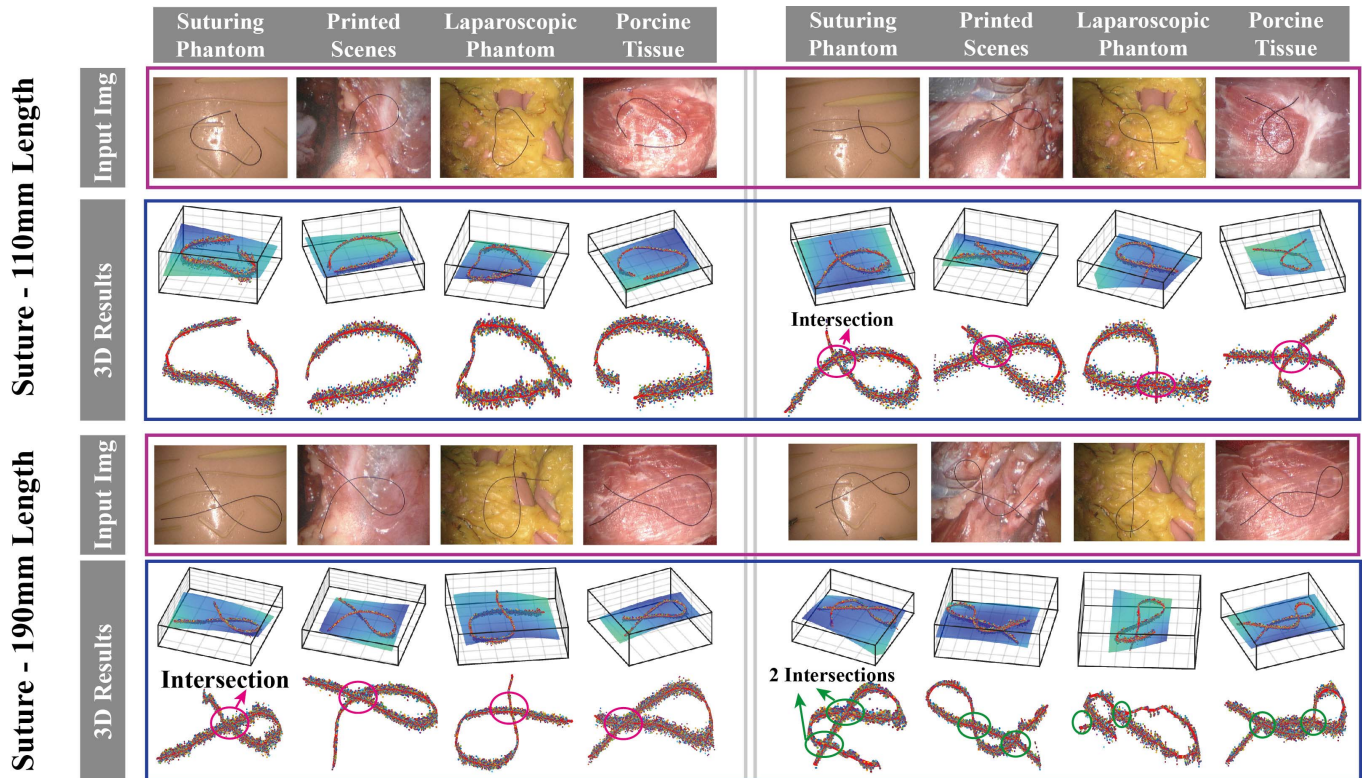
Fig. 9. We put sutures with different lengths on different backgrounds, and show typical results in which sutures are oriented with multiple shapes. The input images are from ECM, and 3D results contain the reconstructed 3D shape with its resting surface, and the zoomed-in shape with multiple sampling points.
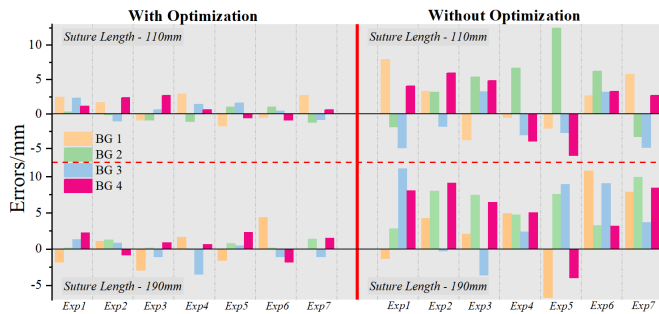

Fig. 10. Error comparisons of suture's 3D length computation with and without SSP optimization under different unstructured environments.

and adopt Pearson Coefficient to evaluate the linear correlation between them. For errors in X/Y directions relative to the time spending, we obtain low correlations which are only 24.7%

and 28.3%. For Z direction and total errors, they are 69.5% and 73.8%. Intuitively, multiple trials are conducted in each depth, hence the time spending may substantially grow when larger errors occur in the depth direction. However, such time spending is essential at the current stage to maintain a reliable grasping, which sets a solid foundation for the following works. Although the visual servoing may potentially save the operational time, it requires quite accurate and real-time tracking of surgical tools under complex environments, which is another popular topic yet out of the scope in this article. Additionally, SIMI values calculated by Eq.(20) are also plotted in Fig. 11(c3), and we can notice they all under 0.8, which proves the SIMI threshold is empirically reliable.

It is worthy noticing that there is one failure case as is highlighted in Table. IV when testing on the laparoscopic torso phantom. The main reason comes from the rugged

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

12   IEEE TRANSACTIONS ON AUTOMATION SCIENCE AND ENGINEERING

TABLE IV

GRASPING RESULTS REGARDING: *E1*: SUCCESSFULLY GRASP THE SUTURE OR NOT; *E2*: GRASPING TIME (UNIT: SECONDS); *E3*: TOTAL ERROR BETWEEN THE TRUE GRASPING POINT AND THE COMPUTED POSITION (UNIT: MM); *E4*: SIMI VALUE WHEN THE SUTURE IS GRASPED

BG	Suturing Phantom 1				Suturing Phantom 2				Suturing Phantom 3				Laparoscopic Phantom				Porcine Tissue			
Num	*E1*	*E2*	*E3*	*E4*	E1	E2	E3	E4	E1	E2	E3	E4	E1	E2	E3	E4	E1	E2	E3	E4
#1	Y	26.6	7.11	61.5%	Y	3.2	1.45	61.7%	Y	28.7	4.72	52.8%	Y	46.2	6.47	69.2%	Y	45.2	9.55	53.8%
#2	Y	25.4	2.76	59.1%	Y	3.5	0.95	71.5%	Y	48.4	6.76	72.0%	Y	25.9	4.47	75.2%	Y	43.9	9.87	60.2%
#3	Y	23.8	4.91	67.2%	Y	24.0	4.88	64.0%	Y	45.7	6.89	54.1%	Y	23.9	4.52	74.4%	Y	30.2	11.51	61.3%
#4	Y	29.7	10.74	62.4%	Y	3.1	1.15	72.4%	Y	29.1	6.45	71.1%	Y	27.4	6.81	61.8%	Y	23.3	4.64	54.5%
#5	Y	31.0	10.47	75.6%	Y	23.8	4.77	67.4%	Y	51.3	11.13	51.3%	Y	23.4	4.76	65.3%	Y	53.8	10.87	79.5%
#6	Y	37.4	6.85	69.7%	Y	23.8	5.07	61.5%	Y	50.2	10.35	66.9%	Y	37.5	6.81	75.9%	Y	43.9	9.97	65.5%
#7	Y	38.8	10.43	71.2%	Y	45.3	10.13	78.6%	Y	42.0	9.52	76.6%	N	N/A	N/A	N/A	Y	27.1	6.34	61.8%
#8	Y	21.6	4.31	73.0%	Y	44.9	9.55	61.2%	Y	38.8	9.48	73.7%	Y	42.4	10.01	63.6%	Y	41.7	9.44	66.9%
#9	Y	34.8	6.53	70.8%	Y	48.1	11.77	76.9%	Y	21.8	4.85	79.2%	Y	31.7	6.93	68.8%	Y	23.2	4.52	73.1%
#10	Y	7.0	5.19	73.8%	Y	57.7	10.68	61.4%	Y	46.9	10.95	72.1%	Y	22.9	4.84	66.1%	Y	46.2	10.82	64.1%

(a)

Suturing Phantom 1   Suturing Phantom 2   Laparoscopic Phantom   Porcine Tissue

X/Y/Z/Total grasping errors in all experiments. (b1)-(b5) accordingly denote on different backgrounds of: suturing phantom 1~3, laparoscopic phantom, and porcine tissue. The broken part in (b4) indicates the one failure case.

(c1). Average X/Y/Z/Total grasping errors and their distributions in all suture grasping experiments. We use the box chart to show their min/max values, 25%~75% ranges and average error values. (c2). We plot X/Y/Z/Total grasping errors with respect to the grasping time. (c3). SIMI values when the suture is successfully grasped in all experiments.
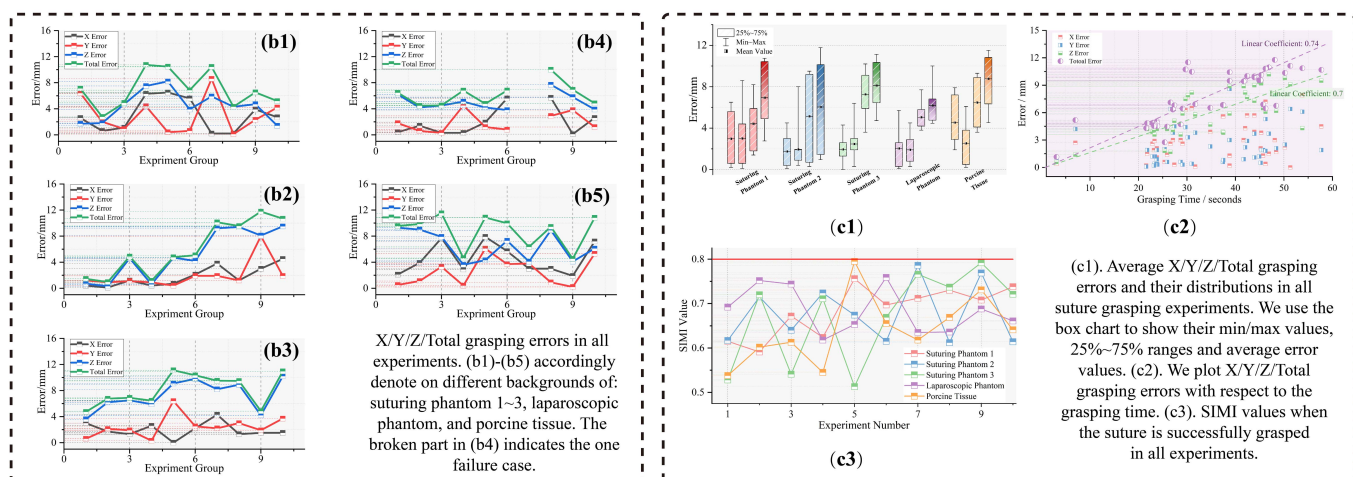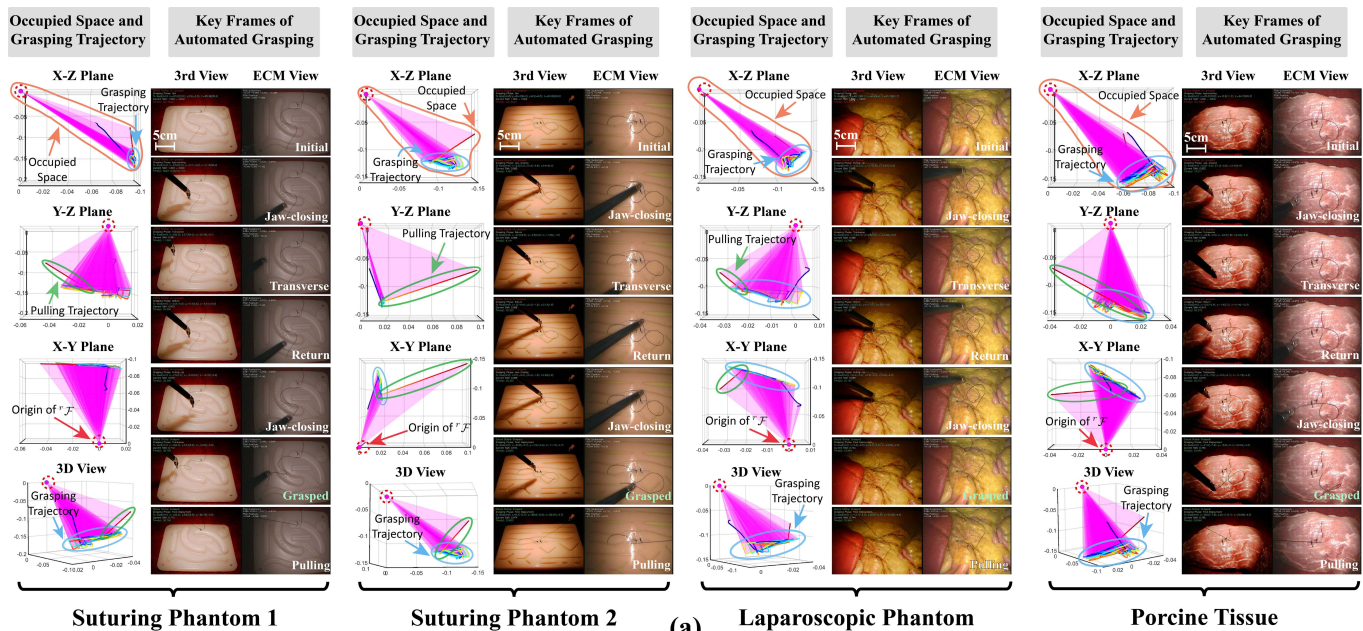
Fig. 11.   (a). Typical snapshots of automated suture grasping under different backgrounds along with their occupied work-space from multiple views; (b1-b5). Grasping errors in all experiments; (c1-c3). Grasping error analysis, grasping time versus grasping errors, and SIMI values when successfully grasp the suture.

and complex structures around the grasping region. When conducting this task, our framework focuses on the perception and manipulation of the suture, without adding the exact sensing and reconstruction of the surrounding space. When the end-effector touches other surrounding tissues/organs, our vision-based mechanism can keenly observe such tiny

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

LU *et al.*: TOWARD IMAGE-GUIDED AUTOMATED SUTURE GRASPING UNDER COMPLEX ENVIRONMENTS

13

environmental changes and terminate the grasping trials to ensure the manipulation safety. In our future works, we can further improve our perception ability in surrounding environment reconstructions, so as to achieve a versatile online planning and grasping approach to handle more challenging conditions.

V. CONCLUSION

In this paper, a learning-driven and online optimization-based framework is introduced to precept the 3D information of a suture. Its segmentation capability and 3D shape computational accuracy for the slender suture have been accordingly tested based on different suturing phantoms, laparoscopic torso, and porcine tissues. Based on satisfactory perceptions, we then propose a novel grasping strategy with a vision-based mechanism to automate the suture grasping operation using the dVRK system. This is a pioneering study that integrates vision, deep learning, and automation techniques to bridge the automated suture stitching and looping procedures. Through extensive experiments, it is proved that our framework is feasible for this sub-task in surgical knot tying with a high successful rate of grasping in practice.

In the next step, automated looping techniques presented in our previous paper [13] can be embedded into the framework and further achieve a double-loop suture through dual arms motion plannings and dynamic manipulations. By exploring a control strategy which considers simultaneous motions of robot end-effectors and the looped suture, we are planning to design a suture's end grasping method by integrating the learning-driven endpoint detection method that we proposed in [28], which hence can facilitate the completion of a single surgical square knot. To repetitively conduct knot tying along the wound, popular needle detection approaches, e.g., [24], can be adopted to automate needle's localization, grasping, and wound stitching operations. Additionally, dynamic 2D segmentation and 3D reconstruction methods that leverage the spatial and temporal information of robotic arms can be further investigated to precept and track the whole suture when it is partially occluded by instruments. By building up these technologies, a fully automated knot tying workflow can be developed to contribute one step further towards a higher-level of task autonomy in robotic surgery.

ACKNOWLEDGMENT

REFERENCES

[1] R. H. Taylor, A. Menciassi, G. Fichtinger, P. Fiorini, and P. Dario, "Medical robotics and computer-integrated surgery," in *Springer Handbook of Robotics*. Cham, Switzerland: Springer, 2016, pp. 1657–1684.

[2] G. I. Barbash and S. A. Glied, "New technology and health care costs—The case of robot-assisted surgery," *New England J. Med.*, vol. 363, no. 8, pp. 701–704, Aug. 2010.

[3] S. A. Scheib and N. F. Amanda, "Gynecologic robotic laparoendoscopic single-site surgery: Prospective analysis of feasibility, safety, and technique," *Amer. J. Obstetrics Gynecol.*, vol. 212, no. 2, p. 179-e1, 2015.

[4] G. S. Weinstein, B. W. O'Malley, Jr., and N. G. Hockstein, "Transoral robotic surgery: Supraglottic laryngectomy in a canine model," *Laryngoscope*, vol. 115, no. 7, pp. 1315–1319, Jul. 2005.

[5] J. P. Hubschman, A. Tsirbas, and S. D. Schwartz, "Robotic surgery in ophthalmology," *Retina Today*, pp. 81–84, May/Jun. 2008.

[6] G.-Z. Yang et al., "Medical robotics—Regulatory, ethical, and legal considerations for increasing levels of autonomy," *Sci. Robot.*, vol. 2, no. 4, p. 8638, 2017.

[7] H. Ren, D. Rank, M. Merdes, J. Stallkamp, and P. Kazanzides, "Multisensor data fusion in an integrated tracking system for endoscopic surgery," *IEEE Trans. Inf. Technol. Biomed.*, vol. 16, no. 1, pp. 106–111, Jan. 2012.

[8] F. Alambeigi, Z. Wang, Y.-H. Liu, R. H. Taylor, and M. Armand, "Toward semi-autonomous cryoablation of kidney tumors via model-independent deformable tissue manipulation technique," *Ann. Biomed. Eng.*, vol. 46, no. 10, pp. 1650–1662, Oct. 2018.

[9] C. Shin, P. W. Ferguson, S. A. Pedram, J. Ma, E. P. Dutson, and J. Rosen, "Autonomous tissue manipulation via surgical robot using learning based model predictive control," in *Proc. Int. Conf. Robot. Automat. (ICRA)*, May 2019, pp. 3875–3881.

[10] Y. Kassahun et al., "Surgical robotics beyond enhanced dexterity instrumentation: A survey of machine learning techniques and their role in intelligent and autonomous surgical actions," *Int. J. Comput. Assist. Radiol. Surg.*, vol. 11, no. 4, pp. 553–568, Apr. 2016.

[11] D. A. Hashimoto, G. Rosman, D. Rus, and O. R. Meireles, "Artificial intelligence in surgery: Promises and perils," *Ann. Surg.*, vol. 268, no. 1, pp. 70–76, 2018.

[12] C. R. Wagner and R. D. Howe, "Force feedback benefit depends on experience in multiple degree of freedom robotic surgery task," *IEEE Trans. Robot.*, vol. 23, no. 6, pp. 1235–1240, Dec. 2007.

[13] L. Bo, H. K. Chu, K. C. Huang, and C. Li, "Vision-based surgical suture looping through trajectory planning for wound suturing," *IEEE Trans. Autom. Sci. Eng.*, vol. 16, no. 2, pp. 542–556, Apr. 2019.

[14] H. Kang and J. T. Wen, "Autonomous suturing using minimally invasive surgical robots," in *Proc. IEEE Int. Conf. Control Appl.*, Sep. 2000, pp. 742–747.

[15] H. Kang and J. T. Wen, "Robotic assistants aid surgeons during minimally invasive procedures," *IEEE Eng. Med. Biol. Mag.*, vol. 20, no. 1, pp. 94–104, Jan./Feb. 2001.

[16] E. C. McLemore et al., "Transanal total mesorectal excision (taTME) for rectal cancer: A training pathway," *Surgical Endoscopy*, vol. 30, no. 9, pp. 4130–4135, 2016.

[17] T. Tulandi and J. I. Einarsson, "The use of barbed suture for laparoscopic hysterectomy and myomectomy: A systematic review and meta-analysis," *J. Minimally Invasive Gynecol.*, vol. 21, no. 2, pp. 210–216, Mar. 2014.

[18] N. Bonaros et al., "Five hundred cases of robotic totally endoscopic coronary artery bypass grafting: Predictors of success and safety," *Ann. Thoracic Surg.*, vol. 95, no. 3, pp. 803–812, Mar. 2013.

[19] T. Osa, N. Sugita, and M. Mitsuishi, "Online trajectory planning and force control for automation of surgical tasks," *IEEE Trans. Autom. Sci. Eng.*, vol. 15, no. 2, pp. 675–691, Apr. 2018.

[20] S. A. Pedram, P. Ferguson, J. Ma, E. Dutson, and J. Rosen, "Autonomous suturing via surgical robot: An algorithm for optimal selection of needle diameter, shape, and path," in *Proc. IEEE Int. Conf. Robot. Automat. (ICRA)*, May 2017, pp. 2391–2398.

[21] D. Navarro-Alarcon et al., "Automatic 3-D manipulation of soft objects by robotic arms with an adaptive deformation model," *IEEE Trans. Robot.*, vol. 32, no. 2, pp. 429–441, Apr. 2016.

[22] F. Alambeigi, Z. Wang, R. Hegeman, Y.-H. Liu, and M. Armand, "Autonomous data-driven manipulation of unknown anisotropic deformable tissues using unmodelled continuum manipulators," *IEEE Robot. Autom. Lett.*, vol. 4, no. 2, pp. 254–261, Apr. 2019.

[23] S. Leonard, K. L. Wu, Y. Kim, A. Krieger, and P. C. W. Kim, "Smart tissue anastomosis robot (STAR): A vision-guided robotics system for laparoscopic suturing," *IEEE Trans. Biomed. Eng.*, vol. 61, no. 4, pp. 1305–1317, Apr. 2014.

[24] S. Sen, A. Garg, D. V. Gealy, S. McKinley, Y. Jen, and K. Goldberg, "Automating multi-throw multilateral surgical suturing with a mechanical needle guide and sequential convex optimization," in *Proc. IEEE Int. Conf. Robot. Automat. (ICRA)*, May 2016, pp. 4178–4185.

[25] A. Kumar Tanwani, P. Sermanet, A. Yan, R. Anand, M. Phielipp, and K. Goldberg, "Motion2Vec: Semi-supervised representation learning from surgical videos," 2020, *arXiv:2006.00545*.

[26] A. Shademan, R. S. Decker, J. D. Opfermann, S. Leonard, A. Krieger, and P. C. W. Kim, "Supervised autonomous robotic soft tissue surgery," *Sci. Transl. Med.*, vol. 8, no. 337, May 2016, Art. no. 337ra64.

[27] B. Lu, H. K. Chu, K. Huang, and J. Lai, "Surgical suture thread detection and 3-D reconstruction using a model-free approach in a calibrated stereo visual system," *IEEE/ASME Trans. Mechatronics*, vol. 25, no. 2, pp. 792–803, Apr. 2020.

[28] B. Lu, X. B. Yu, J. W. Lai, K. C. Huang, K. C. C. Chan, and H. K. Chu, "A learning approach for suture thread detection with feature enhancement and segmentation for 3-D shape reconstruction," *IEEE Trans. Autom. Sci. Eng.*, vol. 17, no. 2, pp. 858–870, Apr. 2020.

[29] B. Lu et al., "A learning-driven framework with spatial optimization for surgical suture thread reconstruction and autonomous grasping under multiple topologies and environmental noises," in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst. (IROS)*, Oct. 2020, pp. 3075–3082.

[30] D. Seita, S. Krishnan, R. Fox, S. McKinley, J. Canny, and K. Goldberg, "Fast and reliable autonomous surgical debridement with cable-driven robots using a two-phase calibration procedure," in *Proc. IEEE Int. Conf. Robot. Automat. (ICRA)*, May 2018, pp. 6651–6658.

[31] A. Hernandez-Vela et al., "Accurate coronary centerline extraction, caliber estimation, and catheter detection in angiographies," *IEEE Trans. Inf. Technol. Biomed.*, vol. 16, no. 6, pp. 1332–1340, Nov. 2012.

[32] M. Hoffmann, A. Brost, M. Koch, F. Bourier, A. Maier, K. Kurzidim, N. Strobel, and J. Hornegger, "Electrophysiology catheter detection and reconstruction from two views in fluoroscopic images," *IEEE Trans. Med. Imag.*, vol. 35, no. 2, pp. 567–579, Feb. 2015.

[33] H. Lee, M. Mansouri, S. Tajmir, M. H. Lev, and S. Do, "A deep-learning system for fully-automated peripherally inserted central catheter (PICC) tip detection," *J. Digit. Imag.*, vol. 31, no. 4, pp. 393–402, 2018.

[34] H. Li and A. Yezzi, "Vessels as 4-D curves: Global minimal 4-D paths to extract 3-D tubular surfaces and centerlines," *IEEE Trans. Med. Imag.*, vol. 26, no. 9, pp. 1213–1223, Sep. 2007.

[35] V. Kaul, A. Yezzi, and Y. C. Tsai, "Detecting curves with unknown endpoints and arbitrary topology using minimal paths," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 34, no. 10, pp. 1952–1965, Oct. 2012.

[36] Q. Li, B. Feng, L. Xie, P. Liang, H. Zhang, and T. Wang, "A cross-modality learning approach for vessel segmentation in retinal images," *IEEE Trans. Med. Imag.*, vol. 35, no. 1, pp. 109–118, Jan. 2016.

[37] Z. Yan, X. Yang, and K.-T. Cheng, "Joint segment-level and pixel-wise losses for deep learning based retinal vessel segmentation," *IEEE Trans. Biomed. Eng.*, vol. 65, no. 9, pp. 1912–1923, Sep. 2018.

[38] S. Izadi et al., "KinectFusion: Real-time 3D reconstruction and interaction using a moving depth camera," in *Proc. 24th Annu. ACM Symp. User Interface Softw. Technol.*, 2011, pp. 559–568.

[39] U. Asif, M. Bennamoun, and F. A. Sohel, "RGB-D object recognition and grasp detection using hierarchical cascaded forests," *IEEE Trans. Robot.*, vol. 33, no. 3, pp. 547–564, Jun. 2017.

[40] F. Endres, J. Hess, J. Sturm, D. Cremers, and W. Burgard, "3-D mapping with an RGB-D camera," *IEEE Trans. Robot.*, vol. 30, no. 1, pp. 177–187, Feb. 2014.

[41] W. Hao Lui and A. Saxena, "Tangled: Learning to untangle ropes with RGB-D perception," in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst.*, Nov. 2013, pp. 837–844.

[42] R. C. Jackson, R. Yuan, D.-L. Chow, W. S. Newman, and M. C. Çavuşoğlu, "Real-time visual tracking of dynamic surgical suture threads," *IEEE Trans. Autom. Sci. Eng.*, vol. 15, no. 3, pp. 1078–1090, Jul. 2018.

[43] N. Padoy and G. D. Hager, "3D thread tracking for robotic assistance in tele-surgery," in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst.*, Sep. 2011, pp. 2102–2107.

[44] L. Zaidi, J. A. Corrales, B. C. Bouzgarrou, Y. Mezouar, and L. Sabourin, "Model-based strategy for grasping 3D deformable objects using a multi-fingered robotic hand," *Robot. Auton. Syst.*, vol. 95, pp. 196–206, Sep. 2017.

[45] C. Papazov, S. Haddadin, S. Parusel, K. Krieger, and D. Burschka, "Rigid 3D geometry matching for grasping of known objects in cluttered scenes," *Int. J. Robot. Res.*, vol. 31, no. 4, pp. 538–553, Apr. 2012.

[46] D. Song, C. H. Ek, K. Huebner, and D. Kragic, "Task-based robot grasp planning using probabilistic inference," *IEEE Trans. Robot.*, vol. 31, no. 3, pp. 546–561, Jun. 2015.

[47] K. Hang et al., "Hierarchical fingertip space: A unified framework for grasp planning and in-hand grasp adaptation," *IEEE Trans. Robot.*, vol. 32, no. 4, pp. 960–972, Aug. 2016.

[48] Y. Bekiroglu, J. Laaksonen, J. A. Jorgensen, V. Kyrki, and D. Kragic, "Assessing grasp stability based on learning and haptic data," *IEEE Trans. Robot.*, vol. 27, no. 3, pp. 616–629, Jun. 2011.

[49] M. Andrychowicz et al., "Learning dexterous in-hand manipulation," *Int. J. Robot. Res.*, vol. 39, no. 1, pp. 3–20, 2020.

[50] C. D'Ettorre et al., "Automated pick-up of suturing needles for robotic surgical assistance," in *Proc. IEEE Int. Conf. Robot. Automat. (ICRA)*, May 2018, pp. 1370–1377.

[51] O. Ozguner et al., "Camera-robot calibration for the da Vinci robotic surgery system," *IEEE Trans. Autom. Sci. Eng.*, vol. 17, no. 4, pp. 2154–2161, Oct. 2020.

[52] M. Hwang et al., "Efficiently calibrating cable-driven surgical robots with RGBD fiducial sensing and recurrent neural networks," *IEEE Robot. Autom. Lett.*, vol. 5, no. 4, pp. 5937–5944, Oct. 2020.

[53] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional networks for biomedical image segmentation," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent.*, 2015, pp. 234–241.

[54] A. Tarvainen and H. Valpola, "Mean teachers are better role models: Weight-averaged consistency targets improve semi-supervised deep learning results," in *Proc. Adv. Neural Inf. Process. Syst.*, 2017, pp. 1195–1204.

[55] W. R. Crum, O. Camara, and D. L. G. Hill, "Generalized overlap measures for evaluation and validation in medical image analysis," *IEEE Trans. Med. Imag.*, vol. 25, no. 11, pp. 1451–1461, Nov. 2006.

[56] M. Berman, A. R. Triki, and M. B. Blaschko, "The Lovász–Softmax loss: A tractable surrogate for the optimization of the intersection-over-union measure in neural networks," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 4413–4421.

[57] D. B. Johnson, "A note on Dijkstra's shortest path algorithm," *J. ACM*, vol. 20, pp. 385–388, Jul. 1973.

[58] R. Y. Tsai and R. K. Lenz, "A new technique for fully autonomous and efficient 3D robotics hand/eye calibration," *IEEE Trans. Robot. Autom.*, vol. 5, no. 3, pp. 345–358, Jun. 1989.

[59] J. H. Mathews and K. D. Fink, *Numerical Methods Using MATLAB*, vol. 4. Upper Saddle River, NJ, USA: Pearson, 2004.

Bo Lu received the B.Eng. degree from the Department of Ship and Offshore Engineering, Dalian University of Technology, Liaoning, China, in 2013, and the M.S. (Hons.) and Ph.D. degrees from the Department of Mechanical Engineering, The Hong Kong Polytechnic University, Hong Kong, in 2015 and 2019, respectively. Afterward, he worked as a Post-Doctoral Research Fellow with the T Stone Robotics Institute, The Chinese University of Hong Kong, Hong Kong. He is currently an Associate Professor with the Robotics and Microsystems Center, School of Mechanical and Electric Engineering, Soochow University. His current research interests include medical robotics, computer vision, and vision-based and learning-driven automation and intervention.

Bin Li received the B.E. degree from the Department of Mechatronics Engineering, Harbin Institute of Technology, China, in 2016, and the M.S. degree in mechanical engineering from Shanghai Jiao Tong University, China, in 2019. He is currently pursuing the Ph.D. degree with the Department of Mechanical and Automation Engineering, The Chinese University of Hong Kong. His research interests include surgical intelligence, robotics, automation and control, and reinforcement learning.

Wei Chen received the B.E. degree in computer science and technology from Zhengzhou University, Zhengzhou, China, in 2012. Since March 2018, he has been a Research Assistant with the Department of Mechanical and Automation Engineering, The Chinese University of Hong Kong, Hong Kong. His research interests include surgical robots, computer vision, and 3-D reconstruction.

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

LU *et al.*: TOWARD IMAGE-GUIDED AUTOMATED SUTURE GRASPING UNDER COMPLEX ENVIRONMENTS   15

Yueming Jin (Member, IEEE) received the B.Eng. degree from the Department of Biomedical Engineering, Northeastern University, China, in 2015, and the Ph.D. degree from the Department of Computer Science and Engineering, The Chinese University of Hong Kong, in 2019. She is currently a Post-Doctoral Research Fellow with The Chinese University of Hong Kong. Her research interests include artificial intelligence and its applications on medical image computing, and robotic video perception.

Zixu Zhao received the B.Eng. degree from the Department of Electronic and Information Engineering, Beihang University, China, in 2017. He is currently pursuing the Ph.D. degree with the Department of Computer Science and Engineering, The Chinese University of Hong Kong. His research interests include deep learning in medical image computing, model efficiency, and robotic video analysis.

Qi Dou (Member, IEEE) received the B.Eng. degree in biomedical engineering from Beihang University, Beijing, China, in 2014, and the Ph.D. degree from the Department of Computer Science and Engineering, The Chinese University of Hong Kong, Hong Kong, in 2018. She is currently an Assistant Professor with the Department of Computer Science and Engineering, The Chinese University of Hong Kong. Her research interests are in the interdisciplinary fields of medical image analysis and artificial intelligence, for improving lesion detection, anatomical structure computation, and surgical robotics perception, with an impact to advance disease diagnosis and robot-assisted intervention via machine intelligence.

Pheng-Ann Heng (Senior Member, IEEE) received the Ph.D. degree in computer science from Indiana University, Indianapolis, IN, USA. He is currently a Professor with the Department of Computer Science and Engineering, The Chinese University of Hong Kong, Hong Kong, where he is the Director of the Virtual Reality, Visualization, and Imaging Research Centre. He is also the Director of the Research Center for Human–Computer Interaction, Shenzhen Institute of Advanced Integration Technology, Chinese Academy of Sciences, Shenzhen, China. His research interests include virtual reality applications in medicine, visualization, medical imaging, human–computer interfaces, rendering and modeling, interactive graphics, and animation.

Yunhui Liu (Fellow, IEEE) received the B.Eng. degree from the Beijing Institute of Technology, the M.Eng. degree from Osaka University, and the Ph.D. degree from The University of Tokyo in 1992. After working with the Electrotechnical Laboratory, Japan, as a Research Scientist, he joined The Chinese University of Hong Kong (CUHK) in 1995, where he is currently a Choh-Ming Li Professor of Mechanical and Automation Engineering and the Director of the T Stone Robotics Institute. He is also an Adjunct Professor with the State Key Laboratory of Robotics Technology and System, Harbin Institute of Technology, China. He has published over 300 papers in refereed journals and refereed conference proceedings, and was listed in the Highly Cited Authors (Engineering) by Thomson Reuters in 2013. His research interests include visual servoing, medical robotics, multifingered grasping, mobile robots, and machine intelligence. He has received numerous research awards in international journals on robotics and automation from international conferences and government agencies.