

Real-Time Vision-Based Surgical Tool Segmentation with Robot Kinematics Prior

Yun-Hsuan Su, Kevin Huang, Blake Hannaford

Abstract—Robot-assisted minimally invasive surgery combines the skills and techniques of highly-trained surgeons with the robustness and precision of machines. Several advantages include precision beyond human dexterity alone, greater kinematic degrees of freedom at the surgical tool tip, and possibilities in remote surgical practices through teleoperation. Nevertheless, obtaining accurate force feedback during surgical operations remains a challenging hurdle. Though direct force sensing using tool tip mounted sensors is theoretically possible, it is not amenable to required sterilization procedures. Vision-based force estimation according to real-time analysis of tissue deformation serves as a promising alternative. In this application, along with numerous related research in robot-assisted minimally invasive surgery, segmentation of surgical instruments in endoscopic images is a prerequisite. Thus, a surgical tool segmentation algorithm robust to partial occlusion is proposed using DFT shape matching of robot kinematics shape prior (u) fused with log likelihood mask (Q) in the Opponent color space to generate final mask (U). Implemented on the Raven II surgical robot system, a real-time performance robust to tool tip orientation and up to 6 fps without GPU acceleration is achieved.

I. INTRODUCTION

Vision-based force estimation for robot manipulated surgical procedures relies on tissue indentation measurements associated with surgical contacts. Endoscopic images can provide information from which indentation can be inferred, and applied force can subsequently be calculated and applied for realistic force feedback. A natural initial step towards this end involves tool segmentation. There are existing methods regarding marker-less surgical tool-tip segmentation with robot kinematics prior [1] [2] [3]. Yet, based on the authors' knowledge, this paper is the first to simultaneously offer:

- low computational complexity at 6Hz without use of GPU acceleration.
- segmentation performance for real-time implementation with average Sørensen-Dice index greater than 0.73.
- surgical instrument tracking algorithm amenable to Raven II tools.
- robustness to partial occlusion by fusing robot kinematics and color filtering.

A. Proposed System Workflow

The proposed entire visual force estimation method can be divided into four stages, as shown in Fig.2. Stage 1 focuses

on classifying image pixels as either tool or tissue pixels. The tissue pixels neighboring the surgical tool tip define the region of interest (ROI). Stage 2 deals with generating tissue depth information of the ROI. Any resultant deformation can then be inferred, and a deformation map centered around the surgical tool can be generated. In stage 3, deformation information can be used in tandem with tissue dynamics models to estimate applied force [4]. Finally, stage 4 will implement this real-time force estimation via haptic feedback in a bilateral teleoperation scheme. The work presented here focuses on stage 1, real-time surgical tool segmentation. Subsequent stages rely on this efficient extraction of tissue data immediately surrounding the tool tip. An overview of this stage is shown below in Fig.1.

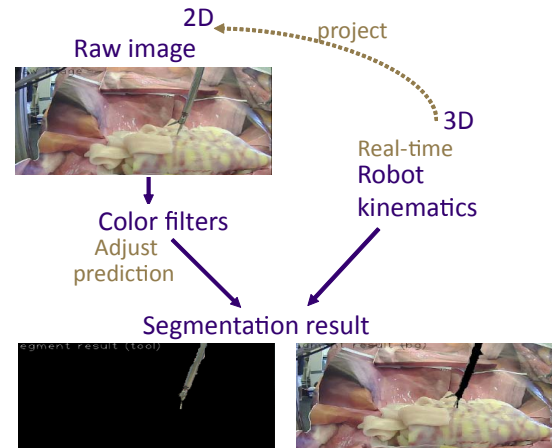


Fig. 1: Surgical tool segmentation overview.

The technical implementation of stage 1 are described in more detail in sections I-D and II.

B. Background

Surgical instrument tracking and isolation is needed for robot-assisted minimally invasive surgery research that focuses on visual deformation analysis at the tool-tissue interface. A mere bounding box around the tissue of interest is insufficient, since tool-pixels will interfere with deformation calculations. To analyze only non-tool pixels, online classification and segmentation of tool pixels within this ROI in real-time is critical.

Online surgical tool segmentation approaches may face challenges including motion blur, partial occlusion, specular reflections on wet tissue surfaces [5], as well as lighting changes. Furthermore, metallic parts of surgical tools reflect tissue colors, which increases the difficulty of color filtering.

Yun-Hsuan Su and Blake Hannaford are with the University of Washington Department of Electrical Engineering, 185 Stevens Way, Paul Allen Center - Room AE100R, Campus Box 352500, Seattle, WA 98195-2500, USA. {yhsu83, blake}@uw.edu

Kevin Huang is with Trinity College, Dept. of Engineering, 300 Summit St, Hartford, CT 06106 USA kevin.huang@trincoll.edu
978-1-5386-2512-5/18/\$31.00 ©2018 IEEE

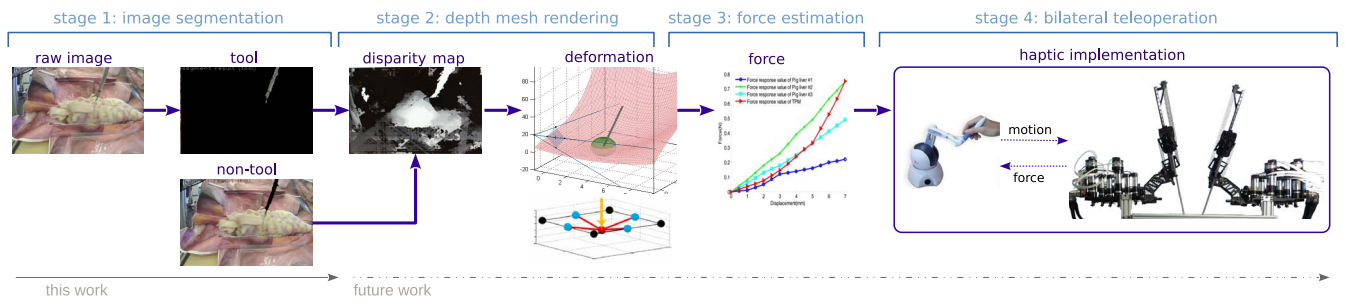


Fig. 2: Illustration of the four stages or subtasks involved in the proposed vision-based force estimation method. The work presented here completes stage 1 and parts of stage 2.

C. Related Work

Existing approaches to address the aforementioned challenges can be broadly classified into four methods. According to Bouget et al., the first distinction separates marker and marker-less [6]. Methods that use markers can be further distinguished by use of either a visual marker or a non-visual marker. Similarly, marker-less techniques can optimally employ tracking algorithms. Fig.3 illustrates the coarse spectrum of these techniques.

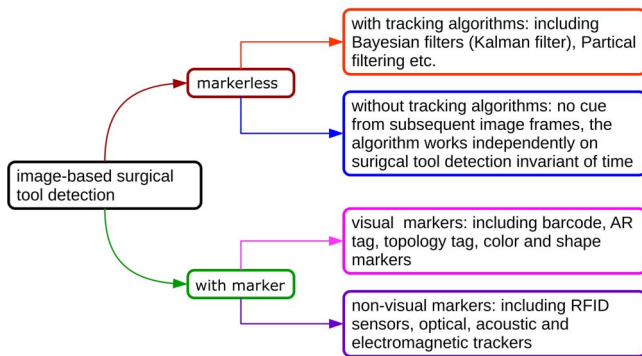


Fig. 3: Different approaches for image-based surgical instrument detection.

Methods using visual markers are very effective. However, the cylindrical shape of surgical tool shafts make distortion inevitable for most 2D markers that conform to the tool surface. Topology markers, however, stand out in instrument detection due to their robustness to tag distortion [7]. Nonetheless, visual markers are subject to occlusion, which are commonplace during surgical operations when the tool is stained with blood. Non-visual markers do not have this problem, and include RFID sensors [8], acoustic [9] or electromagnetic trackers [10]. Concerns with non-visual markers include price, size, and minimizing external electronics placed within patients' bodies. Because of these issues, these sensors or trackers can be undesirable.

Despite the effectiveness of markers, they are not amenable to direct transition to current hospital procedures and clinical setup. Instead, a marker-less surgical instrument segmentation technique based on robotic kinematics and color filtering is proposed in this work.

D. Kinematics Prior Segmentation

In this work, the entire tool configuration estimate (both 3D position of the tool tip and all joint angles) was available from joint encoders and kinematic calculations on the Raven II Surgical Robot System [11]. Given the camera extrinsic parameters with respect to the Raven II base frame, the robot kinematics provided a shape prior (u). This shape prior provided an estimate of the expected surgical instrument location and shape within the camera image frame. This was achieved by projecting the 3D surgical tool links onto the 2D image plane. The predicted shape prior was subsequently modified to match a color filtering mask (Q) in the frequency domain in two steps - modifying translation and modifying rotation and scale. The latter requires transformation to log-polar coordinates. Finally, with this modified shape prior mask, U , a color mask was applied for pixel-wise classification within the surgical tool region. The work done here accomplishes all portions of tasks depicted in Fig.1. Results are promising towards accurate real-time tool segmentation.

II. METHODS

A. Camera Pose Estimation

In order to use robot kinematics to ascertain the surgical instrument's location within the camera image plane, defining the camera frame with respect to the robot base frame is required. In what is often referred to as the perspective-n-point (PNP) problem, the aim is to estimate the pose of an object given n 3D points on the object and their corresponding 2D projections onto the image plane. This process also requires the camera intrinsic parameters. Instead, we require the converse, that is to determine camera pose with respect to the object.

To accomplish this task, the 2D (x, y) projections were generated from the OpenCV function `cvFindChessboardCorners` [12] to detect checkerboard corners in the image frame, as illustrated in Fig.4. The 3D corner locations of the 48 checkerboard corners were obtained by manually measuring the corner positions with respect to the Raven II base frame. Combined with the camera intrinsic parameters, determining the transformation matrix between robot and camera frame is trivial. Fig.4 illustrates the two coordinate frames.

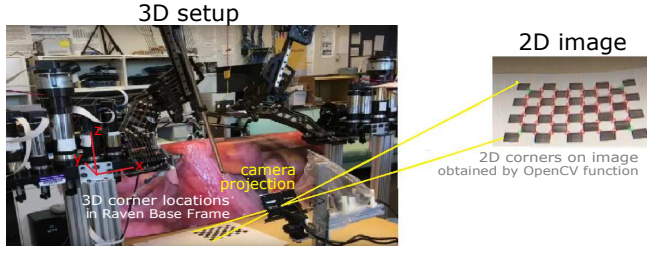


Fig. 4: The 2D and 3D coordinate input of checkerboard corners to the PNP algorithm. From this, the transformation from Raven II base frame to camera frame is obtained.

B. Kinematics Shape Prior Mask

Given robot joint states, forward kinematics, and camera pose, a raw projection of joint locations onto the camera image plane is straightforward. Then, from the physical thickness of each robot link, the perceived thickness on either end of a robot link in the image is inferred respectively. Suppose an object point with known width W is distance D from camera with focal length F . Then the apparent width in camera pixels, P , is defined as

$$P = FW/D \quad (1)$$

The overall shape of the projected robot tool can be obtained with simple trigonometry. The union of these pixels forms the initial shape prior mask, u . This is more computationally efficient than projecting all points on the tool surface.

C. Log-likelihood Color Mask

Once the robot kinematics shape prior u is generated, color filtering across the entire image further refines the prior estimate. The color filtering scheme adapted for this work was based upon work by Van De Sande et al., which claimed that hue and saturation in the HSV colorspace and Opponent₁ and Opponent₂ (denoted O_1 and O_2) are colorspace components providing the most discriminative power to separate surgical tool pixels from background pixels [13], where:

$$\begin{aligned} O_1 &= G - R \\ O_2 &= B - Y = B - (G + R) \end{aligned}$$

The log-likelihood mask (Q) is then defined as

$$Q = w_1 H + w_2 S + w_3 O_1 + w_4 O_2$$

where H and S are the hue and saturation components respectively and the weights w_1, w_2, w_3, w_4 were heuristically tuned. Because HSV is a non-Euclidean colorspace, the coneHSV colorspace [14] was adopted for use during color comparisons, where (H, S, V) values are transformed into $(V, S \cos(H), S \sin(H))$. The used colorspace components are shown in Fig.5. An ideal post filtering image will appear bright for the tissue pixels and significantly darker for surgical tool pixels.

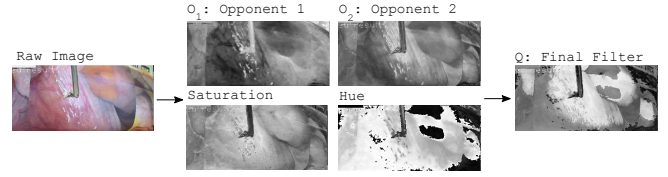


Fig. 5: Colorspace components used to determine color mask. The likelihood map Q is defined by a weighted sum of these components: hue, saturation, O_1 and O_2 .

D. Frequency Domain Shape Matching

The two masks, u and Q , provide two estimates of the surgical tool shape within the image frame. The robot kinematics shape prior mask, u , was derived by projecting surgical tool configurations and thickness onto the image plane. Meanwhile, the log-likelihood mask, Q , was generated via a linear combination of four colorspace components.

In mask u , pixels corresponding to surgical tools are white (255 in 8-bit gray), while the remaining pixels are black (0 in 8-bit gray). The converse is true for Q , that is the determined tool pixels approach 0 (black), and the rest of the image approaches 255 (white). Therefore, ideally the black pixels in u should correspond to white pixels in Q and vice versa. Multiplying ideal masks pixel-wise should result in all zeros.

However, u may not align well with Q , as shown by comparing Fig.6-b and Fig.6-c. This can be due to inaccuracies in camera extrinsic parameters, robot kinematics and joint sensors, or timing mismatches between robot pose and image frame. An objective function defined as the sum of pixel-wise multiplication between u and Q can be interpreted as the energy, E , to be minimized for optimal alignment. Suppose there are N_R rows and N_C columns in both images A, B . Then E is defined as

$$E(A, B) = \sum_{y=1}^{N_R} \sum_{x=1}^{N_C} A(x, y) B(x, y) \quad (2)$$

The mask matching procedure aims to modify shape prior mask u to best match Q and thus minimize E , and is achieved in two optimization steps:

- Translation — finding translated version of u to match Q , generating translated shape prior mask U .
- Rotation and scale — finding the optimal rotation and scaling of U to match Q , generating mask $\mathbb{U}$.

a) *Translation*: Suppose a translational error exists between shape prior u and actual surgical tool image location. To counteract this error, a translational offset which minimizes E is sought. Let $\vec{t} = (t_x, t_y)$, and then define $u_{\vec{t}}$ as the resultant mask of u translated by $\vec{t}$. The optimal translated mask is denoted U

$$U = \arg \min_{u_{\vec{t}}} E(u_{\vec{t}}, Q) \quad (3)$$

The solution to this optimization is achieved efficiently in the frequency domain using duality between spatial and frequency domains [15], namely

$$u \circledast Q = \mathbb{F}^{-1}(\mathbb{F}_Q \mathbb{F}_u^*) \quad (4)$$

where $\otimes$ denotes spatial convolution and $\mathbb{F}$ the Discrete Fourier Transform (DFT). Consider pixel (t_x, t_y) of the spatial convolution (origin is center of image):

$$u \otimes Q(t_x, t_y) = \sum_{y=1}^{N_R} \sum_{x=1}^{N_C} u(x - t_x, y - t_y) Q(x, y)$$

which is precisely $E(u_{\vec{t}}, Q)$. Thus the optimal offset is determined by the minimum pixel of $u \otimes Q = \mathbb{F}^{-1}(\mathbb{F}_Q \mathbb{F}_u^*)$. The time complexity reduces from $\mathcal{O}(N^4)$ to $\mathcal{O}(N^2 \log N)$ using the DFT. Figure 6 outlines the procedure.

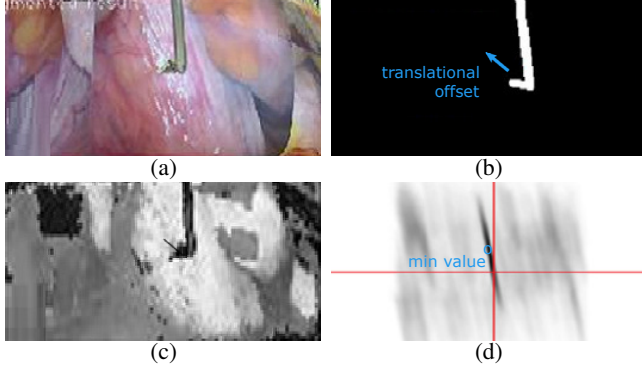


Fig. 6: Translation matching. (a) raw image frame (b) initial shape prior mask, u (c) color filtering mask, Q (d) two masks convolved, minimum value gives optimal translational offset to generate mask U .

Consider Fig.7f, where Q , u , U are respectively marked with red, blue, and green. The green is a translated version of blue that better matches red.

b) Rotation and Scale: The two masks U and Q may also misalign in rotation and scale. To account for this, the masks U and Q were first transformed to log-polar coordinates and zero-padded, forming U' and Q' . A Cartesian coordinate (x, y) is represented in log-polar coordinates as (a, b) where

$$\begin{aligned} a &= \log \sqrt{x^2 + y^2} \\ b &= \text{atan2}(y, x) \end{aligned}$$

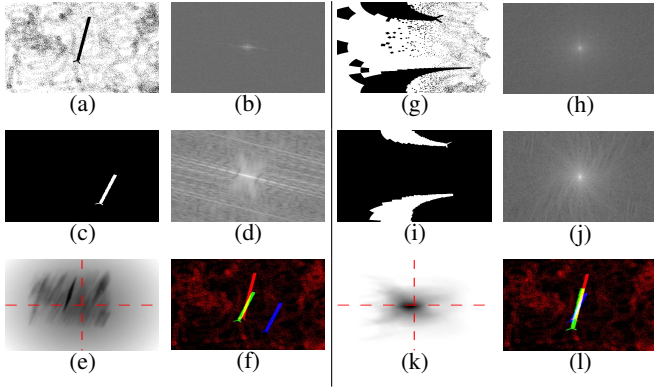


Fig. 7: DFT shape matching. (a)-(f) illustrate translation matching to find U , (g)-(l) rotation and scale matching to find U . (a) color filter mask Q (b) Fourier transform $\mathbb{F}_Q$ (c) shape prior mask u (d) Fourier transform $\mathbb{F}_u$ (e) convolution $u \otimes Q$ (f) red - Q , blue - u , green - U (g) log-polar color mask Q' (h) Fourier transform $\mathbb{F}_{Q'}$ (i) log-polar shape mask U' (j) Fourier transform $\mathbb{F}_{U'}$ (k) convolution $U' \otimes Q'$ (l) red - Q , blue - U , green - U .

a and b correspond to scale and rotation respectively. Finding the minimum pixel of $U' \otimes Q'$ thus determined the optimal scale and rotation of U to best match Q . The scaled and rotated version of U is denoted $\mathbb{U}$.

Consider Fig.7l, where Q , U , $\mathbb{U}$ are respectively marked with red, blue, and green. The green is a scaled and rotated version of blue that better matches red. Theoretically, due to nonlinear coupling of the two steps, the global optimum is achieved by interchangeably applying translation and rotation/scale adjustments until convergence, but only one iteration is applied in this work, under the assumption that the kinematics data is of high accuracy, to improve efficiency.

III. EXPERIMENTAL DESIGN

Fig.8 shows the experimental system setup using the Raven II platform. A 40mm baseline stereo camera with 640×480 pixel resolution was fixed to the Raven II base frame to acquire image data. Realistic tissue images were placed in the background.

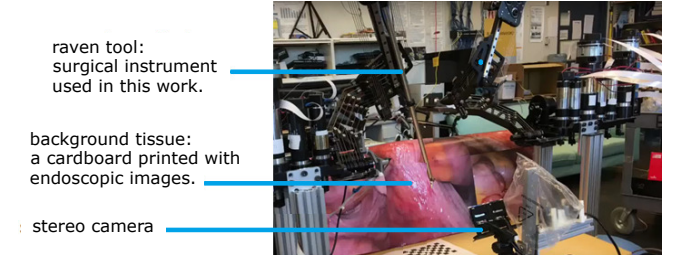


Fig. 8: Experimental setup, includes Raven II and stereo camera hardware.

A. Robot Kinematics Shape Prior

The joint locations of the Raven II platform were obtained from encoder readings and forward kinematics. The positions can be projected onto the camera image plane and shape can be determined via Eq.1. With raw position data, a static positioning error was observed. This was compensated with a static offset added to the initial robot pose estimate. This process is illustrated in Fig.9.

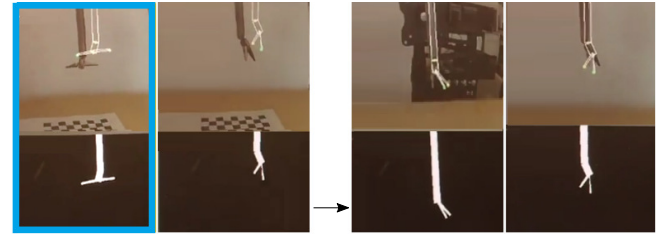


Fig. 9: The left shows raw projection of robot joints and initial shape prior of two poses without static offset. The right shows with static offset.

B. Frequency Domain Shape Matching

The main source of misalignment between color mask and shape prior mask arises from latency between image stream and robot kinematic information. That is, the robot pose is sampled slightly prior to the image frame with some variance. The methods described in Section II-D were used to determine optimal mask shift, scaling and rotation. In

practice, to avoid mismatching different tools in view, a 2D Gaussian distributed penalty map centered at the origin is fused with $u \otimes Q$ to bias the optimal translational solution towards smaller magnitude, under the assumption that initial shape prior u is close to true tool projection.,

C. Color Mask Post Processing

A final color mask was used to account for the effects of partial occlusion from real tissue and trivial ambiguity near segmentation boundaries. As illustrated in Fig.10, there are four steps to turn the nicely aligned shape prior $\mathbb{U}$ into the actual binary segmentation result.

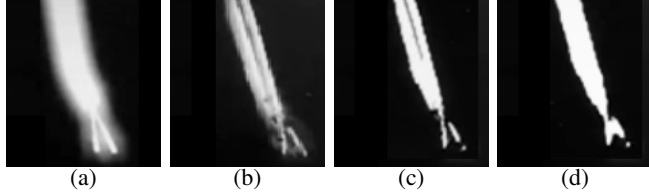


Fig. 10: Final color mask procedure. (a) dilated and blurred edges (b) log-likelihood color mask (c) binary threshold (d) morphological operations, resulting in final mask.

First, mask borders were expanded outward to tolerate trivial edge misalignment using OpenCV functions `dilate` and `blur` [12] (Fig.10a). Then, the same log-likelihood color filter for generating Q was applied. This helps to eliminate tool pixels partially occluded by real tissue (Fig.10b). Observe that this step incorrectly removed some tool pixels, due to the reflective nature of the tool. Next, a simple binary threshold classifies pixels as either tool or non-tool (Fig.10c). Finally, to account for the misclassified pixels due to the tool reflection, the OpenCV function `morphologyEX` was used to dilate boundaries and eliminate noise [12]. This results in the final segmentation mask as shown in Fig.10d.

IV. RESULTS AND DISCUSSION

A. Raven II Tool Segmentation

Fig.11 illustrates the final results of the real-time image-based surgical tool segmentation with robot kinematics shape prior. This was performed with the Raven II surgical robot platform, and the final mask was achieved using the techniques and workflow described in sections II and III. Overlaying the final mask with the raw image allows for segmentation of foreground (surgical tool tip) and background. The results shown here were achieved at a refresh rate of approximately 6 Hz using a commodity workstation.

From the robot kinematics and information about static pose estimation offset, an initial raw shape prior mask, u , is first generated. A log-likelihood color mask, Q , is created from raw image data. These two masks are then convolved (using duality property and DFT) to estimate optimal translation to match the shape prior to color mask, generating translated shape prior, U . Masks U, Q were then converted to log-polar coordinates, where they were again convolved to estimate optimal scale and rotation of the shape prior mask to match the color mask, generating mask $\mathbb{U}$. A post process color mask produces the final shape prior.

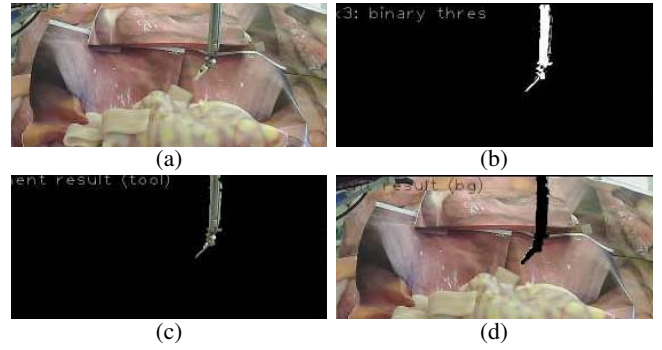


Fig. 11: Raven II tool segmentation. (a) raw image (b) final shape prior mask (c) segmented foreground tool (d) segmented background tissue.

B. Sørensen-Dice Index Analysis

The Sørensen-Dice index was used to measure the accuracy of the automatic real-time surgical tool segmentation. For evaluation, data was collected by actuating the Raven II tool along a trajectory traversing a wide variety of joint configurations while staying within the image frame. The maximum displacement and rotational speed of motion are 10cm/s and $30^\circ/s$ respectively, which meets standard surgical operation requirements [16]. Image frames (640×480 pixels) were captured and processed in real-time as described in sections II and III. 75 of the 2000 frames were randomly selected and manually labeled offline to classify tool from background. These manually labeled masks formed the ground truth X against which the real-time, automatically generated masks Y were evaluated.

The Sørensen-Dice index is a measure of similarity between two datasets, and is defined as

$$QS = \frac{2|X \cap Y|}{|X| + |Y|} \quad (5)$$

When datasets X, Y are identical, $QS = 1$, while disjoint X, Y result in $QS = 0$. For each of the 75 images, the ground truth dataset included pixel locations of the manually labeled surgical tool. The experimental set included segmented tool pixel locations generated by the proposed method. An average dice coefficient of 0.7372 is achieved, which compares well to state-of-the-art graphics-accelerated methods [17]. Sørensen-Dice indices for each individual frame over time are depicted in Fig.12. A broad selection

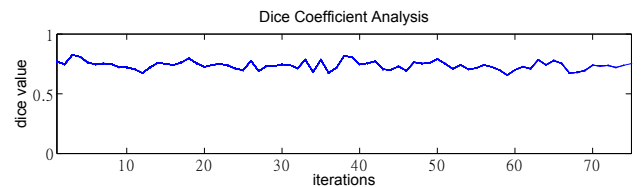


Fig. 12: Sørensen-Dice indices for 75 analyzed frames. Manually labeled mask pixels were compared to real-time generated mask pixels.

of tool poses were captured for this analysis, and Fig.13 demonstrates the very slight dependence that tool configuration bears on the Sørensen-Dice index. This suggests that the method is robust to varying tool configurations.

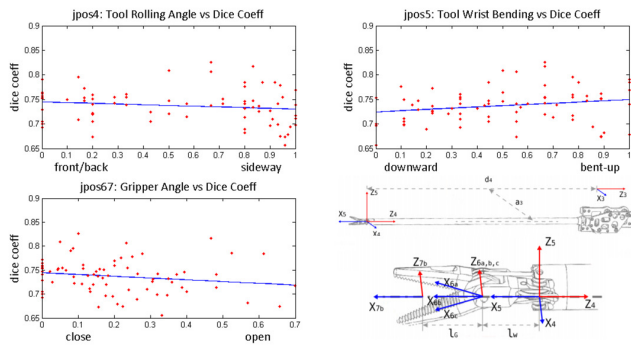


Fig. 13: The correlation between each tool joint and dice coefficient output.

C. Color Spectrum Stochastic Modeling

In this work, the log likelihood map Q was a weighted sum of the Opponent, RGB and HSV color components, and was essential for the shape matching of kinematics prior mask $\mathcal{U}$. While computationally efficient, a more discriminative color filtering scheme is possible through statistical analysis. To that end, a large number of surgical operation images were manually labeled and analyzed for color space components. From this, the probability distributions of tool pixels and non-tool pixels along each color space component can be generated. This statistical representation in color space, as shown in Fig.14, can provide the means for an advanced color filtering scheme.

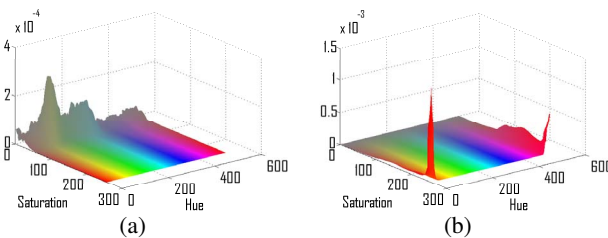


Fig. 14: Color statistics. (a) probability distribution of tool pixels (b) probability distribution of non-tool pixels.

V. CONCLUSION

This work presented a method for real-time vision-based surgical instrument segmentation with kinematic prior. The method affords a notable combination of attributes, including

- results validating use with Raven II tools.
- low computational complexity.
- 6Hz execution rate without GPU acceleration.
- average Sørensen-Dice index > 0.73 .
- robustness to partial occlusion by fusing robot kinematics with color filtering.

The technique was evaluated on the Raven II surgical platform, and segmentation results were compared with manually segmented images. The results were encouraging with high Sørensen-Dice index that is robust to tool configuration. Thus, the method is promising towards the use of kinematic prior and color masking for real-time tool segmentation in a robot-assisted minimally invasive surgical setting.

Future improvements to the proposed image segmentation method include stochastic modeling of surgical tool and tissue pixels, as described in section IV-C. This can greatly improve the generation of log likelihood color mask Q . Furthermore, static offset correction of estimated robot pose can be automated through Kalman filtering. Validating the method in various lighting conditions and with a reduced baseline stereo camera (or endoscopic) setup will further promote this method towards clinical issue. A natural extension of this work includes exploring the remaining subtasks of the vision-based force estimation as illustrated in Fig.2 while integrating the segmentation method described here.

REFERENCES

- [1] A. Reiter, P. K. Allen, and T. Zhao, "Feature classification for tracking articulated surgical tools," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer, 2012, pp. 592–600.
- [2] A. Reiter, P. K. Allen, and T. Zhao, "Marker-less articulated surgical tool detection," in *Computer assisted radiology and surgery*, 2012.
- [3] D. Burschka, J. J. Corso, M. Dewan, W. Lau, M. Li, H. Lin, P. Marayong, N. Ramey, G. D. Hager, B. Hoffman *et al.*, "Navigating inner space: 3-d assistance for minimally invasive surgery," *Robotics and Autonomous Systems*, vol. 52, no. 1, pp. 5–26, 2005.
- [4] L. Cheng, "Computation and measurement of force and tissue damage for the grasper-tissue interface in robot-assisted minimal invasive surgery," Ph.D. dissertation, 2015.
- [5] D. Stoyanov and G. Z. Yang, "Removing specular reflection components for robotic assisted laparoscopic surgery," in *Image Processing, 2005. ICIP 2005. IEEE International Conference on*, vol. 3. IEEE, 2005, pp. III–632.
- [6] D. Bouget, M. Allan, D. Stoyanov, and P. Jannin, "Vision-based and marker-less surgical tool detection and tracking: a review of the literature," *Medical image analysis*, vol. 35, pp. 633–654, 2017.
- [7] E. Costanza and J. Robinson, "A region adjacency tree approach to the detection and design of fiducials," *Video, Vision and Graphics*, 2003.
- [8] S. Parlak, I. Marsic, and R. S. Burd, "Activity recognition for emergency care using rfid," in *Proceedings of the 6th International Conference on Body Area Networks*. ICST (Institute for Computer Sciences, Social-Informatics and Telecommunications Engineering), 2011, pp. 40–46.
- [9] M. Chmarra, C. Grimbergen, and J. Dankelman, "Systems for tracking minimally invasive surgical instruments," *Minimally Invasive Therapy & Allied Technologies*, vol. 16, no. 6, pp. 328–340, 2007.
- [10] M. P. Fried, J. Kleefield, H. Gopal, E. Reardon, B. T. Ho, and F. A. Kuhn, "Image-guided endoscopic surgery: Results of accuracy and performance in a multicenter clinical study using an electromagnetic tracking system," *Laryngoscope*, vol. 107, no. 5, pp. 594–601, 1997.
- [11] B. Hannaford, J. Rosen, D. W. Friedman, H. King, P. Roan, L. Cheng, D. Glozman, J. Ma, S. N. Kosari, and L. White, "Raven-II: an open platform for surgical robotics research," *IEEE Transactions on Biomedical Engineering*, vol. 60, no. 4, pp. 954–959, 2013.
- [12] G. Bradski, *Dr Dobb's Journal of Software Tools*. M & T Pub., 2000.
- [13] K. E. van de Sande, T. Gevers, and C. G. Snoek, "Color descriptors for object category recognition," in *Conference on Colour in Graphics, Imaging, and Vision*, vol. 2008, no. 1. Society for Imaging Science and Technology, 2008, pp. 378–381.
- [14] Z. Pezzementi, S. Voros, and G. D. Hager, "Articulated object tracking by rendering consistent appearance parts," in *Robotics and Automation, 2009. ICRA'09. IEEE International Conference on*. IEEE, 2009, pp. 3940–3947.
- [15] S. Chen, D. Cremers, and R. J. Radke, "Image segmentation with one shape priors template-based formulation," *Image and Vision Computing*, vol. 30, no. 12, pp. 1032–1042, 2012.
- [16] P. Woerdeman, P. Willems, H. Noordmans, and J. van der Sprenkel, "The analysis of intraoperative neurosurgical instrument movement using a navigation log-file," *International Journal of Medical Robotics and Computer Assisted Surgery*, vol. 2, no. 2, pp. 139–145, 2006.
- [17] I. Laina, N. Rieke, C. Rupprecht, J. P. Vizcaíno, A. Eslami, F. Tombari, and N. Navab, "Concurrent segmentation and localization for tracking of surgical instruments," *CoRR*, vol. abs/1703.10701, 2017.