2021 43rd Annual International Conference of the
IEEE Engineering in Medicine & Biology Society (EMBC)
Oct 31 - Nov 4, 2021. Virtual Conference

A Novel Perception Framework for Automatic Laparoscope Zoom Factor Control Using Tool Geometry

Yan-Jun Yang, Arvind N Vadivelu, Charles H. C. Pilgrim,
Dana Kulić, Member, IEEE and Elahe Abdi, Member, IEEE,

Abstract— **In conventional Minimally Invasive Surgery, the surgeon conducts the operation while a human or robot holds the laparoscope. Laparoscope control is returned to the surgeon in teleoperated camera holding robots, but simultaneously controlling the laparoscope and surgical tools might be cognitively demanding. On the other hand, fully automated camera holders are still limited in their performance. To help the surgeon to better focus on the main operation while maintaining their control authority, we propose an automatic laparoscope zoom factor control framework for Robot-Assisted Minimally Invasive Surgery. In this paper, we present the perception section of the framework. It extracts and uses the surgical tool's geometric characteristics to adjust the laparoscope's zoom factor, without any artificial markers. The acceptable range and tooltip's position frequency during operations are analysed based on the gallbladder removal surgery dataset (Cholec80). The common range and tooltip's heatmap are identified and presented quantitatively.**

I. INTRODUCTION

In the past decades, innovations in surgical robotics have provided surgeons with improved efficiency, safety, and convenience compared to traditional surgery, where surgeons cooperate with human assistants [1]. In laparoscopic surgery, using a camera holding robot to replace the human camera holder gives surgeons direct control of the laparoscope and reduces suboptimal communications with assistants [2]. Manual camera holding robots are usually controlled via voice interface [3], head/gaze motion interface [4], foot interface [5], or joystick [6]. However, the additional laparoscope adjustment task increases the surgeon's cognitive load and is distracting. Therefore, upgrading the system from a fully teleoperated robotic arm to a system that can complete certain tasks or sub-tasks autonomously may help to reduce the surgeon's cognitive load [7]. Automation can include in-plane movement, insertion/zoom adjustment [8], [9], and complete automation of the laparoscope [10], [11], [12], [13], [14].

This research has been supported by MIME Seed Funding, the Monash Institute of Medical Engineering and Monash Partners Academic Health Science Centre

Y.-J. Yang, E. Abdi and D. Kulić are with the Department of Mechanical and Aerospace Engineering, Monash University, Clayton, VIC 3800, Australia (e-mail: Yanjun.Yang@monash.edu; Elahe.Abdi@monash.edu; Dana.Kulic@monash.edu

Charles H. C. Pilgrim is with Suite 29 Cabrini Medical Centre, Melbourne, Victoria, Australia, The Alfred Hospital, Commercial Rd, Melbourne Victoria and Central Clinical School, Faculty of Medicine, Monash University. (email: charlespilgrim@hpbsurgery.com)

Arvind N Vadivelu is with the Department of Mechanical Engineering, University of Melbourne, Parkville, VIC 3010, Australia (e-mail: arvindn@student.unimelb.edu.au

A stable and optimal laparoscopic view during minimally invasive surgery (MIS) that meets the surgeon's expectation and preference is an important factor in improving the surgeon's performance [15]. A screenshot of the optimal laparoscope view is shown in Figure 1. A typical control strategy to achieve an optimal view is to i) maintain a certain size ratio between the tool and the background view (hereby shortened as tool-view ratio) as a trade-off between perspective and details [8], [10], [11], [12], [13] and ii) keep the tooltip near the center of the view [10], [16].

A fully automated system would significantly reduce the surgeon cognitive load, but comes with the potential hazard introduced by the increased complexity of the novel system and it would be less capable of dealing with complex or unanticipated conditions. In our research, we only focuses on the zoom ratio control as it give the surgeon physical laparoscope control authority, but also shared their partial workload. Tool-view ratio can be adjusted by changing the distance between the laparoscope and the tooltip or tuning the lens zoom factor. Acquiring information about the distance between the laparoscope and the main tool from a monocular lens is challenging. To estimate the camera-tool depth, either sensors and markers [12], [8], [13], color band on the tool's shaft [10], [11], or neural networks [17], [14] are used. Nishikawa *et al* [8] presented the commonly used distances between the tooltip and the laparoscope of two camera holding assistants in simulated surgery tasks using optical markers. Rivas-Blanco *et al* [13] gave recommended zoom factors for different stages in the simulation with no further explanation, using the robotic system's stored position data. However, using additional accessories could be problematic in real surgery considering sterilization and equipment setup. On the other hand, neural networks-based methods can overcome drawbacks of the accessories-dependent approaches.

Always keeping the tool at the central region can cause frequent laparoscope position changes when the tool is slightly off the center, thus making surgeons feel sick. Therefore, using a wider central optimal region defined by researchers [11] or based on collected data [17] to balance the motion and view quality is adopted. Li *et al* [17] shows a qualitative frequency heatmap of the optimal region in their paper.

To the best of the authors' knowledge, the common tool-view size ratio (distance or zoom factor) and heatmap in real surgeries have not yet been quantitatively identified.

Based on the literature review, the tooltip position with respect to the laparoscope and the size ratio between the

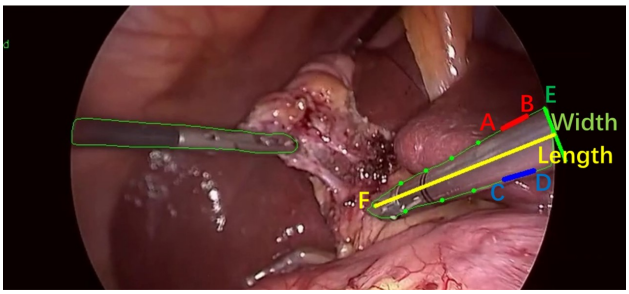

Fig. 1. Intraoperative photo [18] demonstrating target anatomy for cholecystectomy (gallbladder, cystic duct, cystic artery), and relevant adjacent organs for orientation (liver, stomach). Tools' mask contour is drawn on the laparoscopic image using the green line, length and width are annotated as well. The thick blue and red lines connect feature points.

tool and the background view are the two most critical metrics to define an optimal view as they determine whether a surgeon sees sufficient details and reduce unnecessary lens movements. This research aims to develop a new framework for automatic adjustment of the lens zoom factor according to the tool-view size ratio (Figure 2) to work with any manual interface, e.g., [19], [20]. The framework consists of a perception and a control section, the perception section uses the laparoscope image to estimate the tool's geometric features, which are then used as the input for the zoom factor controller. At least four Degrees of Freedom (DoF): Roll, Yaw, Pitch and Insertion/Withdrawal are required to operate the laparoscope with the Remote Center of Motion constraint. The zoom factor adjustment replaces the control of insertion/withdrawal to achieve the optimal tool-view size ratio.

In this paper, we describe and validate the perception components of the framework. The proposed approach is tested on the clipping and cutting stage of the Cholec80 dataset that contains videos of 80 cholecystectomy surgeries performed by 13 surgeons [18]. Cholecystectomy is the gallbladder removal surgery, a relatively common and straightforward procedure [21]. It's main phases have been well-defined [18], with the clipping and cutting stage identified as the most delicate phase in the procedure [18]. Frequent zoom adjustments make this stage suitable to study the human camera holder's preferable zoom operation pattern. The contributions of this paper are:

- A perception algorithm that uses a neural network to extract the tool's geometric features to estimate the tool-view size ratio for the zoom factor control without using a color marker or sensor in contrast to [10], [11].
- The clipper and scissors' common tool-view size ratio and central optimal region in the clipping and cutting phase of cholecystectomy [18] are quantitatively identified based on the Cholec80 dataset using the developed perception block. This provides an initial insight into the optimal zoom factor and central optimal region based on real surgery data.
- A waiting zone is defined as the supplementary of the optimal central region. When the tooltip is inside this region, the laparoscope's zoom factor is not adjusted.

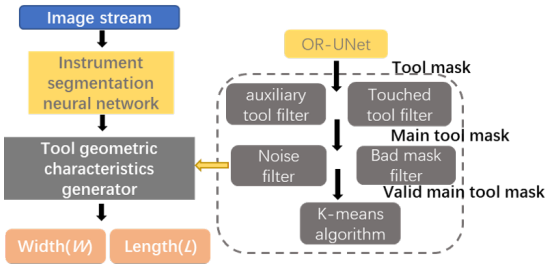

Fig. 2. The flowchart of the proposed laparoscope control framework - perception component.

II. METHODS

Our framework aims to develop a control system that only manipulates the lens zoom factor. Thus, the tool-view ratio, which is the representation of the zoom factor adjustment, is the concern of our framework, rather than the depth and absolute 3D position for fully automatic laparoscope control [17].

In the perception section of the framework, tools such as clipper and scissors in cholecystectomy are defined as the main tool and tools like grasper that assist the operation are defined as auxiliary tools. The proposed perception algorithm uses the main tool's binary mask to estimate the tooltip position and geometric features. The binary masks are produced by an instrument segmentation neural network, while the tool geometric characteristics generator extracts the correct main tool mask's width (W) and length (L) (Figure 1). The width is defined as the distance between the intersecting point of the tool body's longer side and the laparoscopic view's edge (point E), and its symmetric point across the tool's longitudinal axis. The length is defined as the distance from the tooltip to the width line. In the proposed approach, the tool-view size ratio is characterized by $\frac{L}{W^2}$, which is used by the controller to adjust the zoom factor. This ratio is chosen because it exhibits a monotonically decreasing relationship with the zooming operation, i.e., it becomes smaller when zooming in and larger when zooming out.

The relationship between the tool-view size ratio and zoom factor is verified in simulations. Trocar placement for a laparoscopic cholecystectomy (Figure 3(a)) was followed for setting up the simulation experiment. As shown in Figure 3(b), the world frame and camera frame coincide at the lens tip and the z axis is along the lens shaft pointing out. Since the main tool is usually inserted from the right side, the range for the distance a is between [0cm, 20cm], while the allowed moving range for the distance b is between [-5cm, 5cm]. The vertical distance range for c is between [0cm, 30cm] based on the results in [8]. The tool first rotates γ degrees about the tool's y axis (shares the same orientation with the world frame originally) with respect to the insertion point and then rotates θ degrees about the tool's rotated x axis with respect to the insertion point to reach the final target. The range of

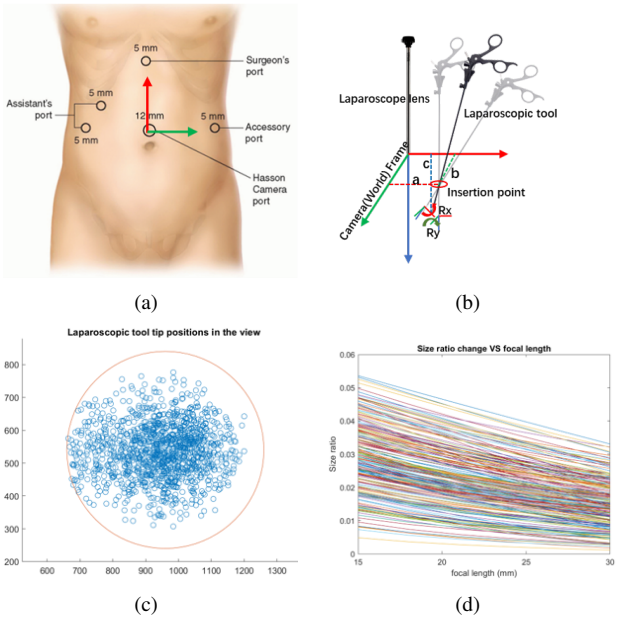

Fig. 3. (a) Trocar placement for the laparoscopic cholecystectomy [22]. (b) Simulation setup paradigm. (c) Tooltip position in the view. (d) Size ratio versus focal length.

γ is between [0, 90] and the range of θ is between [-90, 90]. Evaluated tip positions are shown in Figure 3(c) and almost cover the whole view. The laparoscopic tool used in the simulation has a 5 mm diameter. The selected laparoscopic camera has a 1920*1080 pixel output signal, 300mm radius laparoscopic view and focal length $f = 15 - 31mm(2\times)$. The simulation randomly tested 20000 valid tool insertion configurations (the tooltip is inside the view) and all size ratios become smaller when zooming in and larger when zooming out (Figure 3(d))

An optimal range of the tool-view size ratio is defined as $\frac{L}{W^2}_{Mean} \pm \frac{L}{W^2}_{SD}$ based on collected clinical data. $\frac{L}{W^2}_{Mean}$ gives the base value and $\frac{L}{W^2}_{SD}$ is the acceptable changing range of the size ratio during operations. When the value of $\frac{L}{W^2}$ is in this range, the zoom factor will be kept constant.

The use of $\frac{L}{W^2}$ considers both the tool's size and shape, and clearly identifies the insertion and operation process (Figure 4). Only using the size data causes ambiguity as a slender mask across the view and a larger mask near the edge may have similar areas. Only relying on shape ($\frac{L}{W}$) makes identifying masks difficult, since tools can have similar shape but different sizes. Therefore, the zoom control could be unexpectedly activated if $\frac{L}{W}$ or area is deployed.

A. Instrument segmentation neural network

The segmentation neural network duplicates the OR-UNet [23], which achieved excellent performance in the Robust-MIS 2019 challenge [24]. The model was trained by following the default settings from [23], but using 3-fold cross-validation to achieve real time processing. An ensemble that consists of the three trained models from the cross-validation is used for inference. TensorRT [25] is used to speed up the inference time to 20 fps processing rate on a local PC with Nvidia 2070super GPU and Intel(R) Core i7-10875H CPU.

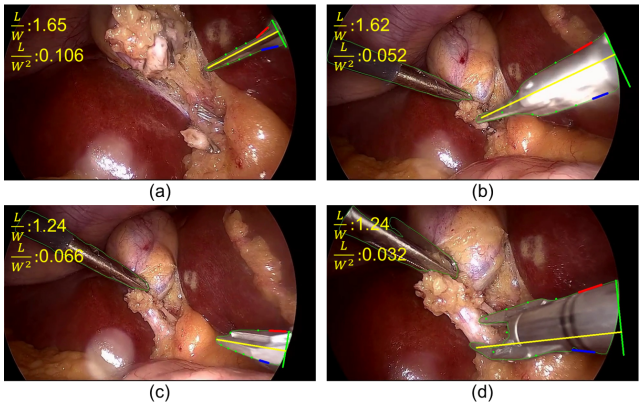

Fig. 4. Screenshots of four operation frames. (a) is a scissors just inserted into the view. (b) is a scissors close to the central region. (c) is a clipper just inserted into the view. (d) is a clipper close to the central region.

B. Tool geometric characteristics generator

The characteristics generator is designed to extract the width and length from a valid mask of the main tool. The touched tools filter (Figure 2) is implemented to detect and handle contact between the auxiliary and main tools. The potential main tool mask will then be checked by the noise filter and the bad-mask filter to make sure the neural network correctly segments the tool.

Finally, the K-means algorithm is used to classify points on the tool contour to get 12 feature points (green points in Figure 1). The distance from the farthest feature point (point F) to the width line will be identified as the tool length. Two sets of two points (points A,B and points C,D) close to the edge are connected to describe the main tool's shape, and the longitudinal symmetric axis is the bisector of the angle formed by line segment AB and CD. The width is calculated by using the intersection point from the longer tool side (point E) and the symmetric axis.

III. RESULTS

The Cholec80 dataset is analysed to collect the tool-view size ratio and tooltip heat map. The optimal size ratio and the central optimal region are inferred using statistical analysis. 13 videos from the dataset are eliminated due to bad segmentation performance. An operation is defined as the process starting immediately after the first zoom operation until just before the tool starts withdrawing from the operating site. In total 121 clipping operations (including 11 invalid data) and 95 cutting operations (including 4 invalid data) were collected.

In the Cholec 80 dataset, the laparoscopic view size varies. Therefore, the width and length are normalised with the curved edge arc length and the circular view's diameter in pixel units respectively for normalised size ratio $\frac{L}{W^2}$ analysis. The tip position coordinates are normalised with the view's radius.

The Cholec80 dataset does not identify the surgeon (i.e., we do not know if the same surgeon performed one or more of the operations), while patients are different in each case. Although some analysed operations are from the same video, we assume all collected operations are independent.

A. Framework performance

The instrument segmentation neural network mean Dice score is 82.06 with 3-fold cross validation on the training set (mean Dice score is 87.41 with 8-fold cross validation in [23]). The tool geometric characteristics generator performance depends heavily on segmented masks. The accuracy of successful identification of length and width is 98.1% tested with 2000 correctly segmented masks.

B. Optimal tool-view size ratio

To estimate the tool-view size ratio's base value and changing range, standard error (SE) is selected to calculate the margin of error. The 95% confidence interval for the base value of $\frac{L}{W^2}$ for the clipper and scissors are in the [0.057, 0.071] and [0.077, 0.099] ranges. The changing range of $\frac{L}{W^2}$ for the clipper and scissors are in [0.0197, 0.0243] and [0.0245, 0.0315] ranges. Histograms of the size ratio's base value and changing range in each operation are shown in Figure 5.

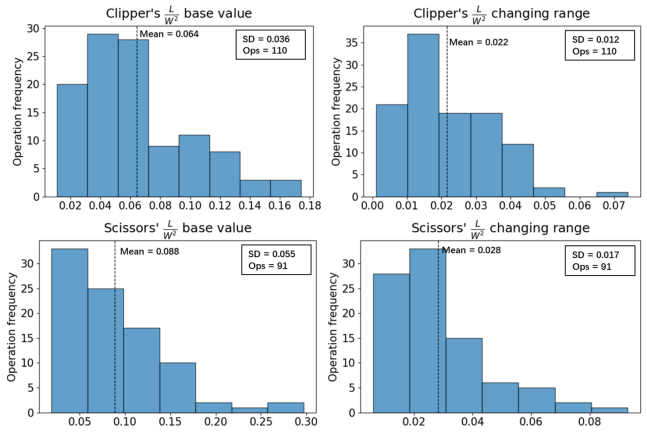

Fig. 5. Histograms of clipper and scissors' $\frac{L}{W^2}$ base value and changing range.

C. Optimal central region

The heatmaps are generated by using the scissors and clipper tip coordinates collected by the perception section. As can be seen from Figure 6, the most activated areas for the clipper and scissors are slightly deviated to the right and upper side of the circle origin, respectively. The most activated region (defined as the area where the probability is higher than 0.8) and the secondary most activated region (defined as the area where the probability is higher than 0.5) of the scissors can be enclosed by two circles located at the center of the view with r_{inside} = 0.23 and $r_{outside}$ = 0.38. The corresponding values for the clipper are r_{inside} = 0.26 and $r_{outside}$ = 0.4. The heatmap can be used as a reference to control the laparoscope's position. Absence of the main tool from the activated region might express the surgeon's intention to change the operating site.

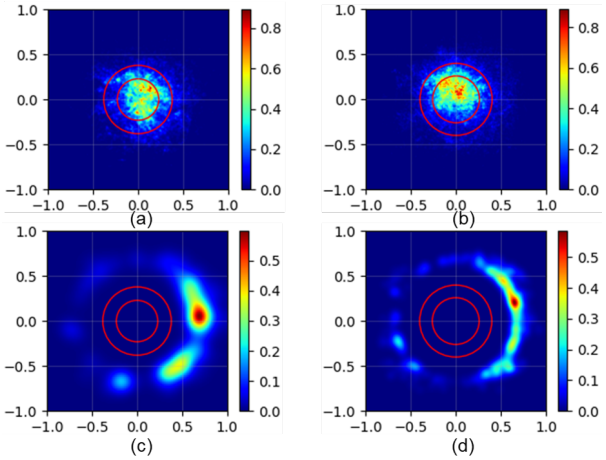

Fig. 6. Generated heatmaps in the clipping and cutting stage (a) is the heatmap of the clipper in the central area. (b) is the heatmap of the scissors in the central area. (c) is the heatmap of the clipper in the waiting zone. (d) is the heatmap of the scissors in the waiting zone.

IV. DISCUSSION & CONCLUSION

This paper presents the perception section of our proposed framework for automatic laparoscope zoom ratio control. The perception part makes use of the symmetrical characteristics of the tool's shape to calculate the width and length. We also present the clipper and scissors' commonly used tool-view size ratio range in the clipping and cutting stage of cholecystectomy, and the recommended optimal central region size for these two tools.

An interesting observation from the data is that except for the central region, surgeons tend to keep the tool at the very edge but not completely withdrawn from the view when they pause and check the target site. After removing all points near the secondary most activated region, the waiting zones for clipper and scissor are shown in Figure 6. The waiting zone should be considered when designing the optimal range, to prevent false actuation while the surgeon checks the target and then resumes the operation.

One challenge in using the collected common tool-view size ratio and central region for laparoscope control is to build the connection between the common value and individual's preference. One surgeon may prefer a wider view while another surgeon may prefer a more detailed view. The difference in $\frac{L}{W^2}$ base value and changing range reflects the individual's preference. Therefore, in the control section development, the common range of $\frac{L}{W^2}$ will first be tested to estimate it's generality for surgeons, but could also be adjusted to adapt to personal preference. This proposed framework has the potential to be adapted to other types of surgeries as long as the perception section could collect enough data to define common moving range and optimal region for all the main tools.

The next step is to analyse the remaining phases of the cholecystectomy. As the current method uses a binary

segmentation neural network, tracking the main tool in real-time is not possible if the main tool is re-inserted from a new position accompanied with other auxiliary tools. In the future, the binary network will be replaced with multi-object segmentation to achieve real-time main tool identification and tracking. Therefore, the perception algorithm will only process the main tool and ignore the other tools. If the main tool is invisible in the laparoscopic view, the zoom factor will hold still, and once the non-detected time becomes longer than a threshold, the system will zoom out to try to search for the main tool. Finally, the whole framework including the control section will be completed, and we will test the real-time performance of the proposed automation system and its influence on the surgeon's workload.

REFERENCES

[1] S.-J. Kim and S. C. Lee, "Technical and instrumental prerequisites for single-port laparoscopic solo surgery: state of art," *World Journal of Gastroenterology: WJG*, vol. 21, no. 15, p. 4440, 2015.

[2] J. E. Jaspers, P. Breedveld, J. L. Herder, and C. A. Grimbergen, "Camera and instrument holders and their clinical value in minimally invasive surgery," *Surgical Laparoscopy Endoscopy & Percutaneous Techniques*, vol. 14, no. 3, pp. 145–152, 2004.

[3] L. R. Kavoussi, R. G. Moore, J. B. Adams, and A. W. Partin, "Comparison of robotic versus human laparoscopic camera control," *The Journal of urology*, vol. 154, no. 6, pp. 2134–2136, 1995.

[4] J.-U. Stolzenburg, T. Franz, P. Kallidonis, D. Minh, A. Dietel, J. Hicks, M. Nicolaus, A. Al-Aown, and E. Liatsikos, "Comparison of the free-hand® robotic camera holder with human assistants during endoscopic extraperitoneal radical prostatectomy," *BJU international*, vol. 107, no. 6, pp. 970–974, 2011.

[5] S. Voros, G.-P. Haber, J.-F. Menudet, J.-A. Long, and P. Cinquin, "Viky robotic scope holder: Initial clinical experience and preliminary results using instrument tracking," *IEEE/ASME transactions on mechatronics*, vol. 15, no. 6, pp. 879–886, 2010.

[6] Z. Rossini, A. Cardia, D. Milani, G. B. Lasio, M. Fornari, and V. D'Angelo, "Vitom 3d: preliminary experience in cranial surgery," *World neurosurgery*, vol. 107, pp. 663–668, 2017.

[7] T. Haidegger, "Autonomy for surgical robots: Concepts and paradigms," *IEEE Transactions on Medical Robotics and Bionics*, vol. 1, no. 2, pp. 65–76, 2019.

[8] A. Nishikawa, H. Nakagoe, K. Taniguchi, Y. Yamada, M. Sekimoto, S. Takiguchi, M. Monden, and F. Miyazaki, "How does the camera assistant decide the zooming ratio of laparoscopic images? analysis and implementation," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pp. 611–618, Springer, 2008.

[9] Y. Sun, B. Pan, Y. Fu, and G. Niu, "Visual-based autonomous field of view control of laparoscope with safety-rcm constraints for semi-autonomous surgery," *The International Journal of Medical Robotics and Computer Assisted Surgery*, vol. 16, no. 2, p. e2079, 2020.

[10] B. W. King, L. A. Reisner, A. K. Pandya, A. M. Composto, R. D. Ellis, and M. D. Klein, "Towards an autonomous robot for camera control during laparoscopic surgery," *Journal of laparoendoscopic & advanced surgical techniques*, vol. 23, no. 12, pp. 1027–1030, 2013.

[11] B. Yang, W. Chen, Z. Wang, Y. Lu, J. Mao, H. Wang, and Y.-H. Liu, "Adaptive fov control of laparoscopes with programmable composed constraints," *IEEE Transactions on Medical Robotics and Bionics*, vol. 1, no. 4, pp. 206–217, 2019.

[12] I. Rivas-Blanco, C. J. Perez-del Pulgar, C. López-Casado, E. Bauzano, and V. F. Muñoz, "Transferring know-how for an autonomous camera robotic assistant," *Electronics*, vol. 8, no. 2, p. 224, 2019.

[13] I. Rivas-Blanco, B. Estebanez, M. Cuevas-Rodriguez, E. Bauzano, and V. F. Muñoz, "Towards a cognitive camera robotic assistant," in *5th IEEE RAS/EMBS International Conference on Biomedical Robotics and Biomechatronics*, pp. 739–744, IEEE, 2014.

[14] M. Wagner, A. Bihlmaier, H. G. Kenngott, P. Mietkowski, P. M. Scheikl, S. Bodenstedt, A. Schiepe-Tiska, J. Vetter, F. Nickel, S. Speidel, *et al.*, "A learning robot for cognitive camera control in minimally invasive surgery," *Surgical Endoscopy*, pp. 1–10, 2021.

[15] H. Mohrmann-Lendla and A. G. Fleischer, "The effect of a moving background on aimed hand movements," *Ergonomics*, vol. 34, no. 3, pp. 353–364, 1991.

[16] A. Nishikawa, K. Ito, H. Nakagoe, K. Taniguchi, M. Sekimoto, S. Takiguchi, Y. Seki, M. Yasui, K. Okada, M. Monden, *et al.*, "Automatic positioning of a laparoscope by preoperative workspace planning and intraoperative 3d instrument tracking," in *MICCAI2006 Workshop proceedings*, pp. 82–91, Citeseer, 2006.

[17] B. Li, B. Lu, Y. Lu, Q. Dou, and Y.-H. Liu, "Data-driven holistic framework for automated laparoscope optimal view control with learning-based depth perception," *arXiv preprint arXiv:2011.11241*, 2020. [Online; accessed 22-December-2020].

[18] A. P. Twinanda, S. Shehata, D. Mutter, J. Marescaux, M. De Mathelin, and N. Padoy, "Endonet: a deep architecture for recognition tasks on laparoscopic videos," *IEEE transactions on medical imaging*, vol. 36, no. 1, pp. 86–97, 2016.

[19] Y.-J. Yang, S. Udatha, D. Kulić, and E. Abdi, "A novel foot interface versus voice for controlling a robotic endoscope holder," in *2020 8th IEEE RAS/EMBS International Conference for Biomedical Robotics and Biomechatronics (BioRob)*, pp. 272–279, IEEE, 2020.

[20] S.-Z. Ye, P. Jain, A. Walley, Y.-J. Yang, and E. Abdi, "A novel four-degree-of-freedom versus a conventional foot interface for controlling a robotic assistive arm," in *2020 8th IEEE RAS/EMBS International Conference for Biomedical Robotics and Biomechatronics (BioRob)*, pp. 1080–1087, IEEE, 2020.

[21] A. Bihlmaier and H. Woern, "Automated endoscopic camera guidance: A knowledge-based system towards robot assisted surgery," in *ISR/Robotik 2014; 41st International Symposium on Robotics*, pp. 1–6, VDE, 2014.

[22] W. L. Hospital, "Port position in laparoscopic general surgery." https://www.laparoscopyhospital.com/drrkmishra/portpositionslide.php. [Online; accessed 22-December-2020].

[23] F. Isensee and K. H. Maier-Hein, "Or-unet: an optimized robust residual u-net for instrument segmentation in endoscopic images," *arXiv preprint arXiv:2004.12668*, 2020.

[24] T. Roß, A. Reinke, P. M. Full, M. Wagner, H. Kenngott, M. Apitz, H. Hempe, D. Mindroc-Filimon, P. Scholz, T. N. Tran, *et al.*, "Comparative validation of multi-instance instrument segmentation in endoscopy: results of the robust-mis 2019 challenge," *Medical image analysis*, vol. 70, p. 101920, 2021.

[25] N. DEVELOPER, "Nvidia tensorrt programmable inference accelerator." https://developer.nvidia.com/tensorrt. [Online; accessed 22-December-2020].