A virtual suturing task: proof of concept for awareness in autonomous camera motion

Nicolò Pasini[1], Andrea Mariani[2], Adnan Munawar[3], Elena De Momi[1] and Peter Kazanzides[3]

***Abstract*— Robot-assisted Minimally Invasive Surgery (MIS) requires the surgeon to alternatively control both the surgical instruments and the endoscopic camera, or to leave this burden to an assistant. This increases the cognitive load and interrupts the workflow of the operation. Camera motion automation has been examined in the literature to mitigate these aspects, but still lacks situation awareness, a key factor for camera navigation enhancement. This paper presents the development of a phase-specific camera motion automation, implemented in Virtual Reality (VR) during a suturing task. A user study involving 10 users was carried out using the master console of the da Vinci Research Kit. Each subject performed the suturing task undergoing both the proposed autonomous camera motion and the traditional manual camera control. Results show that the proposed system can reduce operational time, decreasing both the user's mental and physical demand. Situational awareness is shown to be fundamental in exploiting the benefits introduced by camera motion automation.**

I. INTRODUCTION

A. Research field

Robot-assisted Minimally Invasive Surgery (MIS) is confirming itself as one of the major technological improvements in the surgical scenario over the past two decades [1]. A common characteristic to every surgical robot is the ability to enhance the surgeon's capability to treat patients, by adding features able to improve the surgical outcome. Many different categories of medical robots are available on the market, such as the da Vinci Surgical System, dVSS, (Intuitive Surgical, Sunnyvale, CA, USA). The da Vinci robot is one of the most widely used robot-assisted MIS platforms belonging to the surgeon extender category [2]. The reason behind their success lies on several advantages introduced by robotic assistance: reduced operational and hospitalization time, reduced scars and necessity of further surgical operations for the patient, while motion scaling, tremor filtering and immersive vision extend the surgeons' skills. This last aspect attracts particular interest. Visualization modalities have drastically changed since the introduction of endoscopes with MIS.

Access to soft tissues in both traditional and robot-assisted MIS is permitted by incisions on the patient's skin, allowing surgical tools and camera to be inserted. However, significant differences are introduced with robotic assistance, such as loss of direct control of both tools and camera. In fact, with a traditional MIS approach the presence of an assistant is usually required to perform camera navigation, easing surgical workflow. Human assistance for camera control is not required anymore with the introduction of robotic platforms, placed in between the surgeon and patient. Surgeons are expected to take control over both camera and surgical tools. Specifically tailored consoles are deployed with such systems to enable teleoperation with multiple robotic arms in order to overcome limits deriving from this asynchronous control modality. Devices such as the dVSS are equipped with a pair of Master Tool Manipulators (MTMs) to control robotic arms and a foot pedal tray to allow a quick switch between teleoperation of tools and camera. Surgeons may settle for a suboptimal Field of View (FoV) or allow tools to fall out of view, due to the effort involved for camera repositioning, which can result in soft tissue injuries or surgical inaccuracy [3]. One of the still unanswered questions that researchers tried to address since teleoperated surgical systems started taking over the market is how to reduce the surgeon's mental and physical workload [4], [5] resulting from asynchronous, hence unnatural, tools and camera control. A promising solution, yet not so simple to obtain, is the automation of processes such as camera navigation. At present, automation in surgical robotics does not exist in clinical practice, but extensive studies have been performed to find suitable solutions for camera motion related issues.

B. Related works

Such limitations introduced by new control dynamics motivated the development of specific platforms to analyze camera navigation for MIS. During such studies, the operators are asked to properly center the FoV, maintain a horizon suitable for the performed task, correctly size the range of view and hold a steady image, while specific metrics are recorded.

Virtual Reality (VR) simulators are currently proving themselves as a valid option to train and test surgical skills, not mentioning the related cost effectiveness, ease of deployment and high availability [6], [7]. The high versatility of VR enables also a quicker and easier evaluation of specific metrics that would be otherwise difficult to acquire, or be less precise, in a dry lab scenario.

These camera motion related skills apply to robot-assisted MIS as well; as a result, multiple Human-Machine Interfaces (HMI) have been developed to assist the surgeon in positioning the camera and smoothing the surgical workflow. The first device able to give back to surgeons a direct control over

[1]Department of Electronics, Information and Bioengineering, Politecnico di Milano, Milano, 20133, Italy, (nicolo1.pasini/elena.demomi@mail.polimi.it)

[2]The BioRobotics Institute and Department of Excellence in Robotics & AI, Scuola Superiore Sant'Anna, Pisa, 56121, Italy, (andrea.mariani@santannapisa.it)

[3]Department of Computer Science, Johns Hopkins University, Baltimore, 21218, Maryland, (amunawa2/pkaz@jhu.edu)

978-1-6654-7260-9/22/$31.00 ©2022 IEEE
DOI 10.1109/IRC55401.2022.00073

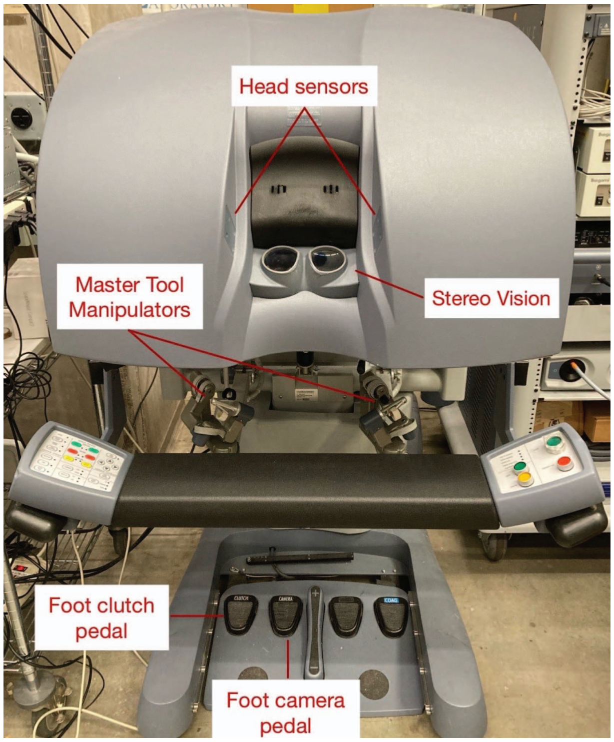

Fig. 1. dVRK Master Console: main elements of the console and their position on the surgeon side. Foot pedal tray and Master Tool Manipulators provide control over camera and tools, only when the head sensors detect the surgeon's presence. The surgical environment is displayed on two stereo displays inside the console.

their FoV was the Automated Endoscopic System for Optimal Positioning (AESOP, Computer Motion Inc., Santa Barbara, California, and later Intuitive Surgical Inc., Sunnyvale, California) [8], belonging to the Robotic-Assisted Endoscopic Manipulators (RAEMs) class. The system integrated a foot interface to control the camera and later introduced voice control, as also implemented for VIKY®(EndoControl Medical, La Tronche, France) and LARS (IBM, Armonk, New York). LAPMAN (Medsys, Gembloux, Belgium) and SOLOASSIST (AKTORmed GmbH, Barbing, Germany) were integrated with joysticks for hand control and EndoAssist (Armstrong Healthcare, High Wycombe, United Kingdom) and Free-Hand (Freehand 2010 Ltd., Guildford, United Kingdom) introduced head movement as a new control modality. Even if significant improvements have been done, for each of the aforementioned devices the direct control of the surgeon over the endoscope is still required.

Thanks to the deployment of open research platforms such as the da Vinci Research Kit (dVRK) [9] and Raven II [10], multiple studies have been conducted and different approaches to endoscope control have been investigated. All the explored modalities can be grouped into three main categories: reactive, proactive and combined systems.

A reactive camera control modality is intended to be a system which has no situation awareness, but rather reacts to the surgeon's input. Several control modalities can be considered reactive, such as eye gaze tracking [11], instrument tracking (both kinematics/dynamics based or image based) [12],[13],[14], external sensors guidance or voice control [8]. A proactive system determines the ongoing phase of the procedure and adjusts the camera's viewpoint by incorporating preexisting knowledge about the visualization requirements and types of movements needed for the procedure. This knowledge can be derived from expert demonstrations, such as in [15] in which they propose a system that applies a Markov model to anticipate the surgeon's tool movements in order to accordingly center the camera, in advance, on the anticipated end-effector midpoint. By merging reactive and proactive systems' characteristics the combined category is obtained, as in [16].

Every existing solution, VR based or not, is a valid option which further closed the gap between direct camera control of traditional and robot-assisted MIS: nevertheless, to the best of our knowledge, none of the proposed methods have analyzed camera motion modalities based on situation awareness during a surgical procedure. The potential of situation aware systems relies on the wide range of actions that the robot can take having rich information about the performed procedure. Otherwise, reactive systems are designed to follow rigid instructions regardless of the situation. A solution has been proposed [16] which combines semantically rich instructions with position hysteresis based on three zones in the endoscope FoV. However, the instrument tracking is performed in 2D, hence the zones of interest which trigger camera motion are defined on a plane inside the FoV of the camera. Furthermore, this approach gives no situation awareness to the system, since these triggering zones are based on camera viewpoint rather than elements of interest inside the FoV. As a result, the effectiveness of situation aware automated camera motion for surgical operations has yet to be assessed.

C. Research hypothesis

Our work is a proof of concept which fits the combined category. Our aim is to provide camera motion based on procedural knowledge during one of the main surgical tasks performed in surgery: suturing. We developed a situation aware autonomous camera motion system that is able to provide an optimal point of view during a suturing task. This was implemented in a virtual environment, and we performed a user study to compare our work with the current foot pedal based camera control. Using the same console setup as dVSS, we analysed both objective improvements in suturing that were introduced by camera automation and the subjective perception of the users.

II. Camera Control

To study camera motion, we first had to develop a camera control modality which resembles the current practice, based on a foot-pedal tray. Then, we implemented an autonomous camera motion modality with situation awareness. These two modalities are detailed in the next two sections.

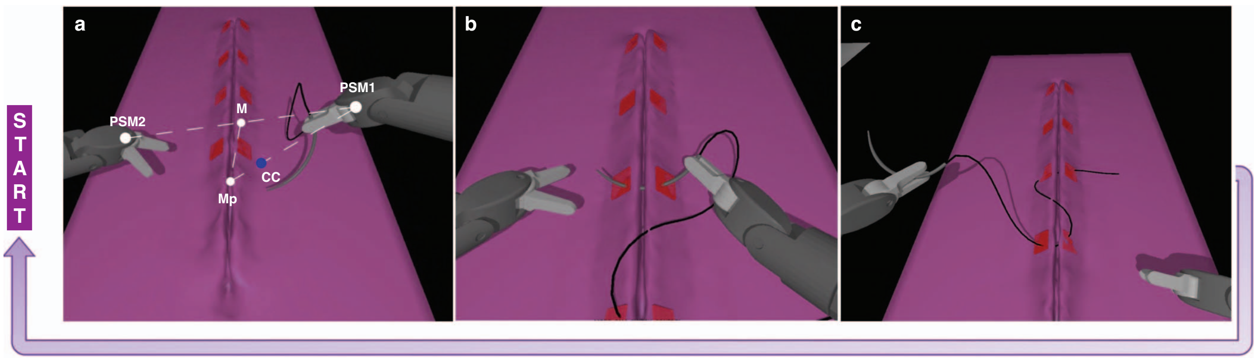

Fig. 2. Needle positioning (a), Tissue bite (b) and Suture throw (c) phases. In (a) the weighted tracking modality is shown: mid-point M is calculated as the geometrical mid-point between PSM1 and PSM2, and projected as Mp on the centerline passing through the stitches. PSM1 is then connected to Mp and the Camera Center (CC) is defined as: $PSM1 + W1 \cdot \mathbf{MpPSM1}$, where **MpPSM1** represents the vector connecting PSM1 to Mp. Similar workflow is applied in (c). Fixed FOV is proposed in (b) during Tissue bite. Stitches are numbered from number 1, closest, to number 4, furthest.

A. *Foot pedal camera motion*

Traditional surgical robotic platforms, such as the da Vinci Robot, are equipped with a foot pedal tray in order to exploit their maximum potential. In the da Vinci, the foot pedals are located right under the master console. In particular, the Camera Pedal, as in Fig. 1, is used to acquire direct control over camera motions: whenever the pedal is pressed, the surgeon loses connection with the tools and gains control over the camera. In particular, the endoscope camera tip (ET) is updated as:

$$ET^{t} = ET^{t-1} - sf\frac{\Delta MTM_{L} + \Delta MTM_{R}}{2} \quad (1)$$

where t represents time, ΔMTM_L and ΔMTM_R denote the variation of left and right master Cartesian position from time *t-1* to t and *sf* stands for scale factor. As a result, only an asynchronous control of tools and camera is permitted.

B. *Autonomous camera motion*

The autonomous camera motion modality is based on tracking of the tools' 3D positions in Cartesian space. The tracking is kinematics-based, hence no image segmentation is needed. This allows us to define the camera's focus during the entire procedure: as [17] reports, different suturing subtasks during surgical procedures require specific adjustments of the FoV. Based on similar assumptions, in Fig. 2 we report the implemented camera motion criteria. A total of 4 gestures have been selected to trigger different camera motions: reaching for the needle, needle positioning, tissue bite and suture throw. Since the main goal of our research is to study the possible beneficial effects introduced by a situation aware navigation system, we decided to rely on Volumes of Interest (VOI) inside the virtual environment, used to segment the suturing task. These volumes are defined as concentric semi-spheres centered on every stitch of the suturing pad and are tailored to trigger a specific camera motion modality whenever the user operates inside of them. As pointed out in [18], specific human gaze patterns can be defined when performing suturing tasks, depicting salient regions inside the FoV. For this reason, we proposed 4 camera motion modalities able to focus on the regions of interest in the environment, respectively:

- Reaching for the needle: whenever the needle is not in between any of the needle drivers' jaws, the camera will hold a steady position, waiting for the task to start or to hold the needle again after having lost its grip.
- Needle positioning: whenever the needle is in between the dominant hand's jaws outside the inner semi-sphere, the camera will track the weighted projected mid-point, as in Fig 2a, with a tailored PSM1 Weight (W1). The selected value for W1 for the user study was 0.9. The position is weighted to reduce motion sickness, a common issue when dealing with automatic camera motion.
- Tissue bite: after having found a suitable insertion position for the needle, the tissue bite will take place inside the inner semi-sphere. For this gesture, we propose a steady zoomed-in position for the camera, as in 2b, to promote a sharp and fixed FoV over the stitch.
- Suture throw: to conclude the suture, needle and thread must be pulled through the stitch. For this gesture, the camera will track the projected mid-point with a tailored PSM2 Weight (W2), as in Fig. 2c. The selected value for W2 for the user study was 0.9. This phase, and its related camera motion, are triggered as soon as the needle and the gripper are outside the outer semi-sphere. The introduction of this second semi-sphere is intended to reduce sudden changes in camera motion modality.

The autonomous camera motion architecture provides a specific tracking modality based on the surgical phase the user is undergoing, meaning the system belongs to the combined class. Knowing the Cartesian position of every element composing the surgical scene, stitches included, gives us the possibility to design a tracking system able to reduce motion sickness with respect to continuous tracking systems. In our work, situation awareness is given by the presence of volumes of interest, which define the procedural sequence based on the global position of the tools, rather than

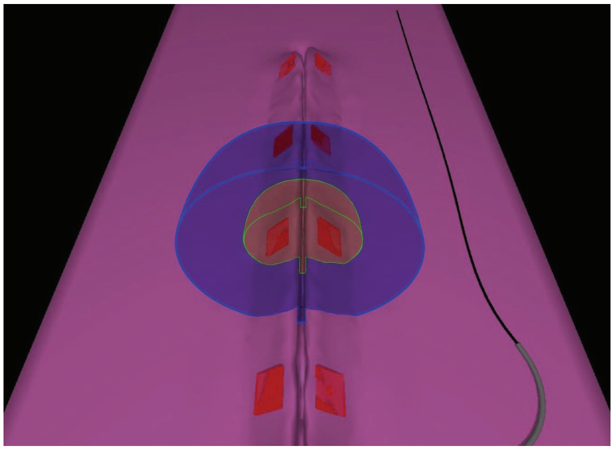

Fig. 3. Volumes of Interest (VOI): in green, the inner semi-sphere which defines the region in which the Tissue Bite phase takes places; in blue, the outer semi-sphere outside which the Needle Positioning and Suture Throw phases take place. Such a design is repeated for every stitch (not shown). Please note that these semi-spheres are here reported and visible only for a better understanding: VOI are not visible while performing the task.

their position with respect to the camera viewpoint. If the algorithm was implemented to track continuously the virtual jaws' positions, or their mid-point, the view would never come to a standstill, resulting in a disturbing experience for the surgeon.

III. Experimental Evaluation

This section describes the experimental setup and the protocol applied during the user study that was carried out to test the hypothesis previously stated. We start with a description of the VR suturing task and performance metrics selected to analyze the final outcome. To conclude, we report performance and workload assessment methodologies and the acquisition protocol.

A. Experimental setup

In order to allow teleoperation, the master console of a da Vinci Research Kit (dVRK) was used for the experimental study. The dVRK is a first generation da Vinci Surgical System, integrated with custom control hardware and software which makes it open access to promote research [9]. The dVRK's master console, which composes the surgeon side of the robot, is displayed in Fig. 1: it is equipped with a foot-pedal tray, two Master Tool Manipulators (MTMs) and a stereo viewer (upgraded to a resolution of 1280×1024 per eye) for the visualization of the surgical environment through the endoscope. The robotic platform was integrated with the virtual environment: it is based on the 2021-22 AccelNet Surgical Robotics Challenge [19] which reproduces a suturing task, one of the most common procedures performed during MIS, thanks to the Asynchronous Multi-Body Framework (AMBF) simulator [20]. AMBF uses the Robot Operating System (ROS) as a control middleware which allows the easy integration of the dVRK MTMs with the virtual environment. Furthermore, AMBF can stream the depth and video data on standardized ROS payloads so that they can be stored using ROS Bags.

B. Virtual Reality task

Thanks to the manipulators inside the Master Console, users were able to control two virtual Patient Side Manipulators (PSMs) equipped with 6 degrees of freedom (DoF) Needle Drivers, as in Fig. 2. The selected scale factor is equal to 0.5 (meaning every Master's movement was halved inside the virtual environment). Using the Surgical Robotics Challenge environment, we designed a suturing task comprehensive of suturing pad, needle, thread and two surgical needle drivers. The shape of the pad resembles the typical shape of suture training pads, with a linear direction and both entry and exit points for every stitch. Every participant was asked to use both hands, hence performing multiple instrument to instrument exchanges, to perform each sub-phase of suturing, as reported in Fig. 2. The stitches' entry and exit points are represented as red squares, with dimensions of $5 \times 5mm$, at an approximate center-center distance of $2.55cm$.

C. Performance Metrics

Both objective and subjective metrics were defined to analyze the user study outcomes. In order to quantify the user's performance from an objective perspective, we considered 5 metrics: PSMs total path length (2), clutch pedal presses, camera pedal presses and completion time. The PSMs total path length refers to the total distance covered by the needle drivers inside the virtual environment while completing the task. The clutch pedal presses addresses how many times the user needed to readjust the position of the MTMs due to bad positioning of the virtual PSMs.

Upon completing the study, every user was asked to complete two NASA Task Load Index (TLX) surveys [21], one for each camera control modality, to assess the subjective workload by incorporating a multi-dimensional rating procedure. The NASA TLX derives an overall workload score based on a weighted average of ratings on six sub-scales: Mental Demand, Physical Demand, Temporal Demand, Performance, Effort, Frustration.

D. Acquisition Protocol

To evaluate the beneficial impact that the proposed autonomous tracking modality could bring to teleoperated suturing, we performed a user study consisting of 10 non-medical users (20 to 26 years old, 3 females and 7 males, only one left-handed). To assess their predisposition to the task, users were asked to complete a pre-experimental survey in which they were asked whether they played any musical instrument, sport, videogame or used any robotic system before. Since no statistical difference was noted, they have been classified as having the same low level of expertise regarding teleoperated systems. The participants were asked to complete the suturing task of picking up the needle, positioning it in the bottom right corner, and performing a total of 7 stitches, from stitch 1 to stitch 4 and back, as depicted in Fig. 3. Both modalities were tested 3 times

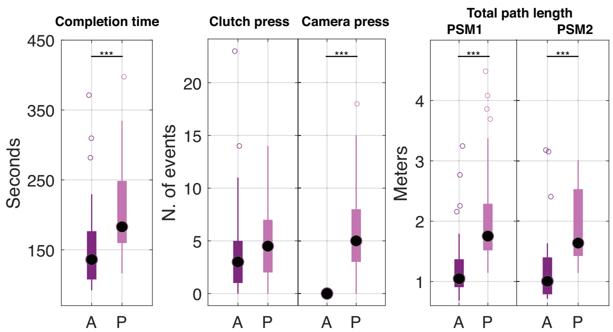

Fig. 4. Objective metrics: statistical difference is shown with $p < 0.001$ for both Time completion, Camera presses and PSMs total path length. Modalities are identified as Autonomous (A) and Foot Pedal (P) camera motion.

each, for a total of 6 repetitions per user, 60 repetitions total. All subjects started the suturing tasks with the same starting conditions and fixed FoV. The proposed view has been selected to allow the visualization of all four stitches, so that the choice of whether or not to move the camera is left to the user. Regarding the autonomous camera modality, the tracking started as soon as the user reached the needle–from that moment the camera was under complete control of the tracking system. Every user was asked to repeat the task with alternate modalities, avoiding three consecutive repetitions with the same camera control modality to not introduce any learning effect. Each participant was given an introductory lecture, in which the main components of the Master Console were described, and 5 minutes of training time, during which they could familiarize themselves with the robotic platform. The suturing environment was displayed on two stereo viewers, placed inside the Master Console, in which the user places his/her head, as in Fig. 1. The experiments were carried out after Institutional Review Board (IRB) approval (protocol number: HIRB00000701), with oral consent from participants. The official NASA Task Load Index (TLX) App has been used on an iOS system for measuring the subjective workload. The app has been designed to ensure the privacy of research participant data: NASA TLX anonymizes all results and does not send any personal identifiable information (PII) to any data servers.

E. Statistical Analysis

Due to the relatively small sample size, we decided to perform non-parametric statistical significance tests to assess the effects introduced by the autonomous camera motion. The Wilcoxon signed rank test was selected, considering repetitions with different modalities as two populations with paired observations. Statistically significant results were assessed at different values of p, as follows: * for $p < 0.05$, ** for $p < 0.01$, *** for $p < 0.001$. The statistical analysis was performed in MATLAB using the *signrank()* command.

IV. Results and Discussion

Every subject enrolled in the experiment succeeded in completing the suturing task in the virtual environment 6 times, except for one user who completed 4 stitches out of 7 before losing track of the needle with the foot pedal camera control modality. The primary aim of this study is to determine whether the proposed autonomous camera motion system introduces benefits for suturing during laparoscopic surgery, both in terms of mental and physical workload reduction. Fig. 4 shows the results associated to the objective metrics, which demonstrate a significant statistical difference between the autonomous and foot pedal camera motion modalities for completion time and total PSMs path length, with respectively $p_t < 0.001$ and $p_L < 0.001$. Regarding clutch pedal total presses, no statistical difference has been noted ($p > 0.05$). Such a result can be explained by looking at the physical constraint introduced by the master console in every teleoperated surgical robot: once the MTMs reach their maximum extent inside the console's free-space, subjects need to reposition their hands using the clutch pedal, which allows to reposition the MTMs without moving the PSMs inside the virtual environment. Since motion of the PSMs is strictly managed by the user only, different camera motion modalities will not affect this metric. The average duration of a single repetition was respectively 158.5 s for the autonomous camera and 207 s for the foot pedal modality, confirming what was previously stated by similar works [13],[22]. A reduction in completion time corresponds to a reduction in operational time, hence less physical stress and effort for the surgeon [23]. This reduction in completion time can be related to the cognitive overload introduced by pressing pedals, and by the additional time required to control the camera. The reduction of distance covered by the PSMs may be linked to a reduction of operation workspace. Indeed, in order to complete the same task less movements are necessary, reducing the physical and mental workload for the surgeon and the potential for harm to the patient. The performance improvement, even though with different impact, affected all the users, allowing them to focus on tools control rather than camera navigation.

To complete the suturing task, the proposed camera motion modalities require different levels of effort from the users, both mentally and physically. As depicted in Fig. 5, this result is confirmed by the subjective evaluation performed through the NASA TLX surveys. We report here the 6 subscales used to assess the overall workload score through a weighted average for both modalities. Both the comparisons show a statistical difference between the two study groups, respectively with $p < 0.05$ for a mental demand comparison and $p < 0.01$ regarding the weighted rating. This result is of particular interest if we consider that among the subscales, mental demand is considered the most relevant in assessing the overall workload, on a scale from 0 to 5, where a higher weight means higher relevance in computing the final weighted rating. This result strengthens the previously stated hypothesis: reduction in mental overload given by an autonomous motion of the camera allowed users to focus on the execution of the task, resulting in an eased and smoother workflow.

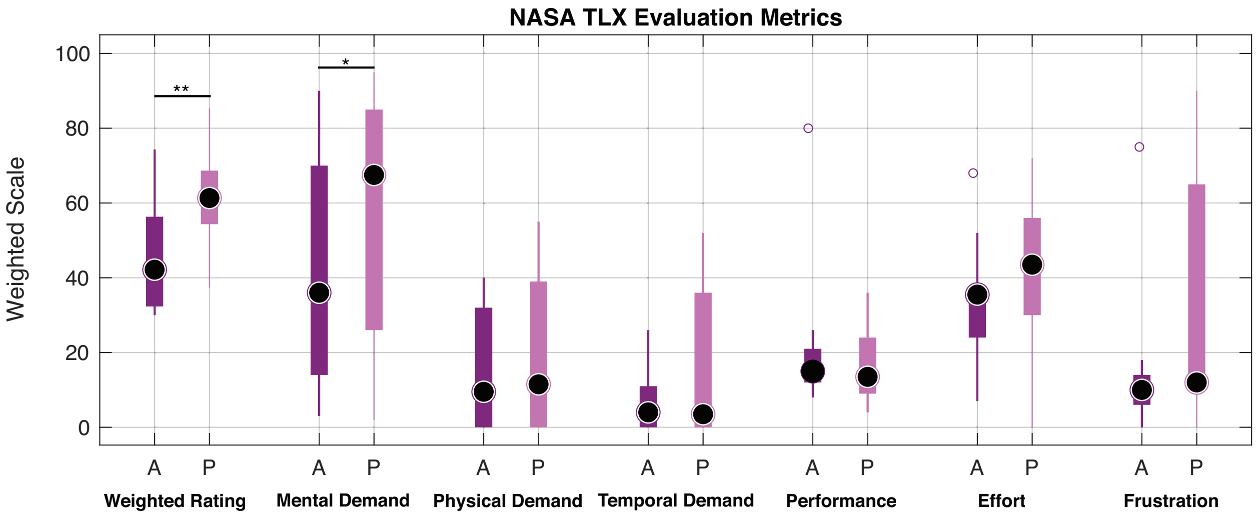

Fig. 5. NASA TLX: subjective workload evaluation. Statistical difference is shown with $p < 0.05$ for Mental Workload and $p < 0.01$ for Weighted Rating. Please note that the weighted values for the sub-scales have been rescaled between 0 and 100, dividing by 5, in order to allow a common representation along with the Weighted Rating. Modalities are identified as Autonomous (A) and Foot Pedal (P) camera motion.

V. Conclusion and Future Works

This work focuses on the effectiveness and benefits introduced by autonomous camera navigation enhanced with procedural knowledge during laparoscopic surgery via robotic assisted minimally invasive surgery. In particular, a situation aware autonomous camera motion system is introduced and compared with current manual teleoperation of the camera by carrying out a user study with non-medical participants performing a suturing task in a virtual environment. The main outcomes are the following:

- Autonomous camera motion allowed users to improve their efficiency both in terms of time and physical effort with respect to manual camera navigation.
- Autonomous camera motion allowed to reduce the mental stress and cognitive overload of users, freeing them from the burden of manual camera control.

Surgeons can directly benefit from these outcomes, since these can result in shorter surgery times and lower cognitive workload.

This work consists in a proof of concept, which aims to pave the way to future dry-lab considerations involving current research platforms for surgical robots. A wider population should be analyzed, involving medical experts. To this extent, a future work of ours will analyze the effect of a situation aware autonomous camera motion system, based on online suturing gesture recognition. In this future work, we will assess the reliability of an online neural network model to classify surgical gestures and we will study the effects of such a system with a dry-lab user study, comparing the outcomes with a pre-existing System for Camera Autonomous Navigation (SCAN) [13]. The aim of our study will be to evaluate the benefits introduced by situation awareness, which could be the next step for an intelligent robot assistant.

VI. Acknowledgements

This research was partially supported by NSF OISE 1927354. The authors are grateful for the support of Anton Deguet.

References

[1] S. Maeso, M. Reza, J. A. Mayol, *et al.*, "Efficacy of the da Vinci surgical system in abdominal surgery compared with that of laparoscopy: A systematic review and meta-analysis," *Annals of surgery*, vol. 252, no. 2, pp. 254–262, 2010.

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

[3] K. Moorthy, Y. Munz, A. Dosis, *et al.*, "Dexterity enhancement with robotic surgery," *Surgical Endoscopy and Other Interventional Techniques*, vol. 18, no. 5, pp. 790–795, 2004.

[4] G. Lee, M. Lee, I. Green, M. Allaf, and M. Marohn, "Surgeons' physical discomfort and symptoms during robotic surgery: A comprehensive ergonomic survey study," *Surgical endoscopy*, vol. 31, no. 4, pp. 1697–1706, 2017.

[5] B. Zheng, M. A. Cassera, D. V. Martinec, G. O. Spaun, and L. L. Swanström, "Measuring mental workload during the performance of advanced laparoscopic tasks," *Surgical endoscopy*, vol. 24, no. 1, pp. 45–50, 2010.

[6] C. Perrenot, M. Perez, N. Tran, *et al.*, "The virtual reality simulator dv-trainer® is a valid assessment tool for robotic surgical skills," *Surgical endoscopy*, vol. 26, no. 9, pp. 2587–2593, 2012.

[7] J. S. Cho, K. Y. Hahn, J. M. Kwak, *et al.*, "Virtual reality training improves da Vinci performance: A prospective trial," *Journal of Laparoendoscopic & Advanced Surgical Techniques*, vol. 23, no. 12, pp. 992–998, 2013.

[8] S. Unger, H. Unger, and R. Bass, "Aesop robotic arm.," *Surgical endoscopy*, vol. 8, no. 9, pp. 1131–1131, 1994.

[9] P. Kazanzides, Z. Chen, A. Deguet, G. S. Fischer, R. H. Taylor, and S. P. DiMaio, "An open-source research kit for the da Vinci® surgical system," in *2014 IEEE international conference on robotics and automation (ICRA)*, IEEE, 2014, pp. 6434–6439.

[10] B. Hannaford, J. Rosen, D. C. W. Friedman, *et al.*, "Raven-II: An open platform for surgical robotics research," *IEEE Transactions on Biomedical Engineering*, vol. 60, pp. 954–959, 2013.

[11] S. Ali, L. A. Reisner, B. King, *et al.*, "Eye gaze tracking for endoscopic camera positioning: An application of a hardware/software interface developed to automate aesop.," *Studies in health technology and informatics*, vol. 132, pp. 4–7, 2008.

[12] 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.

[13] T. Da Col, A. Mariani, A. Deguet, A. Menciassi, P. Kazanzides, and E. De Momi, "Scan: System for camera autonomous navigation in robotic-assisted surgery," in *2020 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, IEEE, 2020, pp. 2996–3002.

[14] K. Hutchinson, M. S. Yasar, H. Bhatia, and H. Alemzadeh, "A reactive autonomous camera system for the Raven-II surgical robot," in *2020 International Symposium on Medical Robotics (ISMR)*, IEEE, 2020, pp. 195–201.

[15] O. Weede, H. Mönnich, B. Müller, and H. Wörn, "An intelligent and autonomous endoscopic guidance system for minimally invasive surgery," in *2011 IEEE international conference on robotics and automation*, IEEE, 2011, pp. 5762–5768.

[16] C. Gruijthuijsen, L. C. Garcia-Peraza-Herrera, G. Borghesan, *et al.*, "Autonomous robotic endoscope control based on semantically rich instructions," *arXiv preprint arXiv:2107.02317*, 2021.

[17] R. D. Ellis, A. J. Munaco, L. A. Reisner, *et al.*, "Task analysis of laparoscopic camera control schemes," *The International Journal of Medical Robotics and Computer Assisted Surgery*, vol. 12, no. 4, pp. 576–584, 2016.

[18] A. A. Awale and D. Sarikaya, "Using human gaze for surgical activity recognition," in *2022 30th Signal Processing and Communications Applications Conference (SIU)*, IEEE, 2022, pp. 1–4.

[19] https://collaborative-robotics.github.io/surgical-robotics-challenge/challenge-2021.html.

[20] A. Munawar and G. S. Fischer, "An asynchronous multi-body simulation framework for real-time dynamics, haptics and learning with application to surgical robots," in *2019 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, IEEE, 2019, pp. 6268–6275.

[21] S. G. Hart and L. E. Staveland, "Development of nasa-tlx (task load index): Results of empirical and theoretical research," in *Advances in psychology*, vol. 52, Elsevier, 1988, pp. 139–183.

[22] A. Mariani, G. Colaci, T. Da Col, *et al.*, "An experimental comparison towards autonomous camera navigation to optimize training in robot assisted surgery," *IEEE Robotics and Automation Letters*, vol. 5, no. 2, pp. 1461–1467, 2020.

[23] K. H. Goodell, C. G. Cao, and S. D. Schwaitzberg, "Effects of cognitive distraction on performance of laparoscopic surgical tasks," *Journal of Laparoendoscopic & Advanced Surgical Techniques*, vol. 16, no. 2, pp. 94–98, 2006.