Reducing Workload during Brain Surgery with Robot-Assisted Autonomous Exoscope

Elisa Iovene[1], Alessandro Casella[1,], Junling Fu[1], Federico Pessina[2], Marco Riva[2], Giancarlo Ferrigno[1], Elena De Momi[1]

***Abstract*— In this paper, a position-based visual-servoing control approach is introduced for a robotic camera holder to improve ergonomics and reduce mental stress during brain surgery. The visual tracking system controls and moves the robotic camera holder by following a selected surgical instrument without the need for artificial markers. The system was validated using a 7 Degree-of-Freedoms (DoFs) redundant robotic manipulator with an eye-in-hand stereo camera configuration and compared with conventional control methods using NASA TLX survey and four objective metrics, including execution time, time out of field of view (FoV), target score, and path length. Experimental results demonstrate that the proposed system can reduce the surgeon's workload during brain surgery-related task execution, improve ergonomics and achieve higher performance than traditional control methods.**

I. Introduction

Work-related musculoskeletal disorders (WMSDs) are critical issues among neurosurgeons, affecting their life quality and career length [1]. These disorders are mainly caused by non-neutral positions during surgical operations, especially when using microscopes and keeping their eyes bound to the oculars. The introduction of exoscopes allows surgeons to maintain a neutral, upright spinal position, leading to better ergonomics [2]. Generally, an exoscope system utilize a scope placed outside the body cavity and projects high-resolution view of the surgical field onto a two-dimensional (2D) or three-dimensional (3D) monitor. This provides the surgeon with improved ergonomics, a greater comfort rate, and less fatigue, especially for long time surgical operational procedures [3].

Although exoscope systems like VITOM-3D and ORBEYE have shown comparable results to conventional microscopes [4], they suffer from limitations due to manual adjustments. Manual repositioning interrupts surgery, leading to longer operation times and reduced efficiency [5]. Moreover, switching between tasks increases cognitive load and error likelihood for neurosurgeons [6].

To deal with these issues, several models attempt to overcome the limitation by incorporating a foot control pedal, but this difficult-to-manage control strategy add complexity, increasing the surgeon's cognitive load [5]. Meanwhile, the inconvenience of repositioning can detract from the benefits of using the exoscope in the operating room, which underscores the importance of hands-free camera movement. Therefore, how to achieve hands-free control of the exoscope system to allow the neurosurgeon to perform uninterrupted bimanual surgery is a critical issue.

[1]E. Iovene, J. Fu*, A. Casella, G. Ferrigno, and E. De Momi are with the Department of Electronics, Information and Bioengineering, Politecnico di Milano, 20133 Milan, Italy. *Corresponding Author. elisa.iovene@polimi.it, junling.fu@polimi.it, alessandro.casella@polimi.it, giancarlo.ferrigno@polimi.it, elena.demomi@polimi.it

[2]F. Pessina and M. Riva are with the Department of Biomedical Sciences, Humanitas University, Pieve Emanuele, Milan, Italy and also with the IRCCS Humanitas Research Hospital, Rozzano, Milan, Italy federico.pessina@hunimed.eu, marco.riva@hunimed.eu

A. Related Work

In recent years, robot-assisted autonomous navigation techniques have been integrated into brain surgery applications, offering superior spatial positioning, dexterity, and non-fatigability. Various robotic camera control strategies, including voice, joysticks, and gaze tracking-based control, have been implemented [7]. Although the level of autonomy (LoA) of robotic exoscope systems has improved, there remains a steep learning curve for surgeons. To further reduce the surgeon's workload, minimizing the need for direct intervention in camera control (LoA 2 [8]) is desirable.

Modus V (Synaptive Medical, Toronto, Canada) developed a hands-free robotic camera system using fiducial markers attached to the tracked surgical instruments [9]. But this method has maneuverability limitations and may require external tracking systems. Marker-based visual-servoing using ArUco codes has been proposed by [10] for da Vinci endoscope movement. Instrument detection via Aruco codes may be unreliable in real scenarios with blood, smoke, or fluid present, and markers occlusion can cause interruptions during brain surgery [11].

Instrument localization has been investigated with various techniques, including robot kinematics-based [12][13] and image-based tracking [14]. Markerless approach provides a fast and accurate reconstruction of the surgical instrument's 3D position from the intra-operative images and can be integrated into the robotic control framework [15]. This approach has been applied to various camera systems, ensuring smooth and controlled movements [16][17][18]. Overall, the markerless-based visual-servoing approach holds great potential for enhancing brain surgery procedures by eliminating the need for external sensors or additional markers.

B. Contributions

This paper presents a robot-assisted autonomous exoscope to reduce the workload of the surgeon during brain surgery. The main contributions of this work are:

- A novel robot-assisted autonomous exoscope system

that employs markerless visual-servoing to reduce workload during brain surgery.
- Comprehensive experiments designed to compare the system's performance with traditional manual and foot remote control methods.
- User study and perceived cognitive workload when using four different exoscope system control modalities in a simulated brain surgery task are explored.

The remainder of this paper is organized as follows. Section II describes materials and methods of the proposed system. Section III depicts the experimental setup for system characterisation and system usability validation. Afterwards, experimental results are illustrated and discussed in Section IV. Finally, conclusions are reported in Section V.

II. Materials And Methods

In this section, the markerless visual-servoing and robot-assisted system for autonomous exoscope is proposed. Section II.A. presents the instrument detection and localisation method. Section II.B. describes the robot position and orientation control strategies. Fig. 1 illustrates the generic diagram of the overall system.

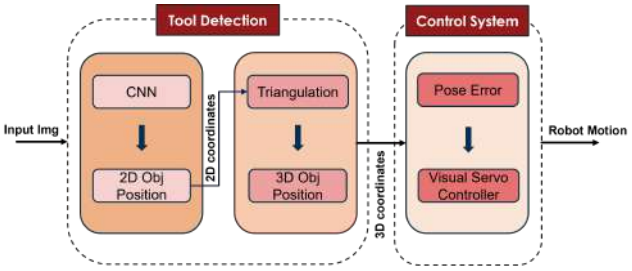

Fig. 1. Architecture of the proposed system. Firstly, image acquired by the camera during brain surgery is send to a CNN which detects the surgical tool. Then, the 3D position of the surgical tool is extracted and feed to a visual-servoing-based robotic controller for autonomous navigation.

A. Surgical Tool Detection

A pre-trained CNN YoloV5 [19] was employed to detect the target in the proposed system. Fine-tuning was conducted to identify a specific surgical instrument. The CNN received downsampled images of size 640x640 and provided the position of the instrument's tip through a bounding box and prediction confidence. The center of the bounding box was considered to be the position of the instrument, $(\mathbf{x_R}, \mathbf{y_R})$. Only the right images from the stereo camera were used for inference and epipolar geometry was used to determine the position of the instrument in the left image [20]. The fundamental matrix, F, was used to find the epipolar line, $l^{'} = Fx^{'}$ associated with the point, $x^{'}$, in the left image. To locate the corresponding point, a sliding window of the same size as the bounding box was moved along the epipolar line until achieving maximum cross-correlation, as shown in Fig. 2. After detecting the instrument's 2D image coordinates in both images, the 3D position of the tip was computed by triangulation with respect to the right camera reference frame, $\boldsymbol{P}^{C}_{tool}$.

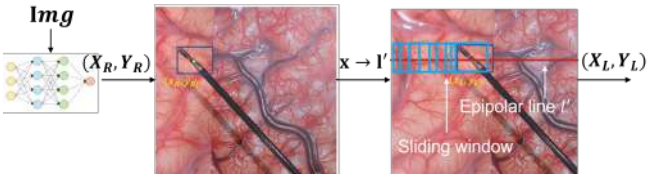

Fig. 2. The right camera image was fed into the neural network which predicted the position of the instrument,$(\mathbf{x_R}, \mathbf{y_R})$. The epipolar line $l^{'}$ (red line) was then found and a window was slid along it to find the corresponding point in the left image,$(\mathbf{x_L}, \mathbf{y_L})$.

B. Robot Control System Design

The coordinate transformations between robot, camera, and surgical instrument are illustrated in Fig. 3.

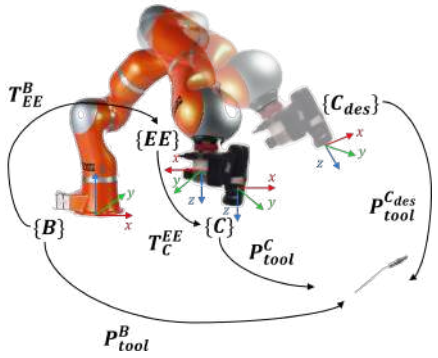

Fig. 3. Frames of interest. $\boldsymbol{T}^{B}_{EE}$, transformation between the robot base $\{\boldsymbol{B}\}$ and the robot end effector $\{\boldsymbol{EE}\}$, $\boldsymbol{T}^{EE}_{C}$, position of the camera $\{\boldsymbol{C}\}$ relative to $\{\boldsymbol{EE}\}$, used to determine the position of the camera in $\{\boldsymbol{B}\}$, $\boldsymbol{T}^{B}_{C}$. $\boldsymbol{P}^{C}_{tool}$ and $\boldsymbol{P}^{C_{des}}_{tool}$, actual and desired position of the tracked object in the camera space, respectively. $\boldsymbol{P}^{B}_{tool}$, position of the tool in the robot's base reference frame.

In this work, a position-based visual servoing (PBVS) control architecture was exploited. Since surgical exoscope movements can be divided into translational and rotational, we employed two control strategies to control the 6-DoFs of the robot end-effector(EE) in Cartesian space. The first strategy controlled the position of the EE, while the orientation remained fixed. On the contrary, the second controlled the orientation while the position remained fixed.

1) Position Control: position control error was given as:

$$\boldsymbol{e_{pos}} = \boldsymbol{P}^{B}_{C_{des}} - \boldsymbol{P}^{B}_{C} \tag{1}$$

where $\boldsymbol{P}^{B}_{C} = \boldsymbol{P}^{C}_{EE} * \boldsymbol{P}^{EE}_{C}$ is the actual position of the camera and $\boldsymbol{P}^{B}_{C_{des}}$ is the desired position of the camera in $\{\boldsymbol{B}\}$. Starting from the target position, the desired position of the camera could be obtained as:

$$\boldsymbol{P}^{B}_{C_{des}} = \boldsymbol{P}^{B}_{tool} * (\boldsymbol{P}^{C_{des}}_{tool})^{-1} \tag{2}$$

where $\boldsymbol{P}^{C_{des}}_{tool} = \begin{bmatrix} 0 & 0 & d & 1 \end{bmatrix}$ as the goal to keep the instrument near the center of the camera image with a distance d. This error, $\boldsymbol{e}_{pos}$, was then fed into a Cartesian PD (Proportional Derivative) controller which calculates the desired joint velocities:

$$\dot{\mathbf{q}}_{\mathbf{d}} = \mathbf{J}^{+}(\mathbf{q})(\mathbf{k_p e_{pos}} + \mathbf{k_d} \dot{\mathbf{e}}_{\mathbf{pos}}) \tag{3}$$

where k_p and k_d are the controller gains found empirically and $\boldsymbol{J}^{+}(\boldsymbol{q})$ is the pseudoinverse of the Jacobian matrix.

2) Orientation Control: in orientation control, the position of the target object, $\boldsymbol{P}^{B}_{tool}$, was used to compute the desired orientation of the EE, which rotated around a fixed point, $\boldsymbol{P}_t$, located in the middle of the tool [21] in Fig. 4.

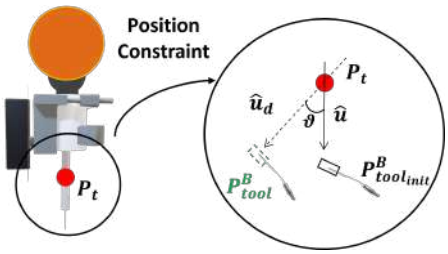

Fig. 4. Position constraint representation. The robot rotates around $\boldsymbol{P}_t$ by an angle θ calculated as the angle between the initial position $\boldsymbol{P}^{B}_{tool_{init}}$ and the desired position $\boldsymbol{P}^{B}_{tool}$.

First, the rotation angle between the initial tool direction $\hat{\boldsymbol{u}} = \boldsymbol{P}^{B}_{tool_{init}} - \boldsymbol{P}_t$ and the desired tool direction $\hat{\boldsymbol{u_d}} = \boldsymbol{P}^{B}_{tool} - \boldsymbol{P}_t$ was defined as:

$$\theta = \mathbf{atan^{-1}} \frac{||\hat{\mathbf{u}}_\mathbf{d} \times \hat{\mathbf{u}}||}{\hat{\mathbf{u}}_\mathbf{d} * \hat{\mathbf{u}}} \tag{4}$$

The vector describing the rotation from $\hat{\boldsymbol{u}}$ to $\hat{\boldsymbol{u_d}}$ in Fig. 4 was computed as:

$$\mathbf{u} = \frac{\hat{\mathbf{u}}_\mathbf{d} \times \hat{\mathbf{u}}}{||\hat{\mathbf{u}}_\mathbf{d} \times \hat{\mathbf{u}}||} \tag{5}$$

Finally, the desired orientation was calculated as:

$$\boldsymbol{R}^{B}_{C_{des}} = (\mathbf{I} + \mathbf{\Lambda sin}(\theta) + \mathbf{2\Lambda^2 sin2}(\frac{\theta}{\mathbf{2}}))\boldsymbol{R}^{B}_{C} \tag{6}$$

where $\mathbf{\Lambda}$ is the rotation matrix given by $\boldsymbol{u}$ and $\boldsymbol{R}^{B}_{C}$ is the actual orientation of the EE computed from the kinematic chain of the robot. From $\boldsymbol{R}^{B}_{C_{des}}$, the orientation error was computed using quaternion notation as:

$$\boldsymbol{e}_{rot} = \boldsymbol{q}_{rotC} * \boldsymbol{q}_{rot}^{-1}_{C_{des}} \tag{7}$$

where $\boldsymbol{q}_{rotC}$ and $\boldsymbol{q}_{rot}^{-1}_{C_{des}}$ are the current and desired quaternions. This error, $\boldsymbol{e}_{rot}$ was then sent to the inverse kinematics controller to calculate desired joint velocities in Eq.(3).

III. Experimental Setup

To simulate the exoscope system and validate the proposed autonomous framework, a 7-DoFs redundant robotic manipulator (LWR 4+ lightweight robot, KUKA, Germany) with an eye-in-hand stereo camera configuration (JVC GS-TD1 Full HD 3D Camcorder) were implemented.

A. Surgical Instrument Detection

A total of 5900 images were involved for constructing the training dataset, 4100 of which were manually recorded and annotated, while the others were extracted from the 2017 EndoVis challenge [22]. Approximately 90% of the dataset was used for training, with the remaining 10% was used for validation testing. The model was trained on a 40GB GPU (Nvidia A100) for 100 epochs with a batch size of 16. The learning rate was set to 0.01 based on the results from the Stochastic Gradient Descent (SGD) network's optimizer. An intersection over union (IoU) $\geq$ 45 %, representing how much the predicted bounding box overlapped with the ground truth, was considered as a true positive (TP). An IoU value below this threshold represented a false positive (FP). The accuracy of the detection was assessed with the average precision (AP), determined by calculating the area under the precision-recall curve $p(r)$: $AP = \int_0^1 p(r)$. A higher value for the area under the curve indicates that the classifier is providing accurate results with both high recall and high precision. As for the detection time, we considered the total amount of time for the model to detect and estimate the position of the target object. Moreover, since the movements being analysed are relatively small, the inference was performed every two frames providing a reasonable compromise between accuracy and time.

B. Surgical Instrument Tracking

The performance of the visual control was investigated as a function of the velocity of the target, i.e., the surgical instrument. To get a realistic estimate of the speed of the instrument in neurosurgery, the velocity was extracted from the neurosurgeon's demonstrations. The neurosurgeon was asked to move the surgical instrument along a 7 cm circular path at two different speeds, one slow and one fast, corresponding to the typical speed used in brain surgery. Three repetitions were recorded for each speed. The velocity was extracted by tracking the position of the instrument over time, and the average value of the repetitions was taken as the reference value, which turned out to be 0.02 m/s for the slow velocity and 0.048 m/s for the fast movement.

The position error, E_{pos} defined as the error between the camera and the target, was used to quantify the performance of the tracking module for the position controller:

$$\boldsymbol{E}_{pos} = ||\boldsymbol{P}^{B}_{C} - \boldsymbol{P}^{B}_{tool}|| \tag{8}$$

The tracking error tolerance was set as $0.005m$. As for the orientation controller, the orientation error was considered to be the distance between the two sets of Euler angles:

$$\boldsymbol{E}_{or} = \boldsymbol{d}(\boldsymbol{\phi}_{C_{des}}, \boldsymbol{\phi}_{C}) \tag{9}$$

where $\boldsymbol{\phi}_{C_{des}}$ and $\boldsymbol{\phi}_{C}$ denote respectively the desired and the actual Euler angles of the camera obtained from $\boldsymbol{R}^{B}_{C_{des}}$ and $\boldsymbol{R}^{B}_{C}$ and $\boldsymbol{d}$ denotes the normalised difference between the two angles so that $0 \leq d \leq \pi$:

$$d(a, b) = min\{|a - b|, 2\pi - |a - b|\} \tag{10}$$

C. Wire Chaser Task

The proposed system was tested on the "wire chaser" task. The task requires precise instrument control and hand-eye coordination, critical skills for performing neurosurgical procedures. Our hypothesis was that the use of the proposed autonomous camera system could lead to improved performance and reduced cognitive load compared to traditional

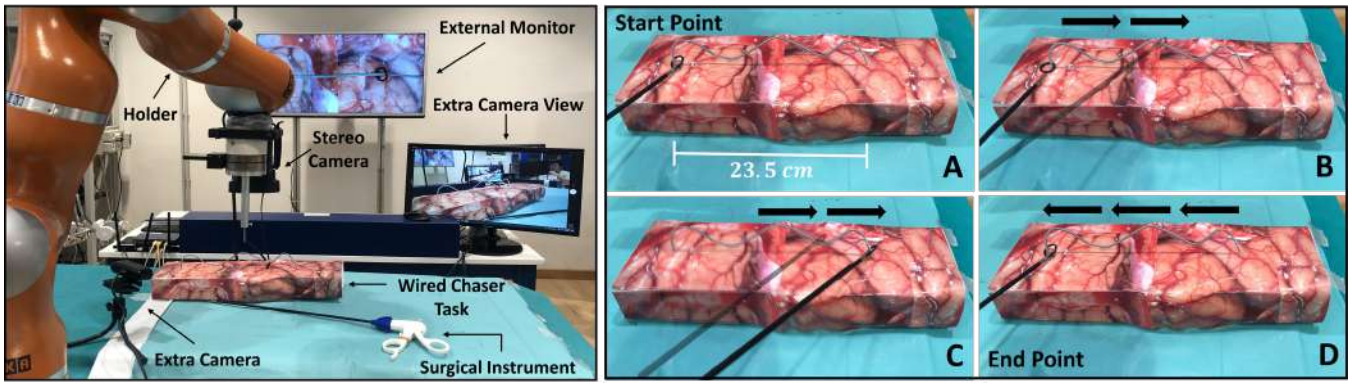

Fig. 5. Overall experimental setup. Left, the task scene: users perform the task while viewing images from the exoscope on an external monitor. A second camera was added to give the user a sense of depth. Right, the task description: users have to grasp the ring with the surgical instrument at a start point (A), move it along the wire (B) and (C) and return to the start position(D).

methods, as a good visualization of the scene is crucial to the task. During the experiment, participants were instructed to move a ring ($\phi = 14$ mm) along a wire (23.5 cm long) placed on a phantom (Level of Clinical Realism - LoCR 1 [23]), using a surgical instrument. The participants performed the task while observing the scene through an external monitor that displayed the image stream. Additionally, an external camera and monitor were set up to provide the participants with depth information. Fig. 5 shows the overall experimental setup. The task was performed with four different modalities and listed as follows:

1) Autonomous Movement - Translation AM_T: the system moved autonomously to keep the instrument in the center of FoV. The activation and deactivation of the motion were done by pressing a foot pedal.
2) Manual Movement MM_T: the system was moved manually. The user had to put down the instrument, move the robotic holder and pick it up again to continue the task.
3) Autonomous Movement - Rotation AM_R: the system rotated autonomously along a fixed point to keep the instrument in the FoV. The activation and deactivation of the motion were done by pressing a foot pedal.
4) Foot-remote Movement FM_R: the user rotated the system using a foot joystick.

where AM_T and MM_T refer to the position control strategy and were compared together as well as AM_R and FM_R that account for the orientation control. 10 non-medical participants were included in the user study. All subjects gave their informed consent. Three repetitions were performed for each modality. Both quantitative and qualitative analyses were conducted to assess users' task load. To determine the measurement tools, the relationship between workload, procedure duration, and performance was considered [25]. Therefore, the following objective metrics were analysed:

- Execution Time [s]: The duration from grasping the ring to placing it back in its original location.
- Time out FoV [s]: Total time the instrument was out of camera's view, and therefore not visible to the user.
- Target Scoring: Penalty of -1 assigned when the user dropped the ring; 0 otherwise.
- Path Length [m]: Sum of distances covered by the instrument between each sampling frame: $\sum_{i=2}^{length(x)} ||P_{tool}(i) - P_{tool}(i-1)||$

where $\boldsymbol{P}_{tool}$ is the position of the instrument. Finally, the workload perceived during the task was assessed using NASA-TLX [26] survey. Users were asked to rank on a scale from 0 to 100 the mental, physical, and temporal demand, performance, effort, and frustration for each modality.

IV. Results & Discussion

A. Surgical Instrument Detection Results

Experimental results show that the AP of the instrument detection model achieves 99.3 % for the chosen confidence threshold, comparable to the results reported in the literature [19]. Besides, the average detection time per frame is 0.066 ± 0.01 s resulting in a processing speed of 15 Hz. Furthermore, a compromise strategy was found by making inferences every 2 frames. This strategy balances detection accuracy and speed, avoiding unnecessary updates and reducing computational loads. As a result, the system can operate in real-time, capturing critical information at a processing speed of 30 Hz.

B. Surgical Instrument Tracking Results

Both position and orientation tracking errors were calculated to evaluate the performance of the proposed control strategy. Mean errors and standard deviations for all three repetitions with different movement speeds are reported in Table I. The mean position and orientation error remains

TABLE I
Position and Orientation Error

	Position Error $E_{pos}[cm]$		Orientation Error $E_{or}[deg]$	
Rep	**Fast**	**Slow**	**Fast**	**Slow**
1	2.31±1.71	1.45±0.71	9.55±9.02	6.34±5.76
2	2.48±0.74	1.47±0.56	9.9±7.39	8.13±6.18
3	2.51±1.38	1.57±0.62	9.34±7.75	8.12±6.11

relatively constant for each repetition under different speed conditions (fast and slow), proving the repeatability and

robustness of the system. The reported tracking errors are mainly caused by the response speed of the robot controller, the setting of error tolerance (0.5 cm), and the surgical tool detection time cost. The maximum position error for slow and fast movement in all 3 repetitions are 2.51 ± 1.38 cm/s and 1.57 ± 0.62 cm/s. Considering the surgeon demonstrated surgical instrument movement velocity with 4.8 cm/s and 2 cm/s for these two speeds level (Section III.B.), the maximum position error may cause a time delay with only 0.23-0.81 s and 0.47-1.09 s, respectively. Moreover, compared to the time consumed for manually adjusting the exoscope system (usually several seconds), the time delay of the proposed robot-assisted autonomous exoscope system is negligible, highlighting its efficiency.

C. Wire Chaser Task Results

The Wilcox signed-rank test was adopted to analyse the performance metrics (Section III.C) of the subjects with statistical significance set at $p < 0.05$.

For the execution time, a significant difference was found among the three repetitions when comparing the camera control modalities, AM_T, MM_T and AM_R, FM_R, as shown in Fig. 6. A mean value of 46.97 ± 14.98 s and

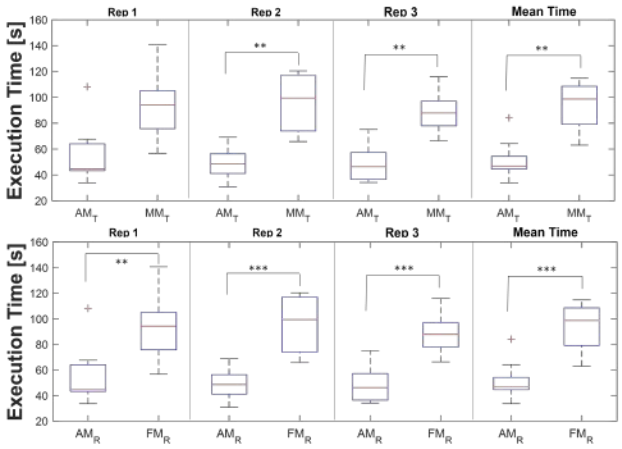

Fig. 6. Execution time for the three repetitions and the mean total execution time. Top: comparison between AM_T and MM_T; Bottom: comparison between AM_R and FM_R. ($*$, $p < 0.05$; $**$, $p < 0.01$; $***$, $p < 0.001$)

98.62 ± 18.18 s of the total execution time was observed for the autonomous and manual modality, respectively. For the autonomous orientation controller and the foot remote control a mean value of 53.47 ± 18.28 s and 93.10 ± 19.82 s was found respectively. The shorter execution time in autonomous modes compared to manual and foot-operated modes is due to users focusing only on the main task while the robotic manipulator handles camera repositioning. Since a long operation time is associated with a high workload, the ability to shorten task duration leads to a lower mental workload [27]. The time out FoV resulted significantly lower (p-value < 0.01) in the autonomous modalities with a mean value of 12.43 ± 22.83 s and 79.30 ± 138.28 s for AM_T and MM_T and 31.32 ± 57.14 s and 121.15 ± 131.92 s for AM_R and FM_R respectively. Again, the results were due to perform a dual task when the manual and foot joystick modes were used. Therefore, each time the position of the robotic camera holder needed to be adjusted, the user's focus was diverted from the main task to the motion of the camera, causing the instrument to leave the field of view. In a real surgical scenario, the risk of complications may increase if the instruments are not visible through the exoscope, as the possibility of inadvertent contact between instruments and delicate structures increases [28]. As for the path length,

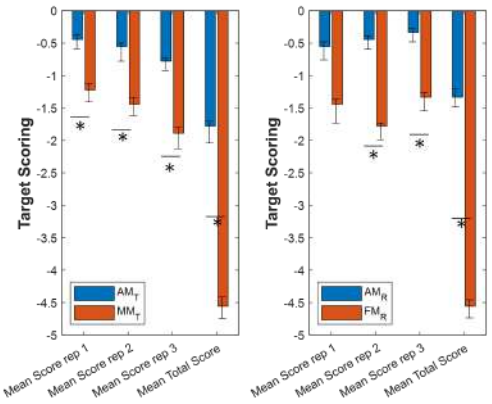

Fig. 7. Comparison of the score obtained during task execution. Left: autonomous and manual control modality comparison; Right: autonomous and foot remote control modality comparison. The "mean total score" item represents the results from all three repetitions. (*, $p < 0.05$)

a significant difference was found ($p - value < 0.001$) when comparing the AM_T mode with the MM_T mode. In Fig. 7 bar plots related to the target scoring for each compared modality are shown. Overall, the performance in terms of targeting score and path length was higher with the autonomous control modality. The drop in performance with the MM_T and FM_R can be attributed to the high mental and physical demands of the systems, as they negatively affect performance due to high cognitive load [29]. Finally,

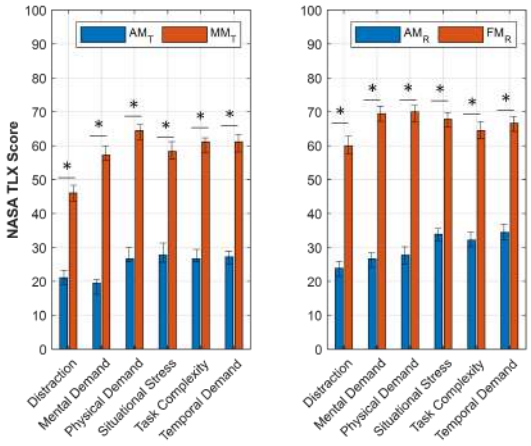

Fig. 8. NASA TLX score. Left: comparison between autonomous and manual control modalities. Right: comparison between autonomous and foot remote control modalities. (*, $p < 0.05$)

the qualitative analysis was conducted using the NASA-TLX survey, with users rating perceived workload using six subscales (Distraction, Mental Demand, Physical Demand, Situational Stress) for all modalities (Fig. 8). A significant difference can be observed between control modalities in all subsections. Users reported lower workload with autonomous control compared to manual and foot remote control, suggesting the proposed system can reduce mental and physical demands, leading to better performance. The lower scores in all the subsections suggest that the system may lead to better performance compared to the traditional control modalities. This is because the high workload and stress can negatively affect the users' decision-making abilities and motor skills, potentially leading to errors during the surgical procedure. Therefore, by reducing the workload and stress of the surgeon, the proposed system may improve surgical outcomes by enabling more accurate and efficient task execution.

V. Conclusion

In this work, we present an autonomous camera control system for robot-assisted neurosurgery that uses an image-analysis based method for instrument localisation. Experimental results have verified that the system can detect and track the moving surgical tool, ensuring convergence of the error to zero. The results of the user study have shown that the proposed framework improves the user experience by providing lower execution time, higher performance, and reduced workload. In future work, the proposed system should be evaluated under more realistic conditions. A more realistic dataset with real clinical images should be considered. Also, a task representative of the clinical scenario will be investigated in the following work.

References

[1] A. Lavé, R. Gondar, A.K. Demetriades, T. R. Meling, Ergonomics and musculoskeletal disorders in neurosurgery: a systematic review, Acta Neurochir, vol. 162, pp. 2213–2220, 2020.

[2] N. Montemurro, A. Scerrati, L- Ricciardi, G.Trevisi, Gianluca, The Exoscope in Neurosurgery: An Overview of the Current Literature of Intraoperative Use in Brain and Spine Surgery, Journal of Clinical Medicine, vol. 11, no. 1, pp. 223, December 2021

[3] F. I. Ahmad, A. F. Mericli, M.V. DeFazio, E. I. Chang, M. M. Hanasono, W. C. Pederson, M. Kaufman, J. C. Selber, Application of the ORBEYE three-dimensional exoscope for microsurgical procedures, Microsurgery, vol. 40, no. 4, pp. 468-472, 2020.

[4] T. Beez, C. Munoz-Bendix, K. Beseoglu,HJ. Steiger, SA. Ahmadi, First Clinical Applications of a High-Definition Three-Dimensional Exoscope in Pediatric Neurosurgery, Cureus, vol. 10, no.1, 2018.

[5] B. Fiani, R. Jarrah, F. Griepp, and J. Adukuzhiyil, The role of 3d exoscope systems in neurosurgery: An optical innovation, Cureus, vol. 13, no. 6, 2021.

[6] H.E. Douglas, M.Z. Raban, S.R Walter, J.I Westbrook, Improving our understanding of multi-tasking in healthcare: Drawing together the cognitive psychology and healthcare literature. Appl Ergon., vol. 59(Pt A), pp. 45-55., 2017

[7] A. Pandya, L.A. Reisner, B. King, N. Lucas, A. Composto, M. Klein, R.D. Ellis, A Review of Camera Viewpoint Automation in Robotic and Laparoscopic Surgery. Robotics, vol. 3, no. 3, pp. 310-329, 2014.

[8] T. Haidegger, Autonomy for Surgical Robots: Concepts and Paradigms, IEEE Transactions on Medical Robotics and Bionics, vol. 1, no. 2, pp. 65-76, 2019.

[9] S. Muhammad, M. Lehecka, N. Niemelä, Preliminary experience with a digital robotic exoscope in cranial and spinal surgery: a review of the Synaptive Modus V system, Acta Neurochir, vol. 161, pp. 2175–2180, 2019

[10] C. Molnár, T. D. Nagy, R. N. Elek, T. Haidegger, Visual servoing-based camera control for the da Vinci Surgical System, IEEE 18th International Symposium on Intelligent Systems and Informatics (SISY), pp. 107-112, 2020

[11] M. Chmarra, C. Grimbergen, J. Dankelman, Systems for tracking minimally invasive surgical instruments, Minimally Invasive Therapy Allied Technologies, vol. 16, pp. 328–40, 2007

[12] T. Da Col, G. Caccianiga, M. Catellani, A. Mariani, M. Ferro, G. Cordima, E. De Momi, G. Ferrigno, O. de Cobelli, Automating endoscope motion in robotic surgery: A usability study on da vinci-assisted ex vivo neobladder reconstruction, Frontiers in Robotics and AI, vol. 8, 2021

[13] I. Avellino, G. Bailly, M. Arico, G. Morel, and G. Canlorbe, Multimodal and Mixed Control of Robotic Endoscopes, in Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems, pp. 1–14, 2020

[14] A. Al-Shanoon, H. Lang, Robotic manipulation based on 3-D visual servoing and deep neural networks, Robotics and Autonomous Systems, vol. 152, 2022

[15] D. Bouget, M. Allan, D. Stoyanov, P. Jannin, Vision-based and marker-less surgical tool detection and tracking: a review of the literature, Med Image Anal., vol. 35, pp.633-654, 2017

[16] C. Gruijthuijsen, L. C. Garcia-Peraza-Herrera, G. Borghesan, D. Reynaerts, J. Deprest, S. Ourselin, T. Vercauteren, and E. Vander Poorten, Robotic endoscope control via autonomous instrument tracking, Frontiers in Robotics and AI, vol. 9, 2022

[17] E. Iovene et al., Towards Exoscope Automation in Neurosurgery: A Markerless Visual-Servoing Approach, IEEE Transactions on Medical Robotics and Bionics, vol. 5, no. 2, pp. 411-420, May 2023, doi: 10.1109/TMRB.2023.3258524.

[18] P.J.M. Wijsman, I.A.M.J Broeders, H.J. Brenkman, et al., First experience with THE AUTOLAP™ SYSTEM: an image-based robotic camera steering device, Surg Endosc, vol. 32, pp. 2560–2566, 2018

[19] Ultralytics-Yolov5, Available online: https://github.com/ultralytics/yolov5 (accessed on 1 Januray 2021)

[20] J. Gedge, M. Gong and Y. -H. Yang, Refractive Epipolar Geometry for Underwater Stereo Matching, Canadian Conference on Computer and Robot Vision, pp. 146-152, 2011

[21] J. Fu et al., Augmented Reality-Assisted Robot Learning Framework for Minimally Invasive Surgery Task, IEEE International Conference on Robotics and Automation (ICRA), London, United Kingdom, pp. 11647-11653, 2023

[22] M. Allan, A. Shvets, T. Kurmann, Z. Zhang, R. Duggal, Y.-H. Su, N. Rieke, I. Laina, N. Kalavakonda, S. Bodenstedt, L. Herrera, W. Li, V. Iglovikov, H. Luo, J. Yang, D. Stoyanov, L. Maier-Hein, S. Speidel, and M. Azizian, 2017 robotic instrument segmentation challenge, 2019

[23] T.D. Nagy, T. Haidegger, Performance and Capability Assessment in Surgical Subtask Automation, Sensors, 2022

[24] M. A., Aghazadeh, M., Mercado, M., Pan, B. Miles, A., Goh. Performance of robotic simulated skills tasks is positively associated with clinical robotic surgical performance. BJU International. vol. 118. 2016

[25] R. D., Dias, M. C., Ngo-Howard, M.T., Boskovski, M.A., Zenati, S.J., Yule. Systematic review of measurement tools to assess surgeons' intraoperative cognitive workload., Br J Surg., vol. 105, no. 5, pp. 491-501, 2018

[26] S. G. Hart, Nasa-task load index (nasa-tlx); 20 years later, Human Factors and Ergonomics Society Annual Meeting, vol. 50, no. 9, pp. 904–908, 2006

[27] J. Wnag, J. Cabrera, K. L. Tsui, H. Guo, M. Bakker, J. B. Kostis, Clinical and Nonclinical Effects on Operative Duration: Evidence from a Database on Thoracic Surgery, Journal of Healthcare Engineering, vol. 2020, 2020.

[28] A. Krupa, M. De Mathelin, C. Doignon, J. Gangloff, G. Morel, L. Sole, J. Leroy, J. Marescaux, Automatic positioning of surgical instruments during laparoscopic surgery with robots using automatic visual feedback, ESAIM: Proceedings, vol. 12, pp. 75-83, 2002

[29] D. Yu, B. Lowndes, C. Thiels, et al., Quantifying Intraoperative Workloads Across the Surgical Team Roles: Room for Better Balance?. World J Surg., vol. 40 no. 7, pp 1565-1574, 2016