Envisioning Robotic Exoscope: Concept and Preliminary Results

Alice Valeria Iordache[1, *], Alessandro Casella[1, 2, *], Elisa Iovene[1], Junling Fu[1], Federico Pessina[3], Marco Riva[3], Giancarlo Ferrigno[1], Leonardo S. Mattos[2], and Elena De Momi[1]

[1]*Department of Electronics, Information and Bioengineering, Politecnico di Milano, Milan, Italy*
[2]*Department of Advanced Robotics, Istituto Italiano di Tecnologia, Genoa, Italy*
[3]*Humanitas Research Hospital, Milan, Italy*
[*]*These authors contributed equally.*

INTRODUCTION

The introduction of a surgical microscope in neurosurgery increased the spectrum and safety of interventions. The surgeon operates with both hands on the patient while also controlling the visualization system to provide the proper view. Despite the benefits of the microscope, low ergonomics may lead to long-term effects on the musculoskeletal system. The bulky binocular system may limit the setup's flexibility, hindering robotics integration into the operating theatre.

Recently, the introduction of the exoscopes overcame those limitations by enabling a flexible and ergonomic working environment, combined with improved image quality. The exoscope is fixed on the surgical field through pneumatic arm holders that the surgeon can move manually or through foot control. Although providing the surgeon with enhanced vision, the complexity of its use has emerged as a limiting factor [1]. The continuous switching from operation theatre and visualization system reduces the smoothness of the surgical procedure, and the steep learning curve of the foot pedal [2] leads to switching to the conventional microscope in most cases. Visual servoing techniques [3] have been primarily studied in the context of camera automation in minimally invasive laparoscopy [4]. Its potential has become even more significant in the context of neurosurgery, as the only solution for tracking instruments is to employ vision sensors. However, visual servoing techniques for exoscope automation are currently limited due to open challenges in neurosurgical practices. This work proposes a framework for an autonomous vision-guided camera holder that tracks and follows a selected surgical instrument based on a visual servoing technique. We envision this solution to automate the exoscope in neurosurgery, providing intuitive control of the system and reducing surgeon workload and operating time.

MATERIALS AND METHODS

A. Visual Servoing Framework

In this study, a markerless position-based visual servoing framework was implemented as illustrated in Fig. 1. A 7 DOF serial manipulator (KUKA LwR 4+) was used with a stereo camera mounted in an eye-in-hand configuration to simulate the exoscope system. An object-detection

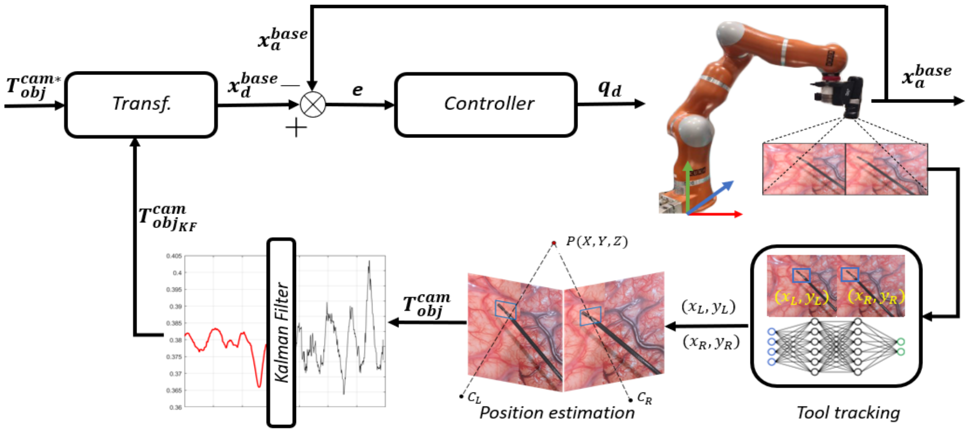

Fig. 1 System's overview: tool coordinates (x_R, y_R) and (x_L, y_L) are identified in the image space. The position $P(X, Y, Z)$ in the 3D space is estimated and filtered by the Kalman Filter $\boldsymbol{T_{obj_{KF}}^{cam}}$. This, together with the desired position $\boldsymbol{T_{obj}^{cam*}}$, are transformed and used to compute the desired position of the camera $\boldsymbol{x_d^{base}}$. The desired and actual position of the camera, $\boldsymbol{x_a^{base}}$, are used to calculate the error signal $\boldsymbol{e}$. The controller computes the required joint positions $\boldsymbol{q_d}$ that compensate for the error.

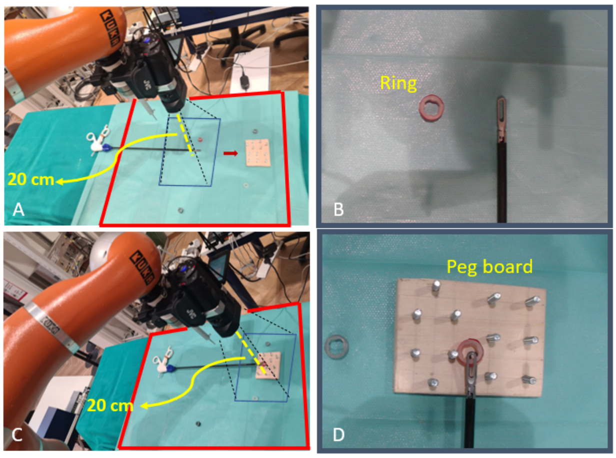

Fig. 2 Design of the task, only one ring is seen by the camera, camera need to be moved to reach the pegboard.

neural network *Yolov3*[5] was trained to identify the tip of a surgical instrument from the images acquired by the 3D camera. The training dataset was composed of 5000 images, 1800 were extracted from the *EndoVis 2017 challenge dataset* [6] while the remaining were recorded and manually annotated.

The position reconstruction was computed by triangulation using DLT, while the Kalman filter was applied, considering a three-dimensional constant velocity model of the instrument's motion. The desired position of the instrument with respect to the camera $\boldsymbol{T}_{\boldsymbol{obj}}^{\boldsymbol{cam*}}$ was defined in such a way that the instrument is kept at the center of the image. Only motions along the XY-plane have been implemented since, in neurosurgery, movement along the depth is only required for focus adjustments. Finally, a position controller was employed to compensate for the error between the desired and actual position of the camera.

B. *Experimental protocol*

In order to test the usability of the system, a user study was carried out on ten non-medical subjects. The designed task was a pick-and-place in which users were asked to use a surgical instrument to pick up, one at a time, four randomly distributed rings in a defined workspace and place them on a target pegboard, as shown in Fig. 2.

The distance in the Z direction between the camera and the task space was $0.2m$ and was kept fixed to provide a reduced Field of View (FoV) and force the user to move the camera to complete the task. Users were asked to perform the task exclusively by observing the scene on an external monitor where the camera FoV was displayed.

The task was executed in two different modalities:

1) *Autonomous Camera Control* (*ACC*): the user was able to activate and deactivate the autonomous motion of the camera by pressing a foot pedal.
2) *Joystick Control* (*JC*): the user was able to move the camera using a foot-controlled joystick each time a different view of the scene was needed.

Three repetitions were performed for both modalities.

RESULTS AND DISCUSSION

The systems' functionality was evaluated in terms of tool's position density inside the image plane and execution time. Fig. 3 shows that in *ACC*, the density is higher at the center of the image than in *JC*. This proves that the *ACC* is more effective at keeping the instrument in the center of the image, providing a better view of the scene, which is crucial for optimal task performance. As for the execution time, *ACC* showed better results, as illustrated in Table I. The mean and standard deviation of the execution time for the three repetitions are reported.

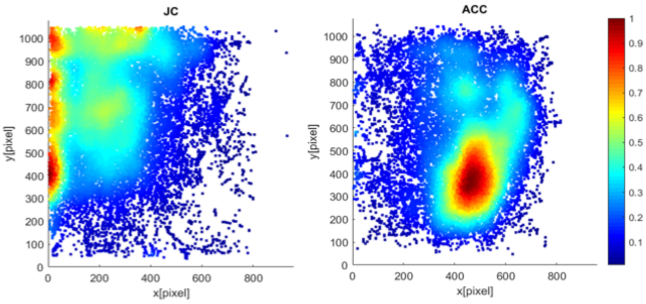

Fig. 3 Normalized instrument's tip density inside the image plane for the *JC* and *ACC* modalities

TABLE I Mean execution time, *std* and p-value for the three repetitions of the *JC* and *ACC* modalities.

Repetition	1		2		3	
Modality	*JC*	*ACC*	*JC*	*ACC*	*JC*	*ACC*
Exec. time[s]	112 ±11.8	82.6 ±24.6	105.3 ±14.4	80.2 ±26.2	104.9 ±18.1	73.1 ±18.5
p-value	**		**		**	

This demonstrates that *ACC* reduces completion time by allowing the user to focus only on the main task and consequently perform better. Statistical analysis was conducted for all repetitions using the Wilcoxon signed-rank test with a 5% level of significance.

CONCLUSIONS AND DISCUSSION

Preliminary results have shown that the proposed markerless visual servoing system has significant advantages over the currently used system. However, translating this system into operating room requires further improvements. Future studies will focus on the improvement of the tracking and control module.

REFERENCES

[1] Mamelak, et al. A high-definition exoscope system for neurosurgery and other microsurgical disciplines: Preliminary report. *Surgical Innovation*, 15(1):38–46, 2008. PMID: 18388000.
[2] Maurer, et al. Evaluation of a novel three-dimensional robotic digital microscope (aeos) in neurosurgery. *Cancers*, 13(17), 2021.
[3] Chaumette and Hutchinson. Visual servo control. i. basic approaches. *Robotics Automation Magazine, IEEE*, 13:82 – 90, 01 2007.
[4] Gruijthuijsen, et al. Robotic endoscope control via autonomous instrument tracking, 2021.
[5] Redmon and Farhadi. Yolov3: An incremental improvement. 04 2018.
[6] Allan, et al. 2017 robotic instrument segmentation challenge, 2019.