POLITECNICO
MILANO 1863

SCUOLA DI INGEGNERIA INDUSTRIALE
E DELL'INFORMAZIONE

Executive Summary of the Thesis

Towards Real-Time Tool Tracking for Autonomous Exoscope Control

Laurea Magistrale in Automation and Control Engineering - Ingegneria dell'Automazione

Author: Diego Cattaneo

Advisor: Prof. Elena De Momi

Co-advisor: Elisa Iovene

Academic year: 2022-2023

1. Introduction

Exoscopes represent a promising visual solution in the neurosurgical field aimed at offering an improved field of view and ergonomics to the surgeons compared with traditional surgical microscopes. However, manual repositioning in existing models can interrupt surgery, leading to longer operation times and reduced efficiency [1]. To further reduce the surgeon's workload, minimizing the need for direct intervention in camera control is desirable. Various techniques have been explored for automatizing the camera motion. Among these, the markerless instrument tracking approach stands out as one of the most widely used methods. It enables fast and precise reconstruction of the surgical instrument's 3D position, seamlessly integrating into robotic control frameworks . This approach has been successfully implemented in multiple camera systems, ensuring smooth and controlled movements [2].

This thesis work presents an innovative hybrid tracking module designed to guarantee real-time tool tracking through a robot-assisted autonomous exoscope.

2. Materials And Methods

The proposed system is divided into three modules: a tool detection module (Sect. 2.1) that can recognize a selected surgical instrument, a hybrid tracking module that tracks and predicts the future position of the target tool (Sect. 2.2), and a visual-servoing controller responsible for zeroing the error between the desired and the actual pose of the robot (Sect. 2.3). The overall system is illustrated in Figure 1.

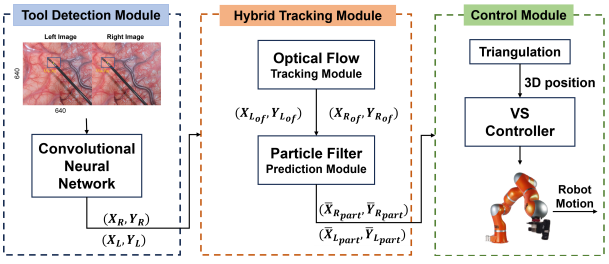

Figure 1: Overall System: images acquired by the stereo camera are sent to a CNN which detects the surgical tool. The 2D position of the tool is sent to a tracking module, then a prediction module estimates the future position of the tool in the image space. Finally, the 3D position of the tool is extracted from the predicted position and is fed to a visual-servoing controller.

2.1. Tool Detection Module

A pre-trained CNN YoloV5 [3] was fine-tuned for target detection. The CNN received downsampled RGB images of size 640x640 from a stereo camera and provided the position of the instrument's tip through a bounding box. The center of the bounding box in the right and left images, (X_R, Y_R), (X_L, Y_L), denoted the position of the instrument in the camera space.

2.2. Hybrid Tracking Module

To address the CNN's low-speed performance in tool detection, a hybrid tracking module was introduced. The module consisted of two key components: the Optical Flow (OF) tracking and a modified particle filter. By leveraging the OF, we achieved efficient tool tracking between consecutive frames. Additionally, the particle filter played a crucial role in predicting the tool's future position, effectively reducing system delays. The OF tracking module offered by OpenCV [4], which exploits the Lukas-Kanade method with pyramids, was chosen for this study. In this work, the position of the instrument in the camera space, (X_R, Y_R), (X_L, Y_L), was sent to the OF together with eight surrounding points to lower the risk of losing the tool position because of partial occlusions or small changes of the pixel intensities. The tool position was thus detected using the CNN on the first frame and then the OF was used to track the tool's position, $(X_{R_{of}}, Y_{R_{of}})$, $(X_{L_{of}}, Y_{L_{of}})$, in the following, with exceptions made when the OF lost the points to follow or appeared to be tracking the wrong point in the image. The CNN was used also every 15 seconds to confirm the precise tracking of the intended target.

A modified particle filter was introduced to get an estimate of the future tool position in the image space, on the basis of the previous position and of the speed and orientation of motion. The particle filter allows to effectively handle complex and non-linear systems such as the motion of a surgical instrument. Every time the tool position, $(X_{R_{of}}, Y_{R_{of}})$, $(X_{L_{of}}, Y_{L_{of}})$, was computed, our particle filter acted as follows:

1. The direction, h_t, and the speed, v_t, of the movement of the tool were computed.
2. The heading of each particle, h_i, was distributed normally around the direction of the movement of the tool, h_t.
3. The future position of the particles was predicted in both images. In particular, if the modulus of the variation of the direction of the tool, $|\Delta_h|$, was under a certain threshold, the Runge–Kutta odometry was used:
$$X_{R,i} = x_t + v_t \cdot dt \cdot \cos\left(h_i + \frac{\Delta_h \cdot dt}{2}\right)$$
$$Y_{R,i} = y_t + v_t \cdot dt \cdot \sin\left(h_i + \frac{\Delta_h \cdot dt}{2}\right)$$
where h_i is the heading of the i^{th} particle, dt is the prediction horizon, x_t and y_t are the x and y coordinate of the tool respectively, coming from the tracking module, v_t is the estimated velocity of the tool, and i is the number of the particle. When $|\Delta_h|$ was above a certain threshold the exact odometry was used:
$$X_{R,i} = x_t + \frac{v_t[\sin(h_i + \Delta_h \cdot dt) - \sin(h_i)]}{\Delta_h}$$
$$Y_{R,i} = y_t - \frac{v_t[\cos(h_i + \Delta_h \cdot dt) - \cos(h_i)]}{\Delta_h}$$
For simplicity, we provided the equation for the right frame; however, the same holds true for the left frame as well.
4. The weighted average of the particles' positions was computed to determine the future position of the tool:
$$\overline{X}_{R_{part}} = \frac{\sum_{i=1}^{N} w_{R,i} \cdot X_{R,i}}{\sum_{i=1}^{N} w_{R,i}}$$
$$\overline{Y}_{R_{part}} = \frac{\sum_{i=1}^{N} w_{R,i} \cdot Y_{R,i}}{\sum_{i=1}^{N} w_{R,i}}$$
with N number of particles and $w_{R,i}$ weight of the i^{th} particle.
5. The weight of the particles was updated as follows:
$$w_{R,i} = \frac{\max(dist_i) - dist_i}{\sum_{i=1}^{N}(\max(dist_i) - dist_i)}$$
where i indicates the i^{th} particle and $dist_i$ takes into account the Euclidean distance between the predicted particle position and both the actual tool position and a potential future position of the tool. This additional term was intended to assign more weight to the direction of the motion being tracked.

The predicted positions in both images $(\overline{X}_{R_{part}}, \overline{Y}_{R_{part}})$, $(\overline{X}_{L_{part}}, \overline{Y}_{L_{part}})$ were used to extract by triangulation the 3D position of the tool that was then sent to the robot controller.

2.3. Robot Control Module

The 3D predicted position of the tool is sent to a visual servoing controller. In this work, the type of camera motions taken into account were two:

- Translational motion with fixed orientation (**Position Control**);
- Rotational motion around a fixed point (**Orientation Control**).

The considered frames are illustrated in Figure 2.

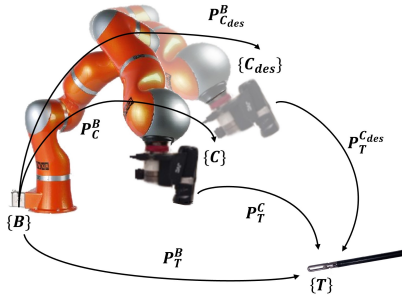

Figure 2: Positions and reference frames. $\{C\}$ indicates the camera reference frame, $\{B\}$ the robot base reference frame, and $\{T\}$ the tool reference frame. $\boldsymbol{P}_T^C$ and $\boldsymbol{P}_T^{C_{des}}$ are the positions of the surgical tool in the actual and desired camera space, respectively. $\boldsymbol{P}_T^B$ is the position of the tool, $\boldsymbol{P}_C^B$ is the actual position of the camera, and $\boldsymbol{P}_{C_{des}}^B$ is the desired position of the camera, all with respect to the robot's base reference frame.

2.3.1 Position Control

In this case, since the goal was to keep the instrument near the center of the camera image with a distance d, the position of the tool with respect to the desired position of the camera was $\boldsymbol{P}_T^{C_{des}} = \begin{bmatrix} 0 & 0 & d & 1 \end{bmatrix}$. So, the desired position of the camera was obtained as:

$$\boldsymbol{P}_{C_{des}}^B = \boldsymbol{P}_T^B (\boldsymbol{P}_T^{C_{des}})^{-1}$$

and the Position error was thus given by:

$$\boldsymbol{e}_{pos} = \boldsymbol{P}_{C_{des}}^B - \boldsymbol{P}_C^B$$

2.3.2 Orientation Control

When the Orientation strategy was used, the concept of remote center of motion [5] was exploited to compute the desired orientation of the camera, $\boldsymbol{R}_{Cdes}^B$.

From $\boldsymbol{R}_{Cdes}^B$, the Orientation error was computed as follows:

$$\boldsymbol{e}_{or} = \boldsymbol{q}_{orC} \cdot \boldsymbol{q}_{orC_{des}}^{-1}$$

where $\boldsymbol{q}_{orC}$ and $\boldsymbol{q}_{orC_{des}}$ were the current and desired quaternions.

The feedback errors $\boldsymbol{e}_{pos}$ or $\boldsymbol{e}_{or}$ were then fed into a resolved-velocity controller [6] which assumes that the manipulator acts as an ideal positioning device and that computes the vector of joints' velocity profiles as follows:

$$\dot{\boldsymbol{q}}_{des}(t) = \boldsymbol{J}_{A_W}^{\#}(\boldsymbol{q}) \boldsymbol{K} \boldsymbol{e} \tag{1}$$

where $\dot{\boldsymbol{q}}_{des}(t)$ is the vector of desired joints' velocity profiles, $\boldsymbol{J}_{A_W}^{\#}(\boldsymbol{q})$ is the weighted pseudo-inverse of the analytical Jacobian of the robot, function of the joints' position, $\boldsymbol{K}$ is a positive-definite gains matrix and $\boldsymbol{e}$ is the vector of the feedback errors. Knowing that $\dot{\boldsymbol{q}}_{des} \approx \dot{\boldsymbol{q}}$, thanks to the assumption of ideal positioning device, and that $\dot{\boldsymbol{e}} = -\boldsymbol{J}_A(\boldsymbol{q})\dot{\boldsymbol{q}}$, the equation 1 becomes:

$$\dot{\boldsymbol{e}} + \boldsymbol{K}\boldsymbol{e} = 0$$

This equality, for a positive definite matrix $\boldsymbol{K}$, allows demonstrating through the Lyapunov method that the equilibrium $\boldsymbol{e} = 0$ is globally asymptotically stable.

3. Experimental Setup

To simulate the exoscope system, a 7-Degrees-of-Freedom (DoF) redundant robotic manipulator (LWR 4+ lightweight robot, KUKA, Germany) with an eye-in-hand stereo camera configuration (JVC GS-TD1 Full HD 3D Camcorder) was used. Moreover, to validate the developed strategies and fine-tune the controller a second 7-DoF redundant robotic manipulator (LBR IIWA lightweight robot, KUKA, Germany) was considered to move the surgical tool with high repeatability, as shown in Figure 3.

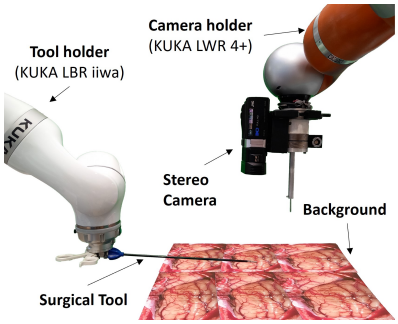

Figure 3: Experimental setup of validation tests

Then, a User Study was organized to evaluate whether the developed autonomous exoscope was more effective than the traditional strategy in reducing the user workload.

3.1. System Validation

The performance of the tracking and the control module was investigated in relation to the target's velocity, representing the surgical instrument's movement. To validate the effectiveness of the hybrid tracking module, an analysis that compared the system with (Hybr) and without (CNN) [2] this module, was conducted. In this validation phase, only the Position Control was used. The strategies were tested within two different velocity scenarios to study the robustness of the system against different conditions. The two velocities were chosen on the basis of a study carried out with neurosurgeons about the typical speeds reached during brain surgeries: the system was thus tested with the tool moving at about 2.5 cm/s (low speed) and about 4 cm/s (high speed). The camera had to follow the tool that was moved in a constant trajectory described in Figure 4.

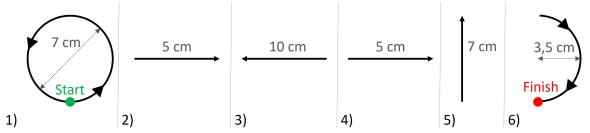

Figure 4: Trajectory travelled by the tool during the experiments

All the tests were repeated five times for each strategy and for every scenario. During these tests, the robot controller was based on a proportional gain equal to 1.
The performance analyzed indexes are the tracking error (TE_{xy}), defined as the distance between the camera and the tool:

$$TE_{xy} = ||\boldsymbol{P}_C^B - \boldsymbol{P}_T^B|| \ [mm]$$

and the center error (CE) defined as the distance between the tool position in the camera frame and the center of the image (C_x, C_y):

$$CE = ||X_t - C_x, Y_t - C_y|| \ [mm]$$

After the validation phase, the robot position and Orientation Controllers were fine-tuned, resulting in the selection of the following proportional gain for the Position Control:

$$\boldsymbol{K}_{pos} = \begin{bmatrix} 4.0 & 0.0 & 0.0 \\ 0.0 & 4.0 & 0.0 \\ 0.0 & 0.0 & 1.0 \end{bmatrix} \quad \boldsymbol{K}_{rot} = 1.0 \cdot \boldsymbol{I}_3$$

and for the Orientation Control:

$$\boldsymbol{K}_{pos} = 0.0 \cdot \boldsymbol{I}_3 \quad \boldsymbol{K}_{rot} = 0.7 \cdot \boldsymbol{I}_3$$

where $\boldsymbol{I}_3$ is the 3×3 identity matrix.
Following that, the system was further tested using the updated parameter in the same experimental setups (in Position and Orientation Control mode). In the Orientation mode, instead of the above described TE_{xy}, the tracking errors TE_r and TE_p, measuring the distance between the desired and actual camera angles, were considered:

$$TE_r = \min(|e_{roll}|, \ 2\pi - |e_{roll}|) \ [deg]$$

where $e_{roll} = roll_{cam} - roll_{des}$ is the difference between the camera's actual roll angle, $roll_{cam}$, and the camera's desired roll angle, $roll_{des}$. The tracking error of the pitch angle, TE_p, was computed in the same way, while it was not calculated for the yaw angle since it was kept fixed.

3.2. User Study

During the User Study, 12 users employed three camera control modes (manual control, Position and Orientation autonomous control) while executing a bimanual pick and place task of a hidden object in the setup shown in Figure 5.

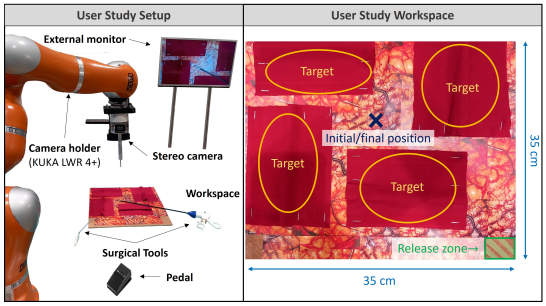

Figure 5: User Study Setup (left): everything the users employed during tests; the robot to move the camera, the stereo camera, the external monitor to look at the scene, the workspace where to accomplish the task, two surgical tools and a pedal, to decide when to move the robot in the automatic control modes. The User Study Workspace (right): it was divided into four different targets, a release zone, and the initial/final position of the surgical instrument.

All the users provided informed consent before participating. The experimental protocol was approved by the ethics committee from Politecnico di Milano, Italy (No.2023-5069). The tests started with the tools in the initial position. Then, the users had to look for a plastic ring hidden under one of the four targets and, once they found it, they had to grasp it with the surgical instrument and release it in the release zone. To conclude the test, they had to take back the tools to the initial position. During the tests, the users were asked to look only at the external monitor showing the images coming from the stereo camera and to keep both the surgical tools in the center of the image. While users were employing the automatic modes, to decide when to move the robot, they had to press a pedal. They were required to accomplish the task using all the camera control modes doing three repetitions for each strategy. To avoid that an eventual learning curve could affect the performance, the order of the strategies followed by each user was chosen from the set of permutations of the three strategies.
The performance indexes evaluated during the tests are:

- A score assigned on the basis of the distance, d, of the tool from the image center:

$$d = ||\boldsymbol{P}_t - \boldsymbol{P}_c||$$

  where $\boldsymbol{P}_t$ and $\boldsymbol{P}_c$ are the positions of the tool and the image center respectively. If d was greater than 5 cm, -1 point was assigned, while if the tool was outside the field of view of the camera, -5 points were assigned. The score was normalized before the data analysis.
- The duration of the movements between two consecutive targets;
- The path length travelled by the tool between two consecutive targets. The length of each movement is computed as:

$$l(M) = \sum_k ||\boldsymbol{P}_t(k) - \boldsymbol{P}_t(k-1)||$$

  where $\boldsymbol{P}_t(k)$ is the position of the tool at the k^{th} instant.
- A NASA Task Load Index survey for each control strategy;
- A questionnaire.

4. Results & Discussion

4.1. System Validation

The performance indexes, TE_{xy} and CE, for the two strategies in the two different scenarios, can be appreciated in Figure 6.

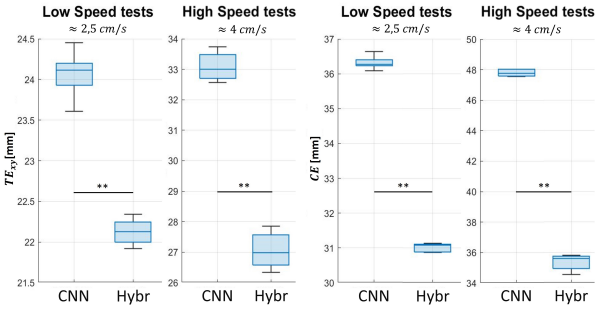

Figure 6: Tracking Error (left) and Center Error (right). (**, p-value < 0.01)

The system with the hybrid tracking module consistently demonstrates the lowest tracking error in both the slow and fast scenarios. This indicates that the Hybr strategy is effective in accurately tracking the surgical instrument's movement compared to the other strategy. Additionally, the Hybr strategy displays low standard deviations in both tracking and center errors, showing its consistency and robustness in tracking and centering of the surgical instrument across varying scenarios.
The evaluation of different performance metrics was examined through the Wilcox signed-rank test, with statistical significance established at a threshold of $p < 0.05$. In Figure 6 can be noticed also a statistical difference in both tracking error and center error. This reinforces the observation that the Hybr strategy consistently delivers superior performance.
The performance of the Hybr strategy after the fine-tuning can be observed in table 1.

Table 1: Metrics' means and standard deviations

Metrics	Low speed	High speed
Position Control		
TE_{xy} [mm]	9.84 ± 0.08	13.11 ± 0.39
CE [mm]	16.14 ± 0.14	20.80 ± 0.16
Orientation Control		
TE_r [deg]	4.29 ± 0.06	5.68 ± 0.08
TE_p [deg]	4.63 ± 0.08	5.65 ± 0.08
CE [mm]	22.41 ± 0.39	27.62 ± 0.43

4.2. User Study

On the data measured during the User Study, a Kruskal - Wallis test was performed (with statistical significance established at a threshold of $p < 0.05$). After computing the average of the repetitions grouped by user and control strategy, the results reported in Figure 7 were obtained.

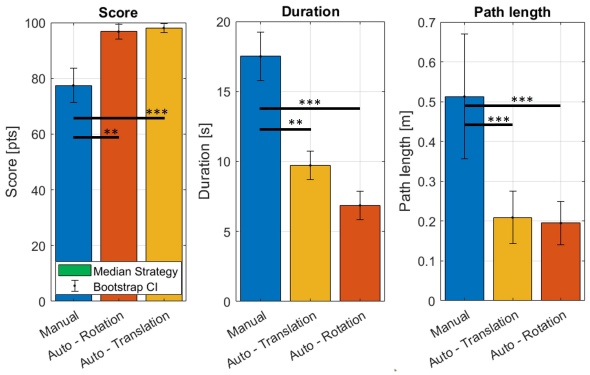

Figure 7: Score (left), Duration (center) and Path length (right). (**, p-value < 0.01; ***, p-value < 0.001;)

These data demonstrate that the automatic strategies allow to keep the tool at the image center more effectively than the manual control and they reduce the duration and the tool path length needed to accomplish the task, which is important in surgical operations that can last many hours. From the NASA - TLX (Figure 8) the automatic strategies result to be less frustrating and less mentally, physically and effort demanding with respect to the manual repositioning of the camera. In the end, the qualitative questionnaire highlights that users think that the automatic strategies facilitate more the task accomplishment (in particular the Position Control mode, that also guarantees the best field of view).

5. Conclusion

This study introduces a novel hybrid tracking module for a position-based visual-servoing control approach applied to a robotic camera holder. The integration of the OF tracking module and a particle filter, was introduced to further optimize system performance by predicting future tool positions. Overall, the tracking system follows the motion of the surgical tool with a relatively low tracking error and its use is appreciated by the users. In future work, the system should be evaluated in scenarios that better mimic the real world.

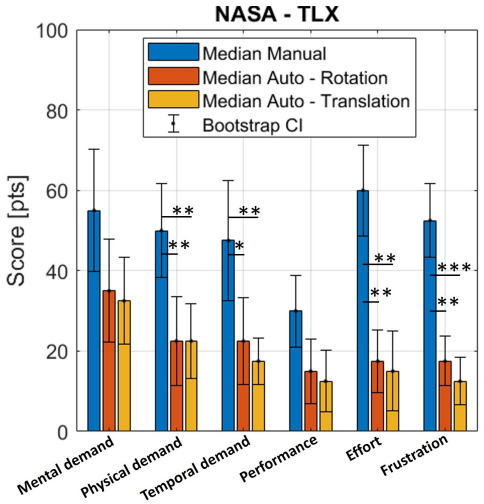

Figure 8: NASA - TLX results

References

[1] B. Fiani et al. The role of 3d exoscope systems in neurosurgery: An optical innovation. *Cureus*, 13(6), 2021.

[2] E. Iovene et al. Towards exoscope automation in neurosurgery: A markerless visual-servoing approach. *IEEE Transactions on Medical Robotics and Bionics*, 5, 05 2023.

[3] Ultralytics. Yolov5 by Ultralytics. `https://github.com/ultralytics/yolov5`.

[4] OpenCV. Optical Flow by OpenCV. `https://docs.opencv.org/3.4/d4/dee/tutorial_optical_flow.html`.

[5] J. Sandoval et al. Collaborative framework for robot-assisted minimally invasive surgery using a 7-dof anthropomorphic robot. *Robotics and Autonomous Systems*, 106:95–106, 2018.

[6] B. Siciliano et al. *Robotics - Modelling, Planning and Control*, pages 447–448. Springer-Verlag London Limited, 2009.