Shared Control Schemes for Middle Ear Surgery

Jae Hun So, Stéphane Sobucki, Jérôme Szewczyk, Naresh Marturi, Brahim Tamadazte

Shared Control Schemes for Middle Ear Surgery

Jae-Hun So[1]*, Stéphane Sobucki[1], Jérôme Szewczyk[1], Naresh Marturi[2] and Brahim Tamadazte[1]

[1]CNRS UMR 7222, INSERM U1150, ISIR, F-75005, Sorbonne Université, Paris, France, [2]Extreme Robotics Lab, University of Birmingham, Birmingham, United Kingdom

This paper deals with the control of a redundant cobot arm to accomplish peg-in-hole insertion tasks in the context of middle ear surgery. It mainly focuses on the development of two shared control laws that combine local measurements provided by position or force sensors with the globally observed visual information. We first investigate the two classical and well-established control modes, i.e., a position-based end-frame tele-operation controller and a comanipulation controller. Based on these two control architectures, we then propose a combination of visual feedback and position/force-based inputs in the same control scheme. In contrast to the conventional control designs where all degrees of freedom (DoF) are equally controlled, the proposed shared controllers allow teleoperation of linear/translational DoFs while the rotational ones are simultaneously handled by a vision-based controller. Such controllers reduce the task complexity, e.g., a complex peg-in-hole task is simplified for the operator to basic translations in the space where tool orientations are automatically controlled. Various experiments are conducted, using a 7-DoF robot arm equipped with a force/torque sensor and a camera, validating the proposed controllers in the context of simulating a minimally invasive surgical procedure. The obtained results in terms of accuracy, ergonomics and rapidity are discussed in this paper.

Keywords: medical robotics, tele-operation, comanipulation, visual servoing, robot control

1 INTRODUCTION

The tasks performed by robots are becoming more and more complex and are no longer limited to positioning or pick and place. A significant amount of progress has been made to increase the autonomy of robots, especially in certain industries like automotive, aerospace etc. However, several scientific and technological obstacles persist to reach a sufficient degree of autonomy to carry out more complex tasks, particularly in unknown, unstructured and dynamic environments. Modern-day robots are expected to share the workspace with humans or other robots in a safe and secure manner (Burgard et al., 2000; Roy and Dudek, 2001). To this extent, external sensors such as force torque sensors, cameras, lidars etc. play a prominent role.

For instance, force sensors provide local contact information while handling or grasping objects during insertion and assembling tasks; vision sensors such as cameras provide rich and global information (which can also be local in some cases) of the robot environment (Espiau et al., 1992; Chaumette et al., 2016). This brings up the question, whether it is possible to effectively combine several modes of perception (e.g., force and vision) within the same control scheme to benefit from their complementary advantages? Of course, this does not concern the sequential controllers that are usually reported in the literature (Baeten et al., 2003). In general, the robot is first controlled using the visual information provided by the camera(s) when the end-effector is far from the target, and then

switches to the proximity sensing (e.g., position or force), when the robot is close or in contact with the target. In most cases, it is trivial to combine measurements from different sensors such as position and force in the same control scheme (Morel et al., 1998; Mezouar et al., 2007; Prats et al., 2007; Liang et al., 2017; Adjigble et al., 2019). However, fusion of information, e.g., provided by the camera and the force sensor, is highly challenging due to the fact that these features are fundamentally different. Very few works have studied this possibility (Morel et al., 1998; Mezouar et al., 2007) and more recent works showed interest on this topic (Abi-Farraj et al., 2016; Selvaggio et al., 2018; Adjigble et al., 2019; Selvaggio et al., 2019).

This paper focuses on the development of two shared control laws, which combine proximal (local) and global measurements. Specifically, local measurements provided by position or force sensors are combined with globally observed visual information, which is acquired from a camera attached to the robot end-effector in an *eye-in-hand* setting. In our previous work, (So et al., 2020; So et al., 2022), have demonstrated how a robotic arm can be controlled in a fully and well-established position-based manner by means of tele-operation with a Phantom Omni system. There, we have developed an end-frame tele-operation controller. This approach has shown promising performances in terms of accuracy, repeatability and intuitiveness, particularly in the context of minimally invasive surgery. However, we have observed a strong coherence between the overall performance and operator's expertise levels.

Thus, in many cases, in particular if the operator is a surgeon, it turned out to be highly challenging to control all the degrees of freedom (DoF) of the surgical instrument (e.g., attached to robot's end-effector) merely using a joystick. Apparently, handling end-effector spatial rotations was particularly difficult for a surgeon who has to focus at the same time on the surgical procedure. Similar conclusions were drawn for the other control scheme presented in (So et al., 2020), which uses comanipulation mode of the robotic arm.

Addressing the aforementioned issues with our previous work, in this paper, we propose a new shared controller where a robot arm can be controlled either fully force-based or by using a shared control law that combines force and vision measurements in the same control loop. The operator, *i.e.*, a surgeon controls the linear (translational) axes of the tool via a human-robot interface while the rotations are automatically controlled by the vision-based controller. Thus, we first expressed two well-established controllers. The first one consists of a position-based end-frame tele-operation controller and the second one is a comanipulation controller. The main contributions of this paper lie in the formulation of new hybrid controllers by consistently combining two physical quantities in the control loop. This allows separating the linear DoF which are controlled in tele-operation (respectively, comanipulation) with the rotational ones that are controlled simultaneously and intuitively by a vision-based controller. To this extent, four shared control modes are studied and evaluated, which are listed as follows: 1) parallel hybrid tele-operation; 2) external hybrid tele-operation; 3) parallel hybrid comanipulation; and, external hybrid comanipulation. These are presented in detail in **Section 3**.

The developed controllers are experimentally validated using a 7-DoF robot arm, in the context of simulating a minimally invasive surgical procedure. We have conducted subjective analysis by recruiting two different groups of people to study and compare the performance of various controllers in terms of accuracy, behaviour, intuitiveness, and the time required to accomplish a target task. First group of people are the novice operators, who have never used the system; and, the second group consists of expert operators, who have developed the system and possess prior experience operating it. In this work, we compare the performance of conventional teleoperation and comanipulation controllers presented in (So et al., 2020) with our proposed methods. It has been demonstrated that the proposed shared controllers outperform the classical ones in multiple points, e.g., ergonomy, accuracy, etc.

The rest of this paper is organized as follows. In **Section 2**, we discuss the clinical framework of this work. **Section 3** deals with the methodology followed to carry out the proposed controllers as well as with the motivations that led us to develop these controllers. **Section 4** presents the implemented experimental setup to assess the different controllers and compare them to the classical tele-operation and comanipulation control modes. A discussion on the pros and cons of each control approach is also reported in this section.

2 CLINICAL MOTIVATIONS AND MIDDLE EAR SURGERY REQUIREMENTS

2.1 Motivation

The work discussed in this paper is part of a larger research and development investigation focusing on an ergonomic and intuitive robotic solutions for middle ear surgery (Dahroug et al., 2017; Dahroug et al., 2018; Dahroug et al., 2020; So et al., 2020). The primary challenge concerns the design and fabrication of a macro-and microscale robotic system, which consists of a conventional collaborative robotic arm fixed with a miniature continuum robot as the end-effector. The tool can enter and operate inside the middle ear through a millimetric-sized incision made in the mastoid. The secondary challenge is to develop intuitive and ergonomic control modes allowing the surgeon to perform several tasks during surgery such as bringing an instrument close to the surgical site, insert it into the middle ear cavity through an entry incision hole, position the surgical tool in front of pathological tissues for mechanical resection or laser burring process (Hamilton, 2005).

The typical surgical treatment we consider in this work involves the total removal of a pathological tissue that develops inside the middle ear cavity known as cholesteatoma (Holt, 1992). Cholesteatoma can be defined as an abnormal skin growth that occurs in the middle ear cavity, see **Figure 1**. It is usually due to blocked ventilation where dead skin cells cannot be ejected out of the ear. Gradually, the cholesteatoma expands in the middle ear, filling in the empty cavity around the ossicles and then eroding the bones themselves (ossicles and mastoid).
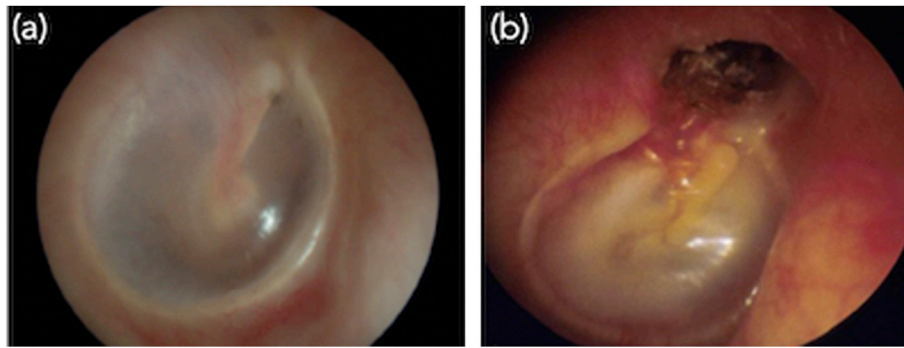
FIGURE 1 | Images of the internal of middle ear. **(A)** Normal ear and **(B)** typical primary acquired cholesteatoma, which destroyed the tympanic membrane.

Cholesteatoma is often infected and results in chronically draining ears. It also results in hearing loss and may even spread through the skull-base into the brain. It was reported that one in 10,000 citizens are diagnosed with cholesteatoma every year in the world (Olszewska et al., 2004; Møller et al., 2020).

Currently, the most effective treatment for cholesteatoma is to surgically resect or ablate the infected tissues (Derlacki and Clemis, 1965; Bordure et al., 2005). The current surgical procedure raises at least two major issues. The first is the degree of invasiveness required for such procedure. To get access to the middle ear cavity, the surgeon performs an incision hole of about 20 mm diameter in the mastoid. Through this, the surgeon visualizes the cholesteatoma as well as pass tools to treat the pathological tissues. The second limitation concerns the clinical efficiency of this procedure, which never ensures an exhaustive removal of the pathological tissues.

Moreover, there is not much literature on aggregate data (by region or country) on the recurrence of cholesteatoma with current surgical methods. However, it has been reported in studies conducted by health care institutions that depending on the surgeon expertise, approximately 25% of cholesteatoma treatments are unsuccessful leading to cholesteatoma persistence or recurrence (Aquino et al., 2011; Møller et al., 2020). These limitations of the current surgical procedures can be tackled by the design of the new micro-and macroscale robotic solution and the development of dedicated control architectures that will make the procedure less invasive and more efficient.

2.2 System Requirements

In one of our previous works (Dahroug et al., 2018), we have evaluated the robotic systems that are currently used for middle ear surgery as well as the control modes that are implemented on these systems. Due to various limitations associated with the existing approaches, we have envisioned the characteristics of an ideal system and the control modes that will improve current practices in cholesteatoma surgery. As observed in (Dahroug et al., 2018), we will need a robotic arm with a minimum of 4-DoFs, a workspace of 1,000 mm × 1,000 mm × 500 mm, a spatial accuracy of at least 0.05 mm. It has been also perceived that the

robotic system must be equipped with exteroceptive vision and force sensors. Besides, the control modes considered for surgical protocol are tele-operation, comanipulation, semi-automatic with operator intervention, and full-automatic mode for certain surgical sub-tasks (e.g., laser ablation of the residual cholesteatoma cells).

3 SHARED CONTROLLERS

Before discussing the proposed control modes, in this section, we first present the technical details of the classical tele-operation and comanipulation approaches. Following, the proposed hybrid tele-operation and comanipulation methods are detailed. **Figure 2** depicts the robotic system, the robot arm and the joystick with their associated frames and the notations that are used in the following. The incision hole and the fiducial marker are also present in the figure with the associated frames, representing the target position of the tool-tip and the camera FoV.
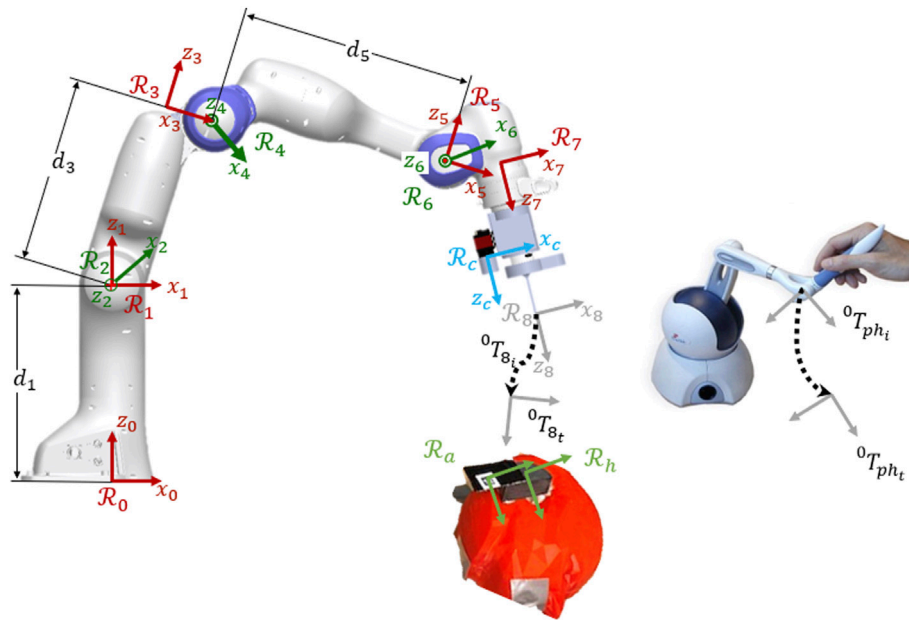
3.1 Classical Tele-Operation

Classical tele-operation is widely used for a variety of robotic applications (Simorov et al., 2012). The tele-operation mode developed in this work is based on a position-based controller allowing the interpretation of the local pose of a Phantom Omni end-effector as the desired pose of the tool tip attached to the robot end-effector. The motion of the joystick is directly mapped into the robotic system's end-effector space, which gives an intuitive position-based control for the user.

To control the robot by tele-operation, the desired position x_d of the joystick expressed in frame $\mathcal{R}_8$ is used as the control input (**Figure 3**). Using the continuous mapping between robot end-effector motion and that of the joystick, the position error $\tilde{x}$ (on the end-effector's position) is expressed as

$$\tilde{x} = x_d - x_e \qquad (1)$$

where, x_e is robot end-effector positions. $\tilde{x}$ goes through a proportional-derivative (PD) controller in order to get the control set-point Δx. Therefore, the robot joint space velocities Δq are obtained from the task-space velocities using the robot's Jacobian pseudo-inverse $J^{\#} \in \mathbb{R}^{8 \times 6}$:

FIGURE 2 | Kinematic model of the robotic system with the associated frames, the joystick for the teleoperation and the target positions (marker and incision hole).
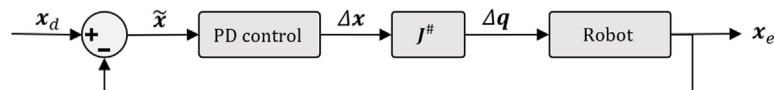
FIGURE 3 | Control scheme of the classical tele-operation.

$$\Delta \boldsymbol{q} = \boldsymbol{J}^{\#} \Delta \boldsymbol{x} \tag{2}$$

Note that internal movements of the robot arm are not exploited in our approach and $\Delta \boldsymbol{q}$ obtained from (2) is simply the minimum norm solution of the robot inverse differential kinematics.

3.2 Classical Comanipulation

Another classical strategy to control the motion of aforementioned robotic system is comanipulation. Such a technique has been widely used in the surgical domain to accomplish a variety of tasks (Zhan et al., 2015). In this work, we have implemented a comanipulation scheme on our robotic system using an external 6-DoF Force/Torque (F/T) sensor. In the presented system, the force sensor is attached to the robotic arm at the frame $\mathcal{R}_8$ where the control point is located. The implemented comanipulation mode is trivial and does not require any registration or specific parameters tuning.

To move the robot in the comanipulation manner, an ergonomic 3D printed wrist is provided to the user (see experimental setup **Figure 7** for details). The wrist is attached to the F/T sensor, which is attached to the distal part of the robotic arm. The user pushes the wrist and the F/T sensor provides the

applied force and torque onto the wrist, represented by $\boldsymbol{f}_{ext}$, expressed in the frame $\mathcal{R}_8$. The measured force and torque are then converted to desired operational velocities, represented by $\dot{x}_d$, thanks to a proportional gain and min/max filter and the output is injected into a PD controller (**Figure 4**). Finally, (2) allows converting the task-operational control vector $\Delta \boldsymbol{x}_8$ to joint-space operational one $\Delta \boldsymbol{q}_8$.

3.3 Developed Shared Controllers

The two control modes discussed above are functional and relatively suitable for performing a large number of tasks in an operating room. However, in order to make the controllers more intuitive and easier to use, especially for a non-specialist (e.g., a novice surgeon), we have developed shared controllers. Indeed, it has been shown that among the 6 DoFs involved in a 3D positioning task, rotations (i.e., θ_x, θ_y, θ_z) are the most difficult to achieve, in particular for a non-specialist (Adjigble et al., 2019). Accordingly, to reduce the cognitive load on the surgeon, we have proposed two decoupled control modes allowing to dissociate translations and rotations. Thereby, the three involved rotations will be managed automatically based on the vision feedback, when the three translations are still controlled by the introduced classical tele-operation or comanipulation modes.
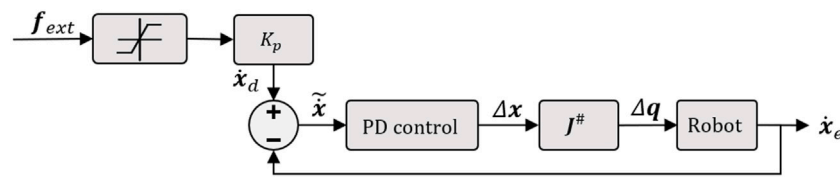
FIGURE 4 | Illustration of the implemented control scheme in case of comanipulation mode.
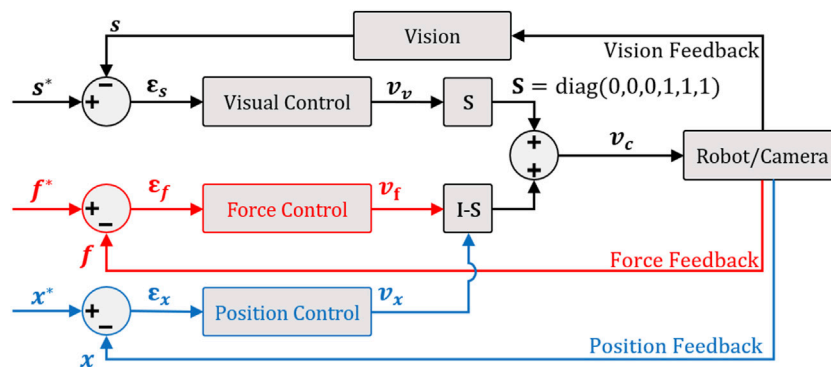
FIGURE 5 | Parallel hybrid force/vision comanipulation (in red) and position/vision teleoperation (in blue) control scheme.

3.3.1 Parallel Hybrid

The first shared controller we develop in this work is the parallel hybrid. It consists of the parallel juxtaposition of two internal loops. The first is a vision feedback loop to manage automatically the angular motion of the robot. The second is a position-based (in case of tele-operation mode) or force-based (in case of comanipulation) loop to control the linear motion. **Figure 5** shows the architecture of the so-called hybrid shared controllers.

To have a decoupled control law, i.e., rotations and the translation are controlled independently without one interfering with the other and then avoiding any conflict at the actuator level, we have opted for a position-based visual controller (Chaumette and Hutchinson, 2006). Thereby, to ensure the orthogonality between force or position and vision controller outputs, it is essential to introduce a selection matrix $\mathbb{S} = \text{diag}(0, 0, 0, 1, 1, 1)$. The selection matrix allows selecting the rotation DoFs to be managed by the vision-based control and the remaining DoFs, i.e., translations to be controlled by the force or position controller.

As previously stated, the context of this work is to execute a middle ear surgery wherein the target corresponds to an incision of millimeter-diameter performed in the mastoid (access to the middle ear cavity). Following the usual way for registration and surgery, a printed *AprilTag code* with known geometry is fixed on the mastoid (Katanacho et al., 2016). Using this, we estimate the full pose of the target (orientations and translations) in the camera frame $\mathcal{R}_c$. It is worth noting that the homogeneous transformation from the frame $\mathcal{R}_h$ attached to the incision hole towards the camera frame $\mathcal{R}_c$ is represented by $^c M_i$.

A feature characterizing the camera-hole relative configuration can be extracted from $^c M_i$ as $s = (t, \theta u)^\top$, where t is the translational part $(t_x, t_y, t_z)^\top$ of the homogeneous matrix $^c M_i$ and θu is the axis-angle representation of the rotational part. Furthermore, to regulate the Cartesian error e between the current position of the robot end-effector noted r and the desired position $r^\star$, corresponding to configuration when the tool is inserted in the incision hole. By considering $s = (t, \theta u)^\top$ as the current position and $s^\star = (t^\star, 0_{1\times3})^\top$ as the desired one, then the Cartesian error e is

$$e = s^\star - s = (\Delta t, -\theta u)^\top \tag{3}$$

where, $\Delta t = t^\star - t$. The time-derivation of s, i.e., $\dot{s} = (\dot{t}, \dot{\theta u})^\top$ allows linking the camera velocity twist $v = (v, \omega)^\top$ to the visual features variation (Marchand et al., 2002):

$$\dot{s} = \begin{pmatrix} I_{3\times3} & 0_{3\times3} \\ 0_{3\times3} & L_w \end{pmatrix} \begin{pmatrix} v \\ \omega \end{pmatrix} \tag{4}$$

where,

$$L_w = I_{3\times3} - \frac{\theta}{2}[u]_\times + \left(1 - \frac{\text{sinc}\theta}{\text{sinc}^2\frac{\theta}{2}}\right)[u]_\times^2 \tag{5}$$

in which $I_{3\times3}$ is the 3×3 identity matrix and L_w is the interaction matrix such that $L_w^{-1}\theta u = \theta u$ as reported in (Malis et al., 1999). Note that the notation θu represents the angle-axis parameterization of the rotation, $\text{sinc}(\theta) = \frac{sin(\theta)}{\theta}$, and $[u]_\times$ being the anti-symmetric matrix associated to the vector u. Therefore, the camera velocity twist can be expressed as follows:
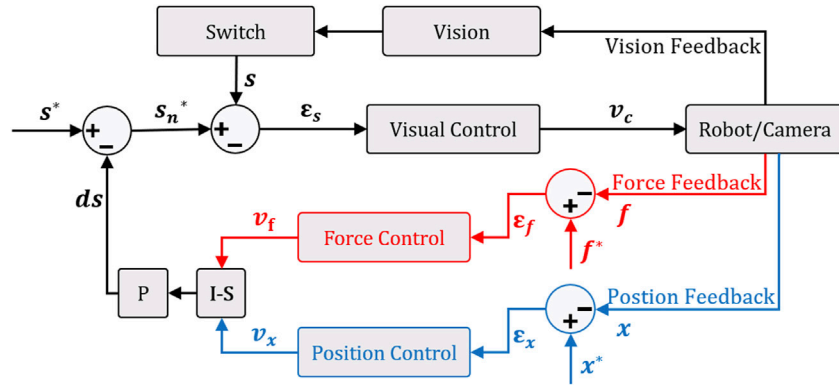
FIGURE 6 | External hybrid force/vision comanipulation (in red) and position/vision tele-operation (in blue) control scheme with the function of keeping the target in the center of the camera field-of-view.

$$\begin{pmatrix} v \\ \omega \end{pmatrix} = -\lambda \begin{pmatrix} t - t^\star \\ \theta u \end{pmatrix} \tag{6}$$

Finally, the selection matrix $\mathbb{S}$ allows expressing the parallel hybrid controller as follows: $v = v_f$ (provided by the force controller) and $\omega = -\lambda \theta u$ (provided by the vision controller).

3.3.2 External Hybrid

When both the force/position and vision-based controllers work in parallel, there is a risk that the tracked visual features (i.e., the *AprilTag*) are lost. This could jeopardize the accuracy of final positioning task. To tackle this issue, we have designed a new hybrid controller as shown in **Figure 6**. The underlying idea is to express the control task as two hierarchical sub-tasks. The first task (priority sub-task) deals with maintaining the target (i.e., *AprilTag*) at the center of the camera field-of-view (FoV), while the second one (secondary sub-task) is devoted to the regulation of the error between the current and the desired poses.

Therefore, to continuously maintain the *AprilTag* in the center of the camera FoV, we introduce a target locking mode based on the angular deviation θu_{lock} that can be defined as follows:

$$\begin{aligned} \theta u_{lock_x} &= \operatorname{atan2}(t_y, t_z) \\ \theta u_{lock_y} &= \operatorname{atan2}(t_x, t_z) \end{aligned} \tag{7}$$

where, t_x, t_y, and t_z are respectively the translation components along the x, y, z directions of the homogeneous matrix ${}^c M_o$ that represents the 3D pose of the *AprilTag* fiducial marker expressed in the camera frame $\mathcal{R}_c$. In the target locking mode, the rotation part of the desired position is then defined as $\theta u_{lock} = (\theta u_{lock_x}, \theta u_{lock_y}, \theta u_z)$. Besides, the Euclidean distance $e_{xy} = \sqrt{t_x^2 + t_y^2}$ expressed in the camera frame $\mathcal{R}_c$, is used to define a threshold to switch between free and locked modes:

$$s(e_{xy}) = \begin{cases} s = (0_{1\times3}, \ \theta u_{lock})^\top, & \text{if } e_{xy} > \epsilon = \sqrt{\frac{1}{2}} d \\ s = (0_{1\times3}, \ \theta u)^\top, & \text{else} \end{cases} \tag{8}$$

where, d is the side-size of the *AprilTag* and ϵ is a predefined threshold triggering the switch between free and locked modes.

In this external hybrid force (or position) vision controller, the controller outputs are used to modify the desired visual features vector $s^\star$. This is associated in the visual servo control by considering that the desired visual features is $s_n^\star = s^\star - ds = (t^\star - t_{ds}, \ 0_{1\times3})^\top$ and the Cartesian error is $e = s_n^\star - s(e_{xy})$. Thereby, the final camera velocity twist is expressed as follows:

$$\begin{pmatrix} v \\ \omega \end{pmatrix} = -\lambda \begin{pmatrix} t_{ds} - t^\star \\ \theta u(e_{xy}) \end{pmatrix} \tag{9}$$

then,

$$v = -\lambda (t_{ds} - t^\star)$$

$$\omega = \begin{cases} -\lambda \theta u_{lock}, & \text{if } e_{xy} > \sqrt{\frac{1}{2}} d \\ -\lambda \theta u, & \text{else} \end{cases} \tag{10}$$

Our system is designed in an *eye-in-hand* configuration. Thus, the relation between the robot velocity $\dot{q}$ and the camera velocity v is obtained as follows:

$$\dot{q} = -J^{\# \, 0} V_c \, v \tag{11}$$

where $J^\#$ is the pseudo-inverse kinematic Jacobian matrix of the 7-DoF robot arm in the base frame $\mathcal{R}_0$, and ${}^0 V_c$ is the transformation matrix associated to the velocity change frame from $\mathcal{R}_c$ to $\mathcal{R}_B$ constructed as:

$${}^0 V_c = \begin{bmatrix} {}^0 R_c & {}^0 t_{c_{[\times]}} {}^0 R_c \\ 0 & {}^0 R_c \end{bmatrix}$$

where ${}^0 R_c$ is the 3×3 rotation matrix from $\mathcal{R}_c$ to $\mathcal{R}_B$, ${}^0 t_c$ is the 3×1 associated translation vector, and ${}^0 t_{c_{[\times]}}$ is the skew symmetric matrix associated to the vector cross-product.

4 EXPERIMENTAL VALIDATION

4.1 Experimental Setup

To assess the proposed controllers, we have developed an experimental setup consisting of a 7-DoF cobot (FRANKA
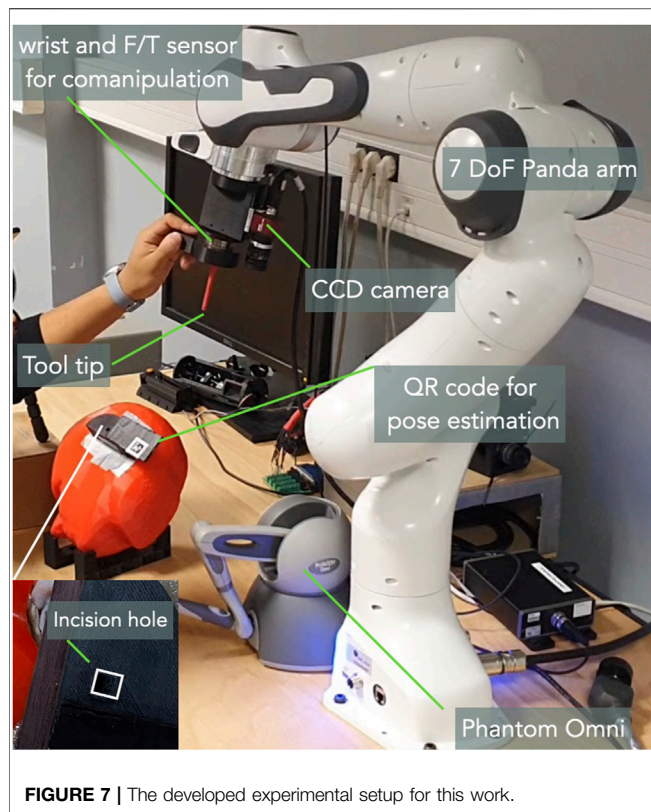
FIGURE 7 | The developed experimental setup for this work.

EMIKA Panda), as shown in **Figure 7**. A 3D printed tool (2 mm of diameter), which mimics a typical surgical instrument used to operate middle ear, is fixed as an end-effector of the robot. A standard CCD camera, AVT Guppy PRO F033b (resolution: 800 × 600 pixels and frame-rate: 25 images/second), is mounted in an *eye-in-hand* configuration on the robot end-effector. A Sensable Phantom Omni haptic device is used to teleoperate the robotic system. For the sake of comanipulation, a 6-DoF force/torque sensor (ATI MINI-40) is fixed at the robot distal part. A head phantom (**Figure 7**) at scale 1 : 1 is positioned to simulate the position of a patient on the operating table. Finally, the tunnel drilled on the head has the shape of the 3D tool fixed on the robot with a tolerance of 0.1 mm. It is worth noting that an *AprilTag QR-code* marker is positioned next to the incision hole, with a distance equal to that of between the tool-tip and the camera. This is to maintain the *AprilTag* inside the camera's FoV until the end of insertion task.

4.2 Validation Scenario

The work is validated by analysing the scenario where a 3D printed tool is inserted into an incision hole. Initially, the robot is placed in an arbitrary position, then the operator jogs it towards the incision so as to insert the tool. The depth of the incision hole is estimated to be 5 *mm*. A small line was drawn at 5 *mm* before the tool-tip. A group of 5 participants (2 expert and 3 novice) are volunteered to carry out the positioning and insertion tasks using different tele-operation (classic, parallel hybrid and external hybrid) and comanipulation (classic, parallel hybrid and external hybrid) control modes. For each of the performed

tasks, Cartesian errors (along each DoF), the twist v_c expressed in the camera frame, the 3D trajectory of the robot end-effector, as well as the time required to achieve the tasks are recorded and analysed.

4.3 Experimental Results

As stated above, the insertion task was performed by 5 different participants. Every participant was asked to repeat the same task three times, alternatively using the developed controllers (3 control modes for the tele-operation and 3 others for the comanipulation).

4.3.1 Accuracy

The first analysis metric concerns the accuracy of each control mode. The final error e_f corresponds to the difference between the final position of the robot end-effector and the reference position (recorded when the 3D printed tool is inserted into the incision hole).

Table 1 summarizes the Cartesian error for each DoF. It can be noticed that the classical tele-operation mode is slightly more accurate than the external hybrid one. However, the parallel hybrid mode is very accurate (**Figure 8A**) with a mean linear error (average of e_x, e_y, and e_z) estimated to be 0.94 mm (2.24 and 4.17 mm for the classical and external hybrid methods, respectively), while the mean angular error (average of e_{θ_x}, e_{θ_y}, and e_{θ_z}) is 0.17° (6.74° and 4.27° for the classical and external hybrid methods, respectively).

Furthermore, concerning the comanipulation control modes, we can point out that both the shared controllers (parllel hybrid and external hybrid) outperform the classical control law (**Figure 8B**). Finally, as for the tele-operation case, the parllel hybrid is more accurate. The average linear error is $e_t = 0.67$ mm (1.32 and 1.49 mm for the classical and external hybrid methods, respectively), while the average angular error is $e_r = 0.45°$ (6.67° and 3.07° for the classical and external hybrid methods, respectively). This conclusion is confirmed by a post-hoc Tukey HSD analysis to evaluate the relevance of the obtained statistical data. As reported in **Table 1**, the comparison of the different control modes is statistically relevant confirming that the parallel hybrid control law outperforms in term of accuracy the other control laws, in both tele-operation (**Figure 8B**) and comanipulation (**Figure 8B**).

4.3.2 Trajectory

The second evaluation criteria concerns the behavior of the robot end-effector while accomplishing the task (positioning and insertion). To do this, we have recorded the spatial trajectory performed by the 3D printed tool for each control mode. An example is shown in **Figure 9A**, which depicts the 3D trajectory performed by the tool tip for each of the three control tele-operation modes achieved by one subject. **Figure 9B** shows the trajectories achieved using the comanipulation modes. It can be pointed out that those executed by the parallel hybrid method are smoother compared to the others. This suggests that the parallel hybrid method is more intuitive and easier to handle. On the other hand, it appears that the major difference between the trajectories lies especially in the second phase of the task,

TABLE 1 | Comparison of the positioning errors obtained with different control modes.

Control	e_x (mm)	e_y (mm)	e_z (mm)	e_{θ_x} (deg)	e_{θ_y} (deg)	e_{θ_z} (deg)
CTo1	1.79 ± 1.57	1.21 ± 1.09	3.47 ± 6.84	6.88 ± 3.56	4.18 ± 3.45	9.16 ± 10.07
PHTo2	**0.78 ± 0.55**	**1.22 ± 0.93**	**0.83 ± 0.77**	**0.34 ± 0.17**	**0.31 ± 0.24**	**0.5 ± 0.28**
EHTo3	2.51 ± 3.65	1.39 ± 1.11	2.61 ± 4.79	6.46 ± 6.83	5.6 ± 5.93	0.76 ± 0.62
CCo4	0.73 ± 0.6	1.73 ± 1.09	1.52 ± 1.18	3.67 ± 3.33	7.49 ± 4.14	8.87 ± 9.22
PHCo5	**0.49 ± 0.52**	**0.97 ± 0.5**	**0.55 ± 0.52**	**0.37 ± 0.27**	**0.46 ± 0.34**	**0.52 ± 0.31**
EHCo6	1.16 ± 0.47	2.19 ± 0.96	1.12 ± 0.69	1.84 ± 1.62	5.88 ± 5.7	1.49 ± 1.37

*Bold represents the best performance obtained in terms of accuracy on all degrees of freedom (translations and rotations) when comparing the different proposed control modes.*Tukey HSD p-value (tele-operation, mean angular error): classical vs external hybrid → p = 0.005 and classical vs parallel hybrid → p = 0.001. Note that p < 0.05, which means statistically significant.*
Tukey HSD p-value (comanipulation, mean angular error): classical vs external hybrid → p = 0.001 6 and classical vs parallel hybrid → p = 0.001.
Tukey HSD p-value (comanipulation, mean linear error): classical vs external hybrid → p = 0.001 and classical vs parallel hybrid → p = 0.003 9.
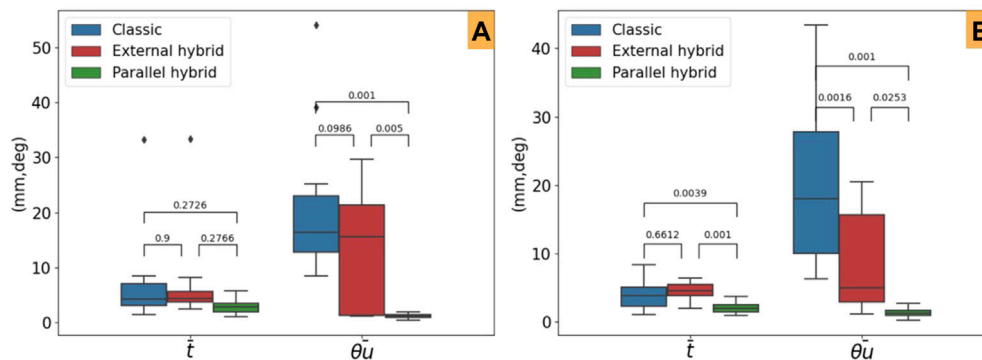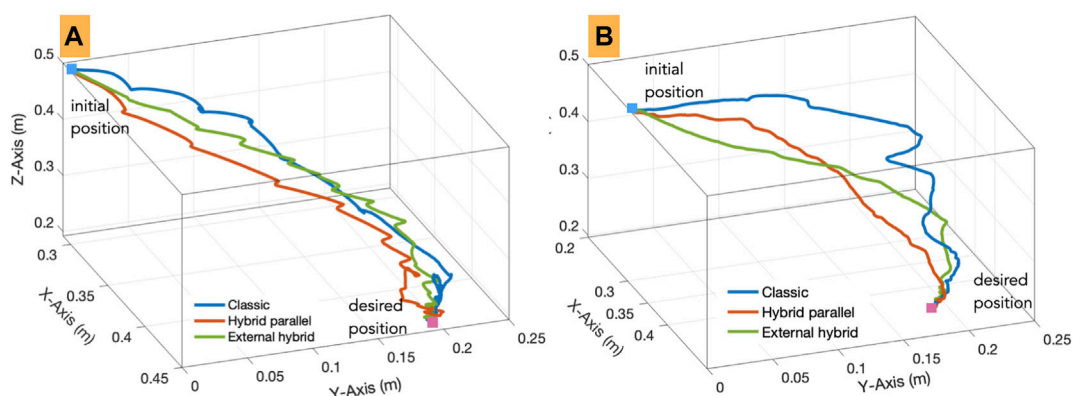[a]*CTo, classical tele-operation.*
[b]*PHTo, parallel hybrid tele-operation.*
[c]*EHTo, external hybrid tele-operation.*
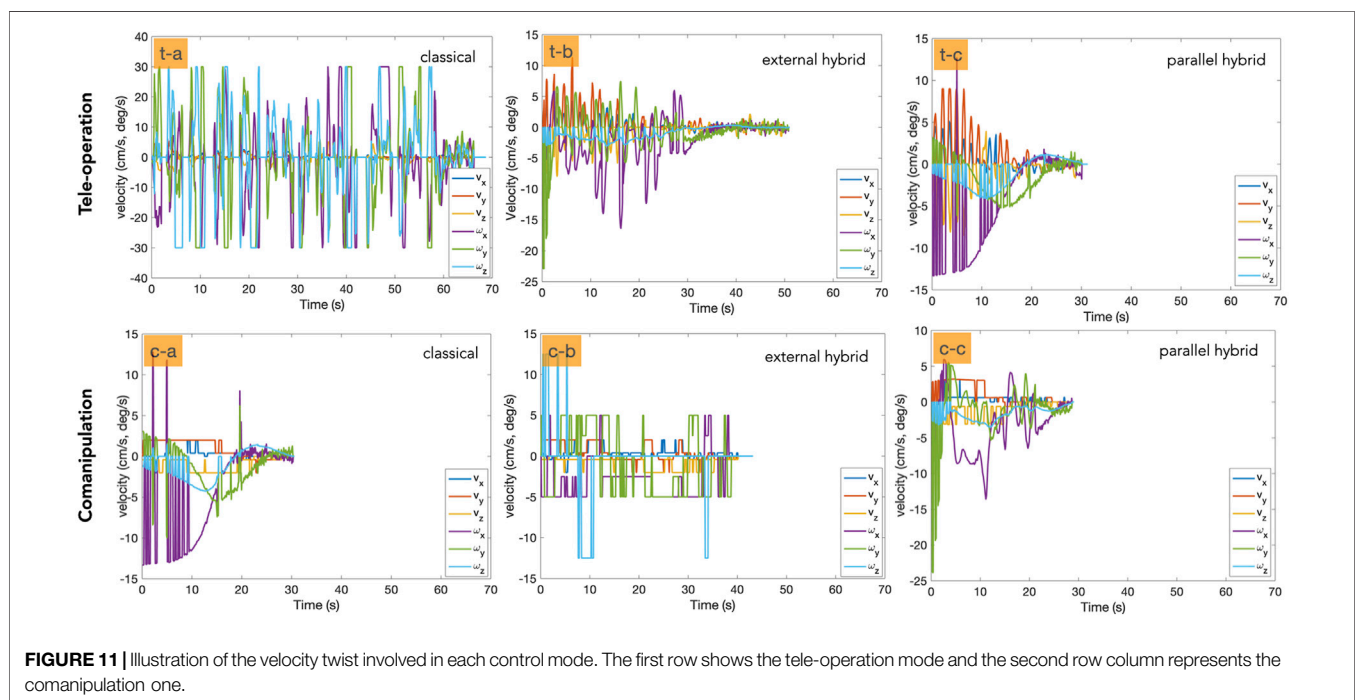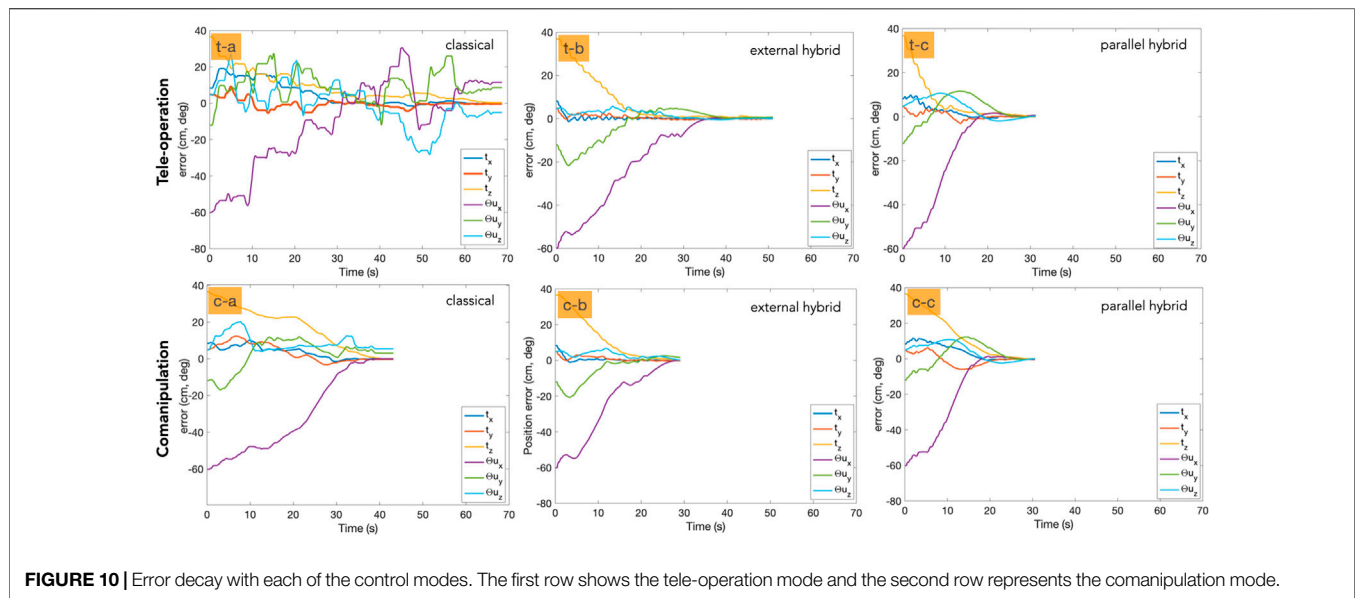[d]*CCo, classical comanipulation.*
[e]*PHCo, parallel hybrid comanipulation.*
[f]*EHCo, external hybrid comanipulation.*

FIGURE 8 | Mean steady-error and post-hoc Tukey HSD *p*-values for the evaluated control laws. **(A)** Tele-operation modes and **(B)** comanipulation ones.

FIGURE 9 | 3D trajectories carried out by an operator using the implemented **(A)** tele-operation and **(B)** comanipulation modes.

i.e., insertion of the tool into the incision hole. Note that the insertion task is considered completed when the tool is inserted 5 mm into the simulated incision hole. In case of classical tele-operation and comanipulation methods, the operator has to repeat the process several times with small movements to find the correct orientation of the tool with

FIGURE 10 | Error decay with each of the control modes. The first row shows the tele-operation mode and the second row represents the comanipulation mode.

FIGURE 11 | Illustration of the velocity twist involved in each control mode. The first row shows the tele-operation mode and the second row column represents the comanipulation one.

respect to the incision hole before starting the insertion, while with the developed hybrid methods the operator can achieve this insertion with minimum effort.

4.3.3 Behaviour

The third evaluation criteria concerns the behaviour of the controllers according to the quality of the error **e** regulation as well as the velocities **v** sent to the robot while accomplishing the task. As can be seen in **Figure 10** for teleoperation, the regulation to zero of the error in each DoF is smoother in

the case of hybrid tele-operation methods (**Figure 10** (t-b) (t-c)) compared to the classical one (**Figure 10** (t-a)). In addition, both the hybrid controllers converged to zero, when the classical mode shows a significant residual error in several DoFs. The same observation can be made for the comanipulation evaluation as depicted in **Figure 10** (right column). In contrast to the classical comanipulation (**Figure 10** (c-a)), the hybrid comanipulation methods show smoother and close to exponential decay of the error (**Figure 10** (c-b) (c-c)).

TABLE 2 | Time required to achieve the task with different control schemes.

Control method	Time duration (s)
Classical tele-operation	64.05 ± 33.84
Parllel hybrid tele-operation	**40.53 ± 10.09**
External hybrid tele-operation	67.82 ± 26.53
Classical comanipulation	39.22 ± 8.38
Parallel hybrid comanipulation	**29.03 ± 6.94**
External hybrid comanipulation	36.53 ± 5.40

Bold represents the best performance obtained in terms of accuracy on all degrees of freedom (translations and rotations) when comparing the different proposed control modes.

We have also analysed the velocity twist **v** sent to the robot. **Figure 11** (left column) depicts the velocity twist evolution with tele-operation controllers, while **Figure 11** (right column) shows the same for comanipulation modes. In the case of tele-operation, the hybrid approaches (**Figure 11** (t-b) (t-c)) obviously show better behavior compared to the classical ones (**Figure 11** (t-a)). Additionally, it can be noticed that the parllel hybrid methods (**Figure 11** (c-c)) outperform both the classical and the external hybrid ones (**Figure 11**(c-a) (c-b)).

4.3.4 Time Required to Achieve the Task

It was observed that the total time required to accomplish the task varies significantly from the hybrid methods comparing to the classical ones, for both the tele-operation and comanipulation modes. **Table 2** summarizes the observed average time. It appears that the parllel hybrid tele-operation approach requires in average 40.53 ± 10.09 s, which means 50% faster than the classical and external hybrid ones. The same conclusion can be made for the comanipulation methods. Indeed, the parllel hybrid controller requires on average 29.03 ± 6.94 s to achieve the task, which means approximately 25% faster than the others control schemes.

5 CONCLUSION

In this paper, we discussed new shared control schemes in both tele-operation and comanipulation modes. The objective was to provide the surgeon with an intuitive, efficient, and precise control approach for minimally invasive surgery using a robotic system. We presented six control schemes: a classical end-frame tele-operation, two shared vision/position tele-operation controllers (called

external/parallel hybrid methods), a classical comanipulation controller, and two shared vision/force comanipulation controllers (also called external and parallel hybrid methods). In order to experimentally analyse the performances of the controllers and compare them to each other, we considered various evaluation criteria: accuracy, Cartesian trajectory of the tool tip, and time required to achieve the implemented positioning task. This evaluation showed that shared controllers (external hybrid and parallel hybrid) outperformed classical methods in both modes, i.e., tele-operation and comanipulation. In addition, the parallel hybrid approaches are much more efficient than the external hybrid ones in almost all the evaluation criteria.

Future work will focus on the implementation of the proposed controllers in real conditions of use, i.e., minimally invasive surgery in middle ear. Senior and junior surgeons will be recruited to evaluated the benefit of such approaches in the operating room.

DATA AVAILABILITY STATEMENT

The raw data supporting the conclusion of this article will be made available by the authors, without undue reservation.

AUTHOR CONTRIBUTIONS

SS and J-HS have contributed equally on conducting the research and prepared the manuscript. All authors contributed to manuscript revision, read, and approved the submitted version.

FUNDING

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/frobt.2022.824716/full#supplementary-material

REFERENCES

Abi-Farraj, F., Pedemonte, N., and Giordano, P. R. (2016). "A Visual-Based Shared Control Architecture for Remote Telemanipulation," in 2016 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), Daejeon, Korea, October 9-14, 2016 (IEEE), 4266–4273. doi:10.1109/iros.2016.7759628

Adjigble, M., Marturi, N., Ortenzi, V., and Stolkin, R. (2019). "An Assisted Telemanipulation Approach: Combining Autonomous Grasp Planning with Haptic Cues," in IEEE/RSJ International Conference on Intelligent Robots and Systems, Macau, China, November 4-8, 2019, 3164–3171. doi:10.1109/iros40897.2019.8968454

Aquino, J. E. A. P. D., Cruz Filho, N. A., and Aquino, J. N. P. D. (2011). Epidemiology of Middle Ear and Mastoid Cholesteatomas: Study of 1146 Cases. *Braz. J. Otorhinolaryngol. (Impr.)* 77, 341–347. doi:10.1590/s1808-86942011000300012

Baeten, J., Bruyninckx, H., and De Schutter, J. (2003). Integrated Vision/force Robotic Servoing in the Task Frame Formalism. *Int. J. Robotics Res.* 22, 941–954. doi:10.1177/027836490302210010

Bordure, P., Robier, A., and Malard, O. (2005). *Chirurgie otologique et otoneurologique.* France: Elsevier Masson.

Burgard, W., Moors, M., Fox, D., Simmons, R., and Thrun, S. (2000). "Collaborative Multi-Robot Exploration," in Proceedings 2000 ICRA. Millennium Conference. IEEE International Conference on Robotics and Automation. Symposia Proceedings (Cat. No. 00CH37065), San Francisco, CA, USA, April 24-28, 2000 (IEEE), 1, 476–481.

Chaumette, F., Hutchinson, S., and Corke, P. (2016). "Visual Servoing," in *Springer Handbook of Robotics*. Editors B. Siciliano and O. Khatib (Cham: Springer), 841–866. doi:10.1007/978-3-319-32552-1_34

Chaumette, F., and Hutchinson, S. (2006). Visual Servo Control. I. Basic Approaches. *IEEE Robot. Automat. Mag.* 13, 82–90. doi:10.1109/mra.2006. 250573

Dahroug, B., Tamadazte, B., and Andreff, N. (2017). "Visual Servoing Controller for Time-Invariant 3d Path Following with Remote centre of Motion Constraint," in IEEE International Conference on Robotics and Automation, Singapour, May 29-June 3, 2017, 3612–3618. doi:10.1109/icra.2017.7989416

Dahroug, B., Tamadazte, B., and Andreff, N. (2020). Unilaterally Constrained Motion of a Curved Surgical Tool. *Robotica* 38, 1940–1962. doi:10.1017/ s0263574719001735

Dahroug, B., Tamadazte, B., Weber, S., Tavernier, L., and Andreff, N. (2018). Review on Otological Robotic Systems: Toward Microrobot-Assisted Cholesteatoma Surgery. *IEEE Rev. Biomed. Eng.* 11, 125–142. doi:10.1109/ rbme.2018.2810605

Derlacki, E. L., and Clemis, J. D. (1965). Lx Congenital Cholesteatoma of the Middle Ear and Mastoid. *Ann. Otol Rhinol Laryngol.* 74, 706–727. doi:10.1177/ 000348946507400313

Espiau, B., Chaumette, F., and Rives, P. (1992). A New Approach to Visual Servoing in Robotics. *IEEE Trans. Robot. Automat.* 8, 313–326. doi:10.1109/ 70.143350

Hamilton, J. W. (2005). Efficacy of the Ktp Laser in the Treatment of Middle Ear Cholesteatoma. *Otology & Neurotology* 26, 135–139. doi:10.1097/00129492- 200503000-00001

Holt, J. J. (1992). Ear Canal Cholesteatoma. *The Laryngoscope* 102, 608–613. doi:10. 1288/00005537-199206000-00004

Katanacho, M., De la Cadena, W., and Engel, S. (2016). Surgical Navigation with QR Codes: Marker Detection and Pose Estimation of QR Code Markers for Surgical Navigation. *Curr. Dir. Biomed. Eng.* 2, 355–358. doi:10.1515/cdbme- 2016-0079

Liang, W., Gao, W., and Tan, K. K. (2017). Stabilization System on an Office-Based Ear Surgical Device by Force and Vision Feedback. *Mechatronics* 42, 1–10. doi:10.1016/j.mechatronics.2016.12.005

Malis, E., Chaumette, F., and Boudet, S. (1999). 2 1/2 D Visual Servoing. *IEEE Trans. Robot. Automat.* 15, 238–250. doi:10.1109/70.760345

Marchand, E., Chaumette, F., Spindler, F., and Perrier, M. (2002). Controlling an Uninstrumented Manipulator by Visual Servoing. *Int. J. Robotics Res.* 21, 635–647. doi:10.1177/027836402322023222

Mezouar, Y., Prats, M., and Martinet, P. (2007). "External Hybrid Vision/ force Control," in Intl. Conference on Advanced Robotics (ICAR07), Daegu, South Korea, November 4-8, 2019, 170–175.

Møller, P. R., Pedersen, C. N., Grosfjeld, L. R., Faber, C. E., and Djurhuus, B. D. (2020). Recurrence of Cholesteatoma - A Retrospective Study Including 1,006 Patients for More Than 33 Years. *Int. Arch. Otorhinolaryngol.* 24, e18–e23. doi:10.1055/s-0039-1697989

Morel, G., Malis, E., and Boudet, S. (1998). "Impedance Based Combination of Visual and Force Control," in Proceedings. 1998 IEEE International Conference on Robotics and Automation (Cat. No. 98CH36146), Leuven, Belgium, November-May 16-20, 1998, 2, 1743–1748.

Olszewska, E., Wagner, M., Bernal-Sprekelsen, M., Ebmeyer, J. r., Dazert, S., Hildmann, H., et al. (2004). Etiopathogenesis of Cholesteatoma. *Eur. Arch. Oto-Rhino-Laryngology Head Neck* 261, 6–24. doi:10.1007/s00405-003-0623-x

Prats, M., Martinet, P., del Pobil, A. P., and Lee, S. (2007). "Vision Force Control in Task-Oriented Grasping and Manipulation," in 2007 IEEE/RSJ International Conference on Intelligent Robots and Systems, San Diego, California, USA, October 29 - November 2, 2007, 1320–1325.

Roy, N., and Dudek, G. (2001). Collaborative Robot Exploration and Rendezvous: Algorithms, Performance Bounds and Observations. *Autonomous Robots* 11, 117–136. doi:10.1023/a:1011219024159

Selvaggio, M., Abi-Farraj, F., Pacchierotti, C., Giordano, P. R., and Siciliano, B. (2018). Haptic-Based Shared-Control Methods for a Dual-Arm System. *IEEE Robot. Autom. Lett.* 3, 4249–4256. doi:10.1109/lra.2018.2864353

Selvaggio, M., Giordano, P. R., Ficuciellol, F., and Siciliano, B. (2019). "Passive Task-Prioritized Shared-Control Teleoperation with Haptic Guidance," in 2019 International Conference on Robotics and Automation (ICRA), Montreal, Canada, May 20–24, 2019 (IEEE), 430–436.

Simorov, A., Otte, R. S., Kopietz, C. M., and Oleynikov, D. (2012). Review of Surgical Robotics User Interface: What Is the Best Way to Control Robotic Surgery? *Surg. Endosc.* 26, 2117–2125. doi:10.1007/s00464-012-2182-y

So, J.-H., Marturi, N., Szewczyk, J., and Tamadazte, B. (2022). "Dual-scale Robotic Solution for Middle Ear Surgery," in International Conference on Robotics and Automation (ICRA), Philadelphia, USA, May 23-27, 2022. (IEEE).

So, J.-H., Tamadazte, B., and Szewczyk, J. (2020). Micro/Macro-Scale Robotic Approach for Middle Ear Surgery. *IEEE Trans. Med. Robotics Bionics* 2, 533–536. doi:10.1109/tmrb.2020.3032456

Zhan, Y., Duan, X.-G., and Li, J.-X. (2015). "Review of Comanipulation Robot in Surgery," in 2015 IEEE International Conference on Mechatronics and Automation, Beijing, China, August 2-5, 2015, 1466–1471. doi:10.1109/icma.2015.7237701