Navigating Expertise Across Space and Time: Augmented Reality Points of Instruction for Remote Mentoring

Title of the PhD Proposal

Navigating Expertise Across Space and Time: Augmented Reality Points of Instruction for Remote Mentoring

Name(s) of PhD Advisor(s)

Gilles Bailly, Directeur de Recherche CNRS, Sorbonne Université (ISIR)
Ignacio Avellino, Chargé de Recherche CNRS, Sorbonne Université (ISIR)
Geoffroy Canlorbe, Professeur des universités-praticien hospitalier (PU-PH), AP-HP et Sorbonne Université, Pitié-Salpêtrière

Host Laboratory

ISIR, Sorbonne Université

Short abstract

Remote collaboration tools have redefined training practices, especially in domains where hands-on apprenticeship is foundational, such as in medicine. Pointers over shared video aim to support verbal instructions (do this *here*), and hand/tool overlays aim to support demonstration (do it *like this*). Yet to guide physical tasks effectively, experts need to provide instructions across large spaces that evolve as novices execute instructions. Current remote tools fail at supporting this practice as they only support "here and now" instructions. This thesis investigates how Augmented Reality Head-Mounted Displays can support experts in authoring spatially anchored gestural instructions, which we call Points of Instruction, that novices can independently navigate across space and time. Through field studies and controlled experiments, we aim to understand and support remote instruction across independent spatial and temporal dimensions in a surgical context.

Short description of hosting research group / lab:

ISIR is a lab with about 200 researchers divided into 5 teams. One HCI group spans researchers across the 5 teams, and studies interaction in robotics, AI, design, the arts, and more, with a strong emphasis on health applications. We publish our results in top-tier ACM venues such as CHI, CSCW, UIST, and IHM.

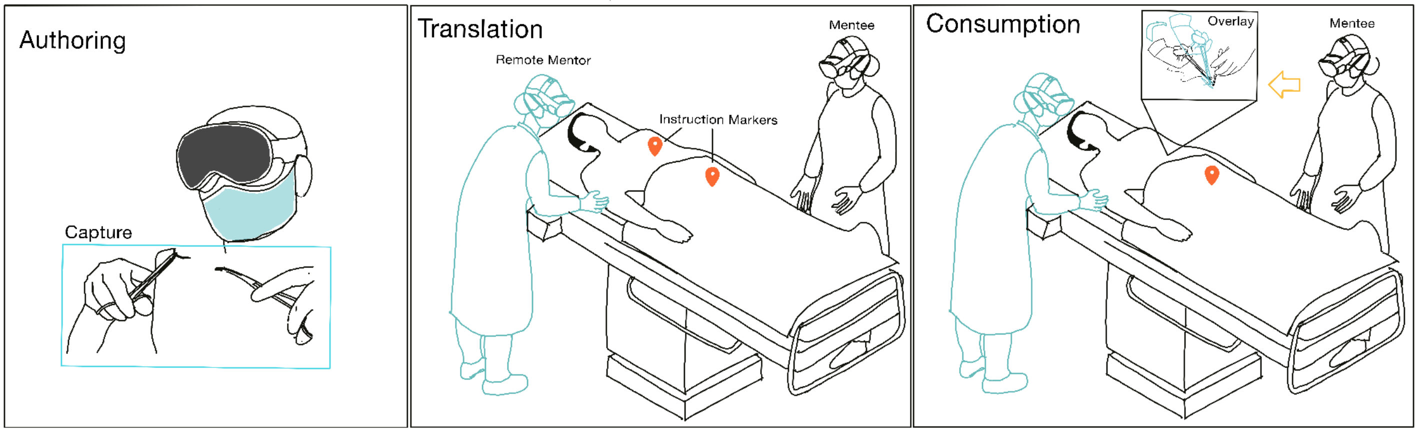

Figure 1. Overview of the Points of Instruction (POI) concept. (Left) The expert authors gestural instructions using an AR head-mounted display, capturing hand movements during demonstration. (Center) Instructions are translated into spatially anchored markers on patient-specific anatomy by a remote mentor, illustrated here as a virtual presence. (Right) The mentee independently accesses these instructions as AR overlays during the procedure.

Context

Remote collaboration tools have redefined training practices, especially in domains where hands-on mentorship is foundational, such as in surgery. Existing video-mediated tools for teleinstruction and telementoring rely on shared visual spaces augmented with pointers, which aim to support verbal instructions (do this *here*), and overlays, which aim to support demonstration (do it *like this*). These approaches assume relatively homogeneous interaction settings, where both an expert mentor and a novice mentee perceive and act upon similar representations of the task. This assumption breaks down in complex physical tasks such as minimally-invasive surgery. Here, the mentor and mentee operate in fundamentally different environments: the mentor can access a virtual reconstruction of the remote environment, while the mentee acts within a constrained intraoperative field (e.g., endoscopic video, evolving anatomy). These environments differ not only in representation but also in interaction modality, temporal constraints, and coupling to the physical world.

Current practice in collocated surgical settings already reveals forms of asynchronous and heterogeneous collaboration. It is common for an expert surgeon to prepare a senior resident in advance using generic video recordings to ground instructions on key procedural steps—e.g., how to dissect a vessel in order to clip it afterward using a particular technique. These instructions are later enacted independently during surgery, in a different context and under different constraints. However, current tools are not adapted to patient-specific anatomy and do not support precise spatial or temporal alignment between instruction and execution. This gap highlights the need for remote collaboration systems that explicitly account for heterogeneous interaction spaces, where instructions are produced in one setting and consumed in another, requiring precise spatio-temporal alignment across these differences.

Problem and Objectives

Many of the remote collaboration tools studied in both HCI and CSCW are designed to support “here and now” instructions and thus fail to fully support remote instruction on physical tasks spread across a complex space that changes over time. One problem is the interactional asymmetry: the remote mentor can only access a virtual representation of the physical space, and preoperative data (imaging, planned trajectories, annotations), while the mentee acts in an intraoperative, embodied space (endoscopic video view, partial visibility, evolving anatomy). These are different views but also different epistemic states of the same patient, potentially creating a mismatch between a predicted/planned state and a real-time observed state. Another problem is that the mentor produces instructions in a reflective, exploratory, and unconstrained space, while the mentee consumes them while engaged in a high-load motor task, often hands-busy, with strict attention constraints. This creates the need for different interaction modalities for producing and consuming instructions. This thesis’ goal is to design novel remote collaboration tools that enable remote experts to author spatio-temporal gestural instructions in virtual environments and local learners to access these instructions in augmented reality as they modify a physical space across time. For this, we will focus on the domain of surgery.

State of the Art

Prior work in HCI and CSCW has extensively studied the benefits of pointers and annotations over shared views (Lambert et al., 2024a and 2024b). However, complex physical tasks such as surgery involve rich non-verbal communication, including gesture demonstration (Lambert et al., 2024b). When mentoring remotely, current telestration tools fail to convey the gestural complexity of surgical instruction, particularly for non-visible instrument positions, force, and tissue handling (Lambert et al., 2024a). Augmented expert feedback (Lebrun et al., 2025 supported by PC1) can enhance learning of tool manipulation skills, although haptic feedback does not provide particular benefits. Digital overlays are a promising avenue, although research has yet to produce tools that support effective instruction across space and time.

Prior work has established that **hand gestures as overlays significantly improve performance in remote collaborative physical tasks**. Kirk & Fraser (2005) showed that when a remote helper guided a local worker through a physical construction task using gestures, the worker could independently rebuild the model faster 24 hours later compared to audio-only instruction. Kirk et al. (2005) also showed that projected hand gestures create mixed-reality ecologies that promote awareness and coordination, allowing helpers to perform object-focused gestures directly in a worker's space. Furthermore, Kirk & Fraser (2006) found that unmediated hand representations outperform digital sketches in speed with no loss of accuracy. In a recent study, Lambert et al. (2026) showed that **surgical tools as overlays** can effectively support instructions with spatial components (e.g., tool position, direction, and motion speed). These studies confirm the instructional value of advanced overlays for remote collaboration on physical tasks, but so far, none have addressed the spatial and temporal constraints of surgical instruction.

There have been limited studies investigating **remote collaboration tools specifically for spatially complex contexts.** Augmented Reality Head-Mounted Displays (ARHMDs) seem a promising avenue, although, in an extensive literature review of Mixed Reality tools for remote collaboration, Ens et al., (2019) point to two fundamental challenges: how to allow remote collaborators to *independently* explore the environment, and how to easily and effectively understand one another's views of the environment. Only Gauglitz et al. (2014) introduced a system that supports remote independent navigation of the view, enabling world-stabilized annotations that remain anchored to 3D spatial coordinates. However, their system supports telepointers, which only partially address gestural instruction, and does not support evolving environments nor asynchronous instructions. Two works supported by PC1 are conducting research in this area. Fèvre et al. (2024) are investigating view coordination across AR, VR, and desktop environments, addressing how on-site and remote collaborators share and synchronize viewpoints across different physicalities, and Lécolinet et al. (2025) are exploring asymmetric remote collaboration for physical therapy, where a VR-equipped therapist guides an AR-equipped patient through movement exercises. These ongoing projects study asymmetric settings, but the instructions remain synchronous and do not account for evolving materials that can create discrepancies between the instruction and the instructed action. We complement these projects by studying asynchronous instructions in changing materials, with the goal of not only transmitting instructions but also knowledge for remote learning.

Finally, **most remote collaboration tools are constrained to synchronous paradigms**, offering few mechanisms for novices to independently navigate instructions across time. Wang et al. (2021) review AR/MR remote collaboration on physical tasks, identifying five types of shared non-verbal cues, including gestures, yet the systems surveyed overwhelmingly assume synchronous co-presence between collaborators. Koh et al. (2025) have shown that people dynamically reposition assistive information during bimanual physical tasks. These insights can directly inform the design of spatially distributed instruction points that novices navigate on demand.

Research Questions

Our overarching question is: how to support remote instruction during physical tasks involving complex materials that evolve over time? We explore the following sub-questions:

- **RQ1.** What challenges do remote experts face when authoring gestural instructions with spatial and temporal components for learners in colocated settings, and how are current remote collaboration tools limited in supporting this practice?
- **RQ2.** How to align and translate between heterogeneous collaboration environments (pre-action vs intra-action, planned vs observed) in asynchronous collaboration?
- **RQ3.** How do mentees consume (navigate and access) spatially anchored instructions during task execution in augmented reality, and how do task constraints shape their usage patterns?

Our hypothesis is that ARHMDs can overcome current limitations by enabling experts to author spatially anchored hand gesture instructions, which we call *Points of Instruction* (POI), that novices can independently navigate across both space and time during remote physical tasks.

Building such a system poses three key challenges. First, capturing expert hand gestures with sufficient fidelity to remain meaningful beyond the moment of production. Second, spatially anchoring these gestures to the physical task environment. In surgery, this entails anchoring instruction to patient anatomy at varying depths, including tissue layers not yet exposed during early phases of a procedure, likely requiring integration with pre-operative imaging data. Finally, temporally decouple the resulting instructions from the authoring step, so novices can navigate them asynchronously from the expert who provided them.

Methodology and Scientific Approach

Methodology. WP1 — Understanding Instruction Authoring and Needs (M1-12) We will conduct field studies with surgeons and surgical trainees to understand (a) how experts conceptualize gestural instructions at anatomical locations, and (b) how novices use and benefit from these instructions during procedures. Then, participatory design workshops to understand how ARHMD overlays can support the role of pre-operative imaging data in enabling anchored instruction placement. We will use qualitative methods similar to Lambert et al. (2024b) and Koh et al. (2025).

WP2 — System Design and Implementation (M10-20). Building on insights from WP1, we will design and implement the system through iterative participatory design with surgical educators and trainees, structured around two heterogeneous interaction environments. A first mentor-side authoring environment, supporting exploration of patient-specific anatomy and expressive gesture input (e.g., VR or 3D interfaces), enabling experts to define spatially and temporally situated instructions. Then, a mentee-side execution environment, based on AR, supporting in-situ, hands-free access to instructions under real-time constraints. Key design challenges include: how to capture and place expert hand gestures from video data, how to enable novices to navigate these gesture demonstrations, and the mechanism for aligning instructions with evolving physical states during task execution.

WP3 — Evaluation (M18-30). We will evaluate the system through controlled experiments with medical residents, using suturing as a representative surgical task. We will compare AR-based gestural instruction against conventional conditions (co-located expert, 2D video), measuring performance with validated instruments such as OSATS and visuospatial ability measures similar to Jaafar et al., (2024).

Evaluation of the Contributions

At the system level, we will assess the accuracy and stability of spatially and temporally anchored overlays, including metrics such as spatial alignment and latency. We will also evaluate the fidelity of gesture interpretation (from authoring to execution). Regarding instruction, we will conduct controlled experiments with medical residents (we have done experiments with 450 residents in the past) to compare the proposed AR-based approach to baseline conditions (e.g., co-located guidance and 2D video), using validated measures such as OSATS to evaluate task performance, including precision, error rates, and completion time, as well as retention through delayed post-tests.

Nature of digital collaboration (1 page max)

Function
The function studied in this thesis is remote instruction and remote apprenticeship training. Our goal is to support these two functions for physical tasks across a complex space, in our case, the anatomy of a patient, which varies across time, in our case, as the surgical procedure unfolds and anatomy is modified.

Type
Both synchronous and asynchronous. Real-time surgical telementoring so that trainees can receive assistance from a remote expert. Current telementoring tools, namely telestration, constrain expert guidance to synchronous interventions: the expert can only direct what is happening in the moment. To better support synchronous collaboration, we will design tools that support instructions across space, and to enable asynchronous collaboration, we will study tools that support trainers in anchoring and trainees in accessing gestural instructions as the surgery evolves.

Time Scale
From Short-Term to Long-Term Interactions. Our first main goal is to support hours-long surgery, but by having the remote trainer interact in a shorter time frame (minutes) by establishing “the plan” of instructions (which ones, where, and when); and by having the trainee interact in bursts as needed (as the surgery unfolds, access instructions). By giving the trainee ownership and the ability to recall Points of Instruction, these resources can be used for learning.

Group Size
We envision a small group size, most of the time 1:1 or 1:2.

Space
The research primarily focuses on remote collaboration, from a medical center of expertise, which can have dedicated remote mentoring rooms, to a surgical room.

Long-Term Vision
The long-term vision is to enable remote instruction of complex physical tasks in a sustainable, scalable way. By enabling mentors to express spatio-temporally aligned instructions in a short time frame, rather than waiting for the trainee to execute actions that can take hours before the next instruction is needed, we free up a valuable trainer resource to assist the next learner. This, in time, will increase the mass of surgical talent and contribute to equitable access to healthcare. Finally, to assess the broader applicability of the POI concept, we will extend and evaluate it in additional healthcare training contexts, such as suturing and patient examination, to determine whether this approach generalizes.

Contribution to digital collaboration: Expected results and Impact (1 page max)

Empirical

Through field studies and controlled experiments, this thesis contributes to the growing literature on the limits of current remote collaboration tools for shared video (telestration) and mentoring practices for physical tasks.

Technological

The implementation of an AR head-mounted display remote training system that combines expert gesture instructions with on-site information, such as pre-operative imaging data, to produce spacio-temporal anchored instructions.

Theoretical

We will study the concept of Points of Instruction, a novel interaction concept that aims at supporting collaboration with a remote mentor across independent spatial and temporal dimensions.

Societal

Enabling asynchronous access to expert surgical guidance will support surgical training, particularly for institutions located far from specialized centers, and advance the democratization of healthcare.

Positioning in the eNSEMBLE program (½ page max)

This thesis addresses the scientific challenges of PC1 by investigating collaboration across heterogeneous interaction spaces in the context of remote surgical training. The work is structured along the core dimensions of collaboration spaces, with a particular focus on their articulation through asynchronous interaction. First, regarding the diversity of users, the project considers asymmetric collaboration between experts and trainees. These users differ not only in expertise, but also in roles, responsibilities, and availability. The collaboration is therefore inherently heterogeneous and often temporally decoupled, combining phases of tightly coupled interaction (instruction authoring) with loosely coupled or independent execution. Second, the project explicitly targets the diversity of interaction devices and modalities. It proposes a dual-environment system in which mentors author instructions using expressive, model-based interaction (e.g., gesture input in virtual or 3D environments), while mentees consume and enact these instructions through constrained, in-situ interaction (e.g., augmented reality under hands-busy conditions). A central challenge is to translate and adapt instructions across these heterogeneous modalities while preserving expert intent and ensuring usability in high-risk contexts. Third, regarding task complexity, the project targets a surgical environment that evolves over time: anatomy changes as tissues are displaced, instructions must be anchored to structures at varying depths, and the mentee operates under attentional strain and motor constraints. These properties create challenges for spatial anchoring, instruction persistence, and timely access, which must be addressed across a transition from the pre-operative planning environment, where instructions are authored, to the intra-operative environment where they are enacted.

This thesis also builds upon ongoing research under PC1. Fèvre et al. (2024, supported by PC1) are investigating in PhyDiViews the view coordination between on-site AR collaborators and remote VR experts in hybrid environments. Our work extends this from *view coordination* to *instruction coordination*. Rather than sharing what collaborators see, we study how to share what experts demonstrate. Lécolinet et al. (2025, supported by PC1) are exploring remote movement guidance for physical therapy between a VR-equipped therapist and an AR-equipped patient. Our project explores similar questions in surgery, a setting distinct from rehabilitation, where guidance involves finer instrument manipulation, greater risk, and an evolving action material.

References

Ens, B., Lanir, J., Tang, A., Bateman, S., Lee, G., Piumsomboon, T., & Billinghurst, M. (2019). Revisiting collaboration through mixed reality: The evolution of groupware. International Journal of Human-Computer Studies, 50 Years of the International Journal of Human-Computer Studies. Reflections on the Past, Present and Future of Human-Centred Technologies, 131, 81–98. https://doi.org/10.1016/j.ijhcs.2019.05.011

Fleury, C. & Nigay, L. (2024). PhyDiViews — Collaboration in Hybrid Environments: Multi-Views with Different Physicalities and Dimensions. PhD project funded by PEPR eNSEMBLE, PC1 CATS.

Gauglitz, S., Nuernberger, B., Turk, M., & Höllerer, T. (2014). World-stabilized annotations and virtual scene navigation for remote collaboration. Proc. UIST '14, 449–459. https://doi.org/10.1145/2642918.2647372

Kirk, D., Crabtree, A., & Rodden, T. (2005). Ways of the hands. Proc. ECSCW '05, 1–21.

Kirk, D. S. & Fraser, D. S. (2005). The effects of remote gesturing on distance instruction. Proc. CSCL '05, 301–310.

Kirk, D. & Stanton Fraser, D. (2006). Comparing remote gesture technologies for supporting collaborative physical tasks. Proc. CHI '06, 1191–1200. https://doi.org/10.1145/1124772.1124951

Koh, Y. S., Bailly, G., & Avellino, I. (2025). How Do People Position Assistive Information During Physical Tasks? Insights for Designing Cross-Domain AR Systems. Proc. IHM '25. https://doi.org/10.1145/3765712.3765733

Lambert, S., Voros, S., Canlorbe, G., Troccaz, J., & Avellino, I. (2024a). Understanding Takeovers and Telestration in Laparoscopic Surgery to Inform Telementoring System Design. Proc. CHI '24. https://doi.org/10.1145/3613904.3641978

Lambert, S., Voros, S., Troccaz, J., Canlorbe, G., & Avellino, I. (2024b). Understanding The Needs of Mentoring in Surgery to Guide the Design of Surgical Telementoring Systems. Proc. ACM Hum.-Comput. Interact., 8(CSCW1). https://doi.org/10.1145/3637420

Lambert, S., Voros, S., Canlorbe, G., Maria, S., Troccaz, J., & Avellino, I. (2026). Virtual Instruments: Beyond Telestration in Surgical Telementoring. To appear in Proc. ACM Hum.-Comput. Interact., (CSCW).

F. Lebrun, C. Simon, A. Boukezzi, S. Otmane and A. Chellali, "Mentor-Guided Learning in Immersive Virtual Environments: The Impact of Visual and Haptic Feedback on Skill Acquisition," in IEEE Transactions on Visualization and Computer Graphics, vol. 31, no. 5, pp. 3547-3557, May 2025, doi: 10.1109/TVCG.2025.3549547.

Lécolinet, E., Medeiros, D., & Serrano, M. (2025). Asymmetric Remote Collaboration for Physical Therapy and Implications on Rehabilitation/Movement Guidance. PhD project funded by PEPR eNSEMBLE, PC1 CATS.

Wang, P., Bai, X., Billinghurst, M., Zhang, S., Zhang, X., Wang, S., He, W., Yan, Y., & Ji, H. (2021). AR/MR Remote Collaboration on Physical Tasks: A Review. Robotics and Computer-Integrated Manufacturing, 72, 102071. https://doi.org/10.1016/j.rcim.2020.102071