Exploring Annotations and Hand Tracking in Augmented Reality for Remote Collaboration

Tiago Madeira[1], Bernardo Marques[1], João Alves[1], Paulo Dias[1] and Beatriz Sousa Santos[1]

[1] IEETA, DETI, University of Aveiro, Aveiro, Portugal
iagomadeira@ua.pt

Abstract. Collaboration among remotely distributed professionals is often required in a maintenance context. Professionals need mechanisms with adaptive capabilities to enable knowledge transfer, since the necessary experience and expertise are usually spread among different professionals. To provide a shared understanding, Augmented Reality (AR) has been explored. In this paper, which is part of ongoing research using a human-centered design approach with partners from the industry sector, we describe a framework using annotations to improve the shared perceived realities of different professionals. The framework allows manually freezing the on-site professional context and sharing it with a remote expert to create annotations. Then, the on-site professional can visualize instructions through aligned and anchored annotations, using a see-through Head Mounted Display (HMD). In addition, annotations based on real-time video stream from a remote expert are also available. Hand tracking is used to manipulate the annotations, enabling the adjustment of their position and scale in the real-world according to the context, thus enriching the on-site professional experience and improving visualization of information while conducting maintenance procedures suggested by a remote expert.

Keywords: Remote Collaboration · Distributed Professionals · CSCW · Maintenance · Framework · Augmented Reality · Annotations · Hand Tracking

1 Introduction

Remote collaboration has important value in the industrial domain, since the necessary experience and expertise are usually distributed among different professionals [1, 2]. The field of Computer-Supported Cooperative Work (CSCW) has been concerned with understanding and designing solutions to support remote collaboration [3]. Augmented Reality (AR) can be used to enable knowledge transfer from professionals unavailable on-site [3, 4], since it can provide a common ground environment, i.e., shared understanding for situation mapping, allowing identification of issues, and making assumptions and beliefs visible [5]. A number of earlier studies have explored different annotation mechanisms for remote collaboration [6]. Gauglitz et al. [7] explored how to stabilize annotations on the real-world using AR tracking in handheld devices. The authors found that users preferred stabilized annotations when compared to the unstabilized alternative.

Kim et al. [8] proposed the use of a manual freeze method to prevent annotations from being anchored to a wrong object, while an on-site user unexpectedly changes the viewpoint. This way, a remote user could manually freeze the live video received from the on-site user and draw on the still video frame, rather than in a live video. Then, s/he can again return to the live video.

Kim et al. [9] introduced a novel auto-freeze method, in which the freezing and unfreezing functions were integrated with the drawing functions to reduce the interactions required. The live video was automatically frozen when a remote user started drawing, and unfrozen when the drawing stopped. Then, a local user could use a Head Mounted Display (HMD) to visualize the annotations, while manipulating physical objects with hand gestures. The authors compared this method with the manual freeze concluding that the auto freeze method was easier to use and speeded up the annotation drawing process.

In this paper, we aim to provide interaction methods to improve cooperation among distributed professionals by using a see-through HMD and hand tracking for visualization and manipulation of different types of AR annotations. Our work, which is part of ongoing research using a human-centered design approach with partners from the industry sector, follows a set of requirements for the design of collaborative solutions described in previous work [reference blind for revision purposes]. The paper makes the following contributions:

- Describes an AR framework developed for remote collaboration;
- Presents different types of annotations as additional layers of information;
- Proposes the use of hand tracking for manipulation of spatial annotations.

2 An Augmented Reality Framework supporting Annotations

The framework developed within this work is designed to support scenarios that require know-how and additional information from professionals unavailable on-site, as is the case of maintenance scenarios. Therefore, it focuses on two types of users: on-site technicians and remote experts.

Framework Description

Maintenance procedures might be performed in multiple contexts i.e., different equipment and/or environments. Therefore, the framework was designed to function in a diverse set of conditions, using shared 2D annotations that can be placed onto existing equipment, instead of 3D virtual models, thus providing a more generic approach, not depending on the existence of pre-defined 3D models.

Figure 1 presents an overview of the framework. When facing unfamiliar problems, on-site technicians can use the camera of a HMD to manually capture (freeze) the context of the problem and send it to a remote expert for context understanding and to enable the instructions/feedback accordingly i.e., inform where to act and what to do using annotation features (in a laptop or desktop computer), allowing the creation of layers of additional information to illustrate difficulties, identify specific areas of interest or indicate questions. Afterwards, the on-site technician receives the instructions showing the suggestions from the remote expert, which can be visualized using the see-through HMD, overlapping the real-world context, while facilitating the performance of procedures intervention through a handsfree approach. The instructions can be presented using aligned and anchored annotations or through video stream displaying the creation process of the annotations. At any moment, the position of the annotations can be moved using hand tracking with the intent to provide a natural interaction mechanism. The overall collaborative process can be repeated iteratively until the task is successfully accomplished. The proposed framework also supports audio communication despite not being the focus of this research.

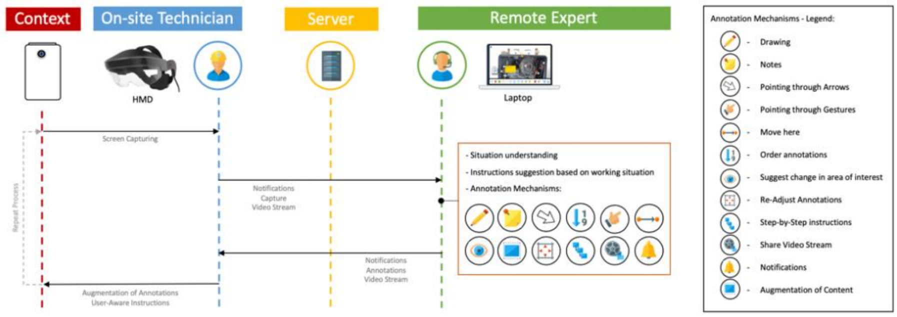

Fig. 1. Framework Overview. Goal: Allow an on-site technician to capture the real world using a see-through HMD and share with a remote expert for analysis and to provide instructions through the creation of annotations. Finally, the on-site technician can view the real world augmented with the annotations and perform an intervention.

Currently, the framework can be used on devices running Android, MAC OS, or Windows. Since on-site technicians are constantly moving, it seems adequate to equip them with HMD or handheld devices. Regarding the remote expert, we support multiple types of devices, including a laptop or desktop computer, an interactive projector, or an handheld device (Figure 2). The framework was developed using the Unity 3D game engine, based on C# scripts. To place the AR virtual content in the real-world environment, we used the Vuforia library. Communication between the different devices was performed over Wi-Fi through specific calls to a PHP server responsible for storing and sharing the enhanced pictures accordingly.

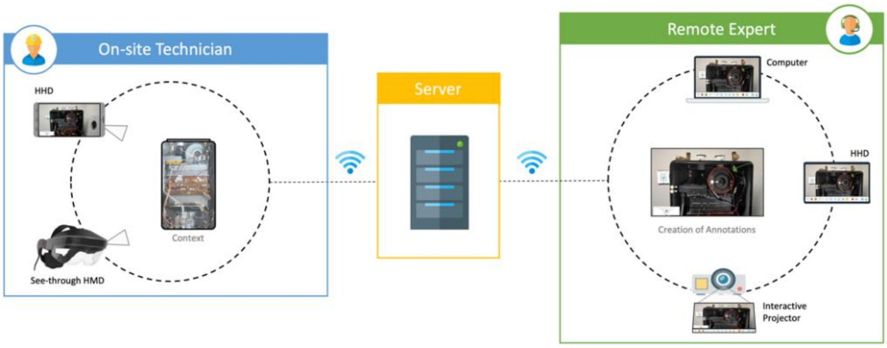

Fig. 2. Framework multiplatform possibilities for the On-site Technician (left: HHD of HMD) and for the Remote Expert (right – Computer, HHD or projection).

All annotations we use are stable, in the sense that they are all placed in the world. The difference is one kind of annotation is meant to be aligned with reality (and is interactive) and the others aren't (and are anchored to a specific pose in the world). Another possibility could be using unstabilized annotations, i.e., display in the corner of the user's field of view.

Aligned Annotations
This type allows annotations to be aligned with the real-world (Figure 3 - 3). The on-site technician may interact with the annotation and adjust its pose and size to better match reality.

Anchored Annotations
This type of annotation allows an image object to be anchored in a designated pose in the real-world to be consulted by the on-site technician while conducting a given procedure (Figure 3 - 2). It is also possible for the on-site technician to create one of these annotations by capturing their context through a snapshot saving feature (Figure 3 - 1).

Anchored Annotations based on video stream
This type of annotation allows a video to be displayed in a designed pose in the world to be consulted by the on-site technician, following the same method previous described (Figure 3 - 4).

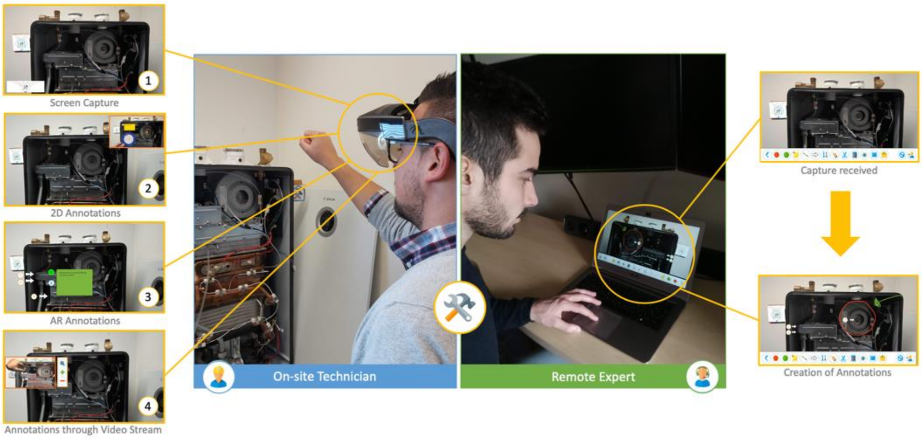

Fig. 3. Example of functions associated to the annotations of the framework for the on-site technician (left – capture and visualization of annotations) and the remote expert (right – situation understanding and creation of annotations).

Manipulation of annotations using Hand Tracking
We used a see-though HMD featuring gesture-based hands-tracking that enables the technician to interact with the annotations in a natural way. Closing one hand within the collision box of an annotation allows the on-site technician to grab it and move it in 6DOF. If two hands are used, this also enables the rotation and scaling of the annotation, by changing the distance between the user's hands and rotating the hands in relationship to each other, respectively.

3 Discussion

When on-site technicians use a hand-held device to visualize annotations, the need to hold the device with a particular orientation and position for proper observation of the information entails more cognitive and physical demands which may further hinder the ability to perform the task at hand [10]. The handsfree nature of the see-through HMD approach creates a setting where the interaction with the world is facilitated, enabling more complex tasks involving the use of two hands, while not compromising the access to information. Furthermore, the user interface design of the HMD, which leverages gesture-based hand tracking is thought to reduce the learning curve for unexperienced users by using hand interactions and allowing virtual objects to have physical properties of their real-world counterparts.

Allowing the annotations to be aligned with the real-world provides added value by complementing physical objects with information related to them in a straightforward way. The relationship between the target object and the information provided is inferred by enabling the on-site technician to perceive the indications of the remote expert as given directly within the physical context, creating a consistent view of the shared workspace. On the other hand, when using annotations that are not aligned with reality, the relationship between the information provided and the real-world context it refers to must be explicitly expressed, for instance.by presenting the information over a still picture of an object

This type of remote interaction may also be adequate for the medical field (e.g. particularly in surgery), simulators (e.g. as flight or driving simulators), military operations, and many other scenarios where cooperation or tutoring are required in a set of tasks involving dexterous use of the two hands.

4 Final Remarks and Future Work

In this study, we describe a framework that allows the use of different kinds of spatialized annotations, facilitating remote collaboration, while leveraging the use of hand tracking for manipulation.” We followed requirements for the design of collaborative tools using AR in remote scenarios, based on a human-centered design approach with partners from the industry sector. We integrated the use of a See-through HMD with hand tracking capabilities into an existing framework based on sharing of enhanced annotations. Through the use of different types of annotations, the framework allows the customization of the annotations position according to the context, allowing for the enrichment of each user’s experience and the merging of the shared perceived realities of different users.

This study is being expanded by conducting a user study to evaluate usability and acceptability and to understand if these new features of the framework can be robust in a real remote setting. Furthermore, we are also planning a focus group to present and discuss the most recent updates to the framework. Then, we intend to conduct formal user studies with domain experts to test our findings and validate the framework.

Acknowledgments. This research was developed in the scope of the PhD grant [SFRH/BD/143276/2019], funded by FCT - Foundation for Science and Technology. It was also supported by IEETA - Institute of Electronics and Informatics Engineering of Aveiro, funded by National Funds through FCT, in the context of the project [UID/CEC/00127/2019]. The context of this research is an industrial partnership collaboration under the Smart Green Homes Project [POCI-01-0247-FEDER-007678], a co-promotion between Bosch Termotecnologia S.A. and the University of Aveiro. It is financed by Portugal 2020 under the Competitiveness and Internationalization Operational Program, and by the European Regional Development Fund.

References

1. E. Arias, H. Eden, G. Fischer, A. Gorman, and E. Scharff, “Transcending the Individual Human Mind-Creating Shared Understanding through Collaborative Design”, ACM Transactions on Computer-Human Interaction, vol. 7, no. 1, pp. 84–113, 2000.
2. S. Johson, M. Gibson, and B. Mutlu, “Handheld or Handsfree? Remote Collaboration via Lightweight Head-Mounted Displays and Handheld Devices”, Computer Supported Cooperative Work and Social Computing, 2015, pp. 1825–1836.
3. B. Ens et al., “Revisiting Collaboration through Mixed Reality: The Evolution of Groupware”, International Journal of Human-Computer Studies, vol. 131, pp. 81–98, 2019.

4. H. Jalo, H. Pirkkalainen, O. Torro, H. Kärkkäinen, J. Puhto, and T. Kankaanpää, "How Can Collaborative Augmented Reality Support Operative Work in the Facility Management Industry?", Conference on Knowledge Discovery, Knowledge Engineering and Knowledge Management 2018, 2018, pp. 41–51.
5. Kangsoo Kim, Mark Billinghurst, Gerd Bruder, Henry Been-Lirn Duh, and Gregory F. Welch, "revisiting Trends in Augmented Reality Research: A Review of the 2nd Decade of ISMAR", IEEE Transactions on Visualization and Computer Graphics, vol. 24, pp. 2947–2962, 2018.
6. S. Kim, Mark Billinghurst, and G. Lee, "The Effect of Collaboration Styles and View Independence on Video-Mediated Remote Collaboration", Computer Supported Cooperative Work: CSCW: An International Journal, vol. 27, no. 3–6, pp. 569–607, 2018.
7. S. Gauglitz, B. Nuernberger, M. Turk, and T. Hollerer, "World-Stabilized Annotations and Virtual Scene Navigation for Remote Collaboration", ACM Symposium on User interface software and technology, 2014, pp. 449–459.
8. S. Kim, G. Lee, N. Sakata, and Mark Billinghurst, "Improving Co-Presence with Augmented Visual Communication Cues for Sharing Experience through Video Conference", IEEE International Symposium on Mixed and Augmented Reality 2014, ISMAR 2014, pp. 83–92, 2014.
9. S. Kim, G. A. Lee, S. Ha, N. Sakata, and Mark Billinghurst, "Automatically Freezing Live Video for Annotation during Remote Collaboration", ACM Conference Extended Abstracts on Human Factors in Computing Systems, CHI EA '15, pp. 1669–1674, 2015.
10. Rehman, Umair & Cao, Shi, "Augmented-Reality-Based Indoor Navigation: A Comparative Analysis of Handheld Devices Versus Google Glass", IEEE Transactions on Human-Machine Systems. pp. 1-12, 2016.