A Survey on Remote Assistance and Training in Mixed Reality Environments

Catarina G. Fidalgo, Yukang Yan, Hyunsung Cho, Maurício Sousa,
David Lindlbauer and Joaquim Jorge

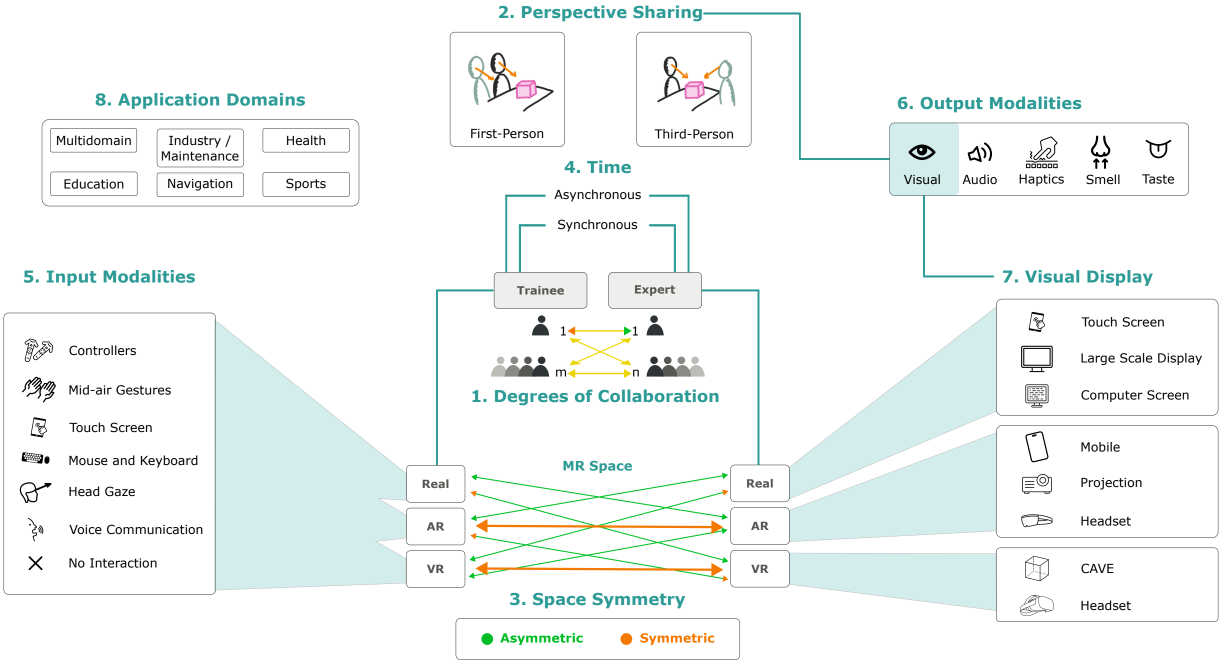


Fig. 1: Overview of our taxonomy for remote assistance and training in Mixed Reality (MR) environments.

Abstract—The recent pandemic, war, and oil crises have caused many to reconsider their need to travel for education, training, and meetings. Providing assistance and training remotely has thus gained importance for many applications, from industrial maintenance to surgical telemonitoring. Current solutions such as video conferencing platforms lack essential communication cues such as spatial referencing, which negatively impacts both time completion and task performance. Mixed Reality (MR) offers opportunities to improve remote assistance and training, as it opens the way to increased spatial clarity and large interaction space. We contribute a survey of remote assistance and training in MR environments through a systematic literature review to provide a deeper understanding of current approaches, benefits and challenges. We analyze 62 articles and contextualize our findings along a taxonomy based on degree of collaboration, perspective sharing, MR space symmetry, time, input and output modality, visual display, and application domain. We identify the main gaps and opportunities in this research area, such as exploring collaboration scenarios beyond one-expert-to-one-trainee, enabling users to move across the reality-virtuality spectrum during a task, or exploring advanced interaction techniques that resort to hand or eye tracking. Our survey informs and helps researchers in different domains, including maintenance, medicine, engineering, or education, build and evaluate novel MR approaches to remote training and assistance. All supplemental materials are available at https://augmented-perception.org/publications/2023-training-survey.html.

Index Terms—Mixed Reality, Virtual Reality, Augmented Reality, Extended Reality, Remote, Collaboration, Training, Assistance

1 Introduction

The environmental impact, cost, and time of air travel, emerging war and oil crisis, and the COVID-19 pandemic made the ability to work remotely very popular, if not essential. While video-conferencing has flourished, many opportunities lie ahead for more advanced approaches to remote interaction. We present a survey on remote assistance and training in Mixed Reality (MR) and contribute a taxonomy that contextualizes the main characteristics of this area based on aspects such as space, time, degree of collaboration, interaction modalities, and application area. Our work aims to identify common trends and open research areas in the field and provide an overview and in-depth discussion of this critical sub-area of Mixed Reality.

Collaboration is generally defined as the "*mutual engagement of participants in a coordinated effort to solve a problem together*" [64].

- *Catarina Fidalgo is with INESC-ID, Instituto Superior Técnico, University of Lisbon and Carnegie Mellon University. E-mail: cfidalgo@andrew.cmu.edu*
- *Yukang Yan, Hyunsung Cho and David Lindlbauer are with Carnegie Mellon University. E-mail: {yukangy | hyunsung | dlindlba}@andrew.cmu.edu*
- *Maurício Sousa is with University of Toronto. E-mail: mauricio.sousa@utoronto.ca*
- *Joaquim Jorge (Senior Member IEEE) is with INESC-ID and Instituto Superior Técnico, University of Lisbon. E-mail: jorgej@tecnico.ulisboa.pt*

Collaboration is essential for different professional areas, from healthcare and mechanical maintenance to professional education. Traditionally, much collaborative work relied on physically co-located people to enable efficient communication. However, the future of collaboration is increasingly digital and distributed across space and time [51]. The ability to work remotely is essential for resilient, scalable, and effective collaboration.

Remote *assistance* and *training* constitute particular cases of remote collaboration in which *remote (expert) users* typically train or guide *local (non-expert) users* through accomplishing some specific task. Examples include helping to operate complex machinery or training for a specific healthcare procedure. In this paper, among the broad collaborative problem solving domain, we focus on cases where remote users have greater expertise than local users. Remote assistance and training are particularly important due to their applicability across application domains, e. g., for industrial maintenance, surgical telemonitoring, or navigation, and the high variance in expert-to-trainee configurations. Therefore, we focus on these key terms to provide more in-depth information, rather than broadly surveying the topic of collaboration, which includes large portions of the area of computer-supported cooperative work (CSCW).

Systems that implement remote assistance and training approaches usually need to address two key challenges. First, remote users require a (virtual) representation of the local users' environment to identify all details necessary to complete the task in question. Second, all people involved, remote and local, need to communicate efficiently, which is typically enabled through advanced interaction techniques for exploring and annotating the shared environment (cf. Mohr et al. [59]). This survey provides details about how such virtual representations typically appear, what details are involved, and what communication techniques are used.

Enabling efficient collaboration through 2D displays, such as desktop computers, smartphones, and tablets, can be challenging since many application scenarios inherently require establishing situational awareness and a joint frame of reference within a complex 3D context. The immersive qualities of MR, which encompasses both Virtual Reality (VR) and Augmented Reality (AR) technology (cf. Milgram et al. [55, 56]), have the potential to tackle these challenges, as it opens a way for increased spatial clarity and larger interaction space [46]. In the context of this survey, we use established definitions of VR, which completely immerses users in a synthetic environment composed solely of virtual objects, and AR technology, which increases the sense of reality by superimposing computer-generated information such as virtual objects and cues on the real world [12]. We detail what technologies are employed within different application contexts and how those influence the interaction between users.

Previous work has conducted surveys related to understanding assistance and training, mostly focusing on co-located settings or the general scenario of collaboration. Sereno et al. [71] focus on surveying both co-located and remote collaborative work in AR but do not consider VR. Pidel and Ackerman [65] provide an overview of general collaboration in AR and VR, not specific to assistance and training. Wang et al. [89] explore remote collaboration in AR and VR with a specific focus on physical tasks. Lapointe et al. [46] review AR-based remote guidance tasks in combination with Artificial Intelligence (AI). Vaughan et al. [87] overview self-adaptive technologies within virtual reality training, and Kenoui [38] discusses AR in the context of healthcare training. None of the aforementioned approaches specifically focus on training and assistance with different combinations of AR and VR technologies. Given the prevalence of training and assistance, specifically in professional applications, and the inherent role differences between remote and local users in these scenarios, it warrants a more in-depth look to provide a better understanding of the unique opportunities and challenges. To our knowledge, we contribute the first survey that holistically considers remote assistance and training in MR.

Given recent progress in interaction and display techniques, we contribute an overview of the current landscape of assistance and training in MR between the years 2000 and 2022. We perform a systematic search for articles in different databases: *ACM Digital Library*, *IEEE Xplore*, and *Science Direct*. After screening 931 articles, we identified 62 articles related to our research questions, following PRISMA guidelines [58]. We characterize the articles along a taxonomy based on time, degree of collaboration, MR space symmetry, input and output modality, visual display, perspective sharing, and application domain. The proposed taxonomybuilds upon existing research and analysis on approaches for remote collaboration in AR and VR, to enable researchers to understand and explore areas that have not yet seen much research and to identify underexplored areas that might provide promising outcomes. As examples for promising directions, we identify opportunities for future solutions to enable multiple experts to jointly provide assistance to one or more trainees, in contrast to the typical focus on individual expert; or to provide more opportunities for users to move across the reality-virtuality spectrum during a task. Our work provides key insights for the future of remote assistance and training, including general trends and opportunities for further research.

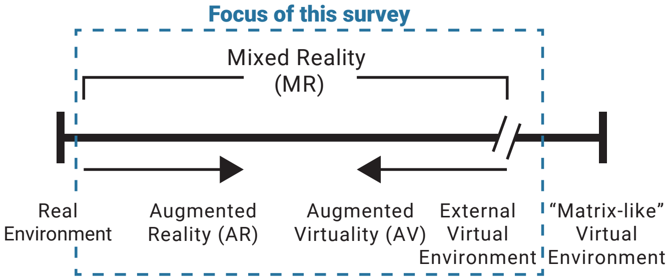


Fig. 2: Revised reality-virtuality continuum. MR encompasses everything in between a fully real environment and a "Matrix-like" virtual environment [72]. Within the context of this survey, we focus on the full spectrum except the two extremes.

2 Scope and Research Questions

We first define the research objectives of our work. We are concerned with approaches that enable remote assistance and training in an MR environment. Throughout the survey, we refer to the user providing training expertise or assistance insights as the *expert*, and to the user receiving such training or assistance as the *trainee*. We consider approaches where one or multiple experts and trainees interact with each other.

2.1 Definition: Mixed Reality

There is no universal agreement on the definition of *Mixed Reality* (MR). We refer readers to Speicher et al. [75] for in-depth discussion around working definitions of MR from the literature by interviewing experts in the area. In this paper, we understand MR as encompassing everything on the Reality-Virtuality continuum, including augmented reality, augmented virtuality, virtual reality and everything in between, according to the definition by Milgram and Kishino's [55, 56], visualized in Figure 2. Based on that, we exclude papers on in-person collaboration and remote collaboration where all collaborators are in reality mediated by video conferencing as they have been extensively investigated by existing research [17, 86, 92]. As a result, we include approaches where at least one collaborator is either in AR or VR.

2.2 Definition: Remote vs Co-located collaboration

Collaborative scenarios can be classified according to where they occur in space: remote collaboration, where users are located in different physical spaces; and co-located collaboration, where all users share the same physical space [53]. Although co-located users can resort to real-world metaphors for collaboration without the need for technological approaches, AR and VR have been used in co-located scenarios for improved immersion and presence [53]. As an example, Herder et al. [28] leveraged avatars in co-located business-oriented applications with a "guide-user-scenario" for large-scale location-based experiences related to virtual product presentation or industrial training with a local group of users. Funk et al. [20] presented HoloCollab for co-located training where physical hardware was substituted by virtual tools rendered in AR within the trainee's environment. This can help in cases where physical tools might be unavailable or too expensive.

Although the use of AR and VR in both remote and co-located scenarios is interesting to explore, we focus on **remote** approaches since these are essential for resilient, scalable, and effective global collaboration. Effective ways of providing assistance and training remotely drastically reduce the environmental impact caused by travelling, and were proven to be essential in scenarios where it is not possible for users to be collocated, e. g., during the COVID-19 pandemic.

2.3 Definition: Remote assistance and training.

Remote assistance and training approaches usually allow experts to help, train, or guide trainees to accomplish some specific task. This typically involves at least one expert and one trainee and bidirectional interaction and communication between parties. Generally and within the context of MR, assistance and training can happen synchronously or asynchronously in time, e. g., through recorded instructions. Here, we understand remote in the sense that one of the users is not physically located in the place where the task is happening. We do not consider assistance or training provided by software without human intervention. Hence, we exclude prerecorded or automated fully immersive VR simulators without any interaction/communication between the parties. We also exclude general-purpose methods for MR (e. g., display devices, rendering techniques, 3D selection techniques, haptic controllers, etc.) that could potentially be used for assistance and training, but are not developed for this context. While we acknowledge that many of the approaches would be applicable for remote assistance and training, we excluded those as they were not tested in the specific contexts. Surveying general purpose methods such as haptic controllers, and identifying their applicability for remote assistance and training would be an interesting extension for future research.

2.4 Research questions

Within the context of this survey, we aim to answer the following research questions.

RQ1: How do users communicate and interact in MR during assistance and training tasks? Due to technical constraints and opportunities, we believe that communication between users is different between physical on-site scenarios and remote MR scenarios. We aim to understand how experts and trainees are enabled to communicate in the challenging conditions of MR environments. More specifically, we want to understand the following sub-questions:

- How many users are enabled to take part?
- How is the workspace shared between users involved?

RQ2: What are the configurations in which remote assistance and training take place? We aim to understand how the remote settings differ from co-located assistance and training in terms of spatial and temporal configuration. More specifically, we want to understand the following sub-questions:

- What are the spatial configurations for users in terms of MR space?
- How does the communication happen in terms of time?

RQ3: How are assistance and training approaches implemented? In order to support assistance and training, we aim to gain an understanding of the technological solutions that are available to experts and trainees, specifically:

- Which interaction methods are available to users?
- Which visual displays and other output technologies are employed?

RQ4: In which domains are remote MR assistance and training used? We aim to identify which domains currently take advantage of remote MR assistance and training and which are underexplored.

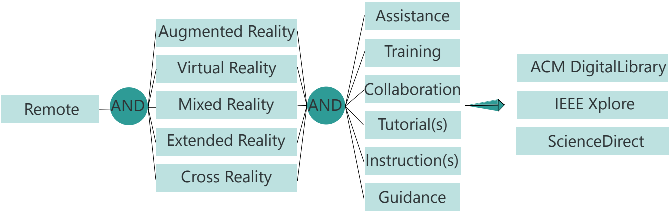


Fig. 3: Keywords included in the individual search strings and the databases that we retrieve the articles from.

3 Methodology

In order to answer our research questions, we conducted a systematic literature review. Our selection process was made according to the PRISMA methodology (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) methodology [58]. We first define the keywords that translate our research, and select which digital libraries to search for publications.

3.1 Keywords

As a basis for the literature search, we included the keywords *Remote*, as well as *Augmented Reality, Virtual Reality, Mixed Reality, Extended Reality* and *Cross Reality*, as those are common instances of MR implementation and synonyms. We then conducted a primary search including only the keywords *Assistance* and *Training*. During the analysis of the results, we found a variety of known relevant works that were not included. We therefore decided to extended the search keywords to include *Collaboration*, *Guidance*, *Tutorial(s)* and *Instruction(s)*, as these are intrinsic to assistance and training. This resulted in the keywords illustrated in Figure 3, that were separated into search strings to look for each main term individually.

3.2 Time frame

Although different surveys on general MR [40, 98], as well as specific aspects such as evaluation in MR [8, 54] typically address the preceding decade of research, we decided to expand on those, as we found it important to not exclude influential papers because of when they were published. Hence, we included articles from the year 2000 to 2022. We did, however, observe a strong increase in activity in the last five to seven years, which goes in line with the commercial availability of headsets such as the Oculus Quest.

3.3 Database and search parameters

We leveraged the keywords to search for publications in the three largest libraries for literature in computer science, specifically *ACM Digital Library*, *IEEE Xplore*, and *Science Direct*, as common for this type of survey. We chose not to include other platforms such as Google Scholar due to the increased noise in the results (e. g., non-archival works, websites, undergraduate theses, duplicates), and since we believe that the most important venues in the field are covered by the three libraries, as described above. Keyword search was performed in the *Title*, *Abstract* and *Author Keywords* of the publications. We searched each of the databases separately, taking into account the inclusion and exclusion criteria described above.

3.4 Inclusion and exclusion criteria

We included articles with English as primary language and focused on archival papers that were peer-reviewed. We did not include non-archival articles such as demos and posters. During the screening of the articles, we included approaches that enabled bidirectional communication between minimum of two humans. We excluded approaches that relied solely on software guidance (e. g., video tutorials) and general-purpose methods that were not developed or tested for remote assistance and training. We focused primarily on venues from the fields of computer graphics, computer vision and human-computer interactions (e. g., ACM CHI, IEEE VR, ACM SIGGRAPH). Venues were manually filtered by the authors after the initial search.

3.5 Selection of relevant articles

The PRISMA flow diagram in the Appendix, Figure A. 1, summarizes the selection process. The first search in all three databases resulted in a raw total of 931 articles. From these we eliminated duplicates, non-archival papers, and papers that did not fit our target venues. This reduced the count to 200 articles, which were screened by title and abstract. In this phase, 112 articles were eliminated. The remaining 88 articles were distributed among the authors, marked for inclusion if appropriate, or discussed if ambiguities arose. In this phase, we eliminated 32 articles that were out of our defined scope, which resulted in a total of 56 articles. Lastly, we added 6 extra articles that did not appear in our systematic approach but are relevant to our research scope, based on author's expertise. This is a typical last step in surveys for extending the number of papers which goes in line with the PRISMA methodology, done by searching the references of the papers found and articles citing the found papers that are relevant [90]. These papers likely did not turn up due to challenges with keyword search in the databases. The final 62 articles were read in full by at least one author, and all relevant data were collected. The research questions were refined, and taxonomy was developed in a holistic discussion-based manner during the data generation.

4 Taxonomy

To answer the research questions, we build a taxonomy based on the information collected from the final 62 papers. The taxonomy consists of eight relevant dimensions. We analyze the paper distribution on each dimension and pair those with qualitative observations, to gain insights about research gaps and opportunities discussed in the remainder of the paper. The final taxonomy is illustrated in Figure 1. In the following, we summarize the individual dimensions and leverage them to answer and discuss the research questions in Section 5. We categorize the individual dimensions whether they are centered around **users** (degree of collaboration, perspective sharing), the **configuration** (MR space symmetry, time), **implementation** (input modality, output modality, visual display), or **application**.

Degree of Collaboration We identify different degrees of collaboration between users as one differentiating dimension between approaches. This includes interaction between 1 expert and 1 trainee only $(1 \leftrightarrow 1)$, 1 expert that interacts with multiple trainees $(1 \leftrightarrow m)$, multiple experts that interact with one trainee alone $(n \leftrightarrow 1)$, and multiple experts that collaborate with multiple trainees at once $(n \leftrightarrow m)$.

Perspective sharing This dimension informs how the workspace is shared between trainees and experts in terms of perspective. We categorize this into first-person view (FPV) or third-person view (TPV). An FPV allows the expert to share the trainee's exact perspective of the workspace, for example, to avoid occlusions and errors in reference frames. A TPV allows the expert to freely explore the workspace, independent of the trainee's perspective. This can be beneficial in terms of communication as it allows seeing the trainee's facial expressions and body posture. For a number of approaches, users can choose or alternate between both perspectives, as discussed below.

Time This dimension identifies whether the environment and information users share are handled synchronously or asynchronously. Synchronous communications happen when users exchange information in real-time. Asynchronous communications involve users' accessing information after it is generated in time, for example, rewinding instructions and replaying sessions.

MR space symmetry This dimension refers to the interaction paradigm within the Reality-Virtuality Continuum [55, 56] in which expert(s) and trainee(s) are located during the task. We consider the MR space as symmetric if users share the same paradigm, i. e., if all are either in AR or VR. Note that we exclude symmetric real-world to real-world interactions, as those typically do not involve any MR approaches. Asymmetric MR spaces include users interacting on different parts of the spectrum (R ↔AR, R ↔ VR, or AR ↔ VR).

Input modality We classify the type of input modality used by experts and trainees to interact with the remote environment. These commonly include tracked controllers, mid-air gestures, touch screens, keyboard and mouse, head gaze, eye gaze, audio communication, or no interaction.

Output modality We identify different types of output modalities that users leverage to communicate with each other, such as visual, audio, haptics, smell, and taste. Note that for both input and output modalities, we only consider approaches that are specific to remote assistance and training. We do not consider general-purpose methods such as haptic controllers or smell & taste interfaces that could be used but were not developed or tested for our context.

Visual display Given the dominance of visual output devices in the MR space, we add visual display as an additional dimension. We identify visual displays used in the different parts of the Reality-Virtuality continuum: real, AR, and VR. Real displays include touch screens, computer screens, or large-scale displays; AR displays include AR handheld displays (phone or tablet), projection-based AR, and head-mounted displays (HMDs); and VR displays include cave automatic virtual environments (CAVEs), i. e., rooms in which each surface is use as projection screens, and VR HMDs.

Application domain We identify application domain as one additional dimension that informs our survey. We categorize the articles into different application domains, including health, sports, education, industry, maintenance, or navigation; as well as general-purpose methods that span multiple domains. This enables us to identify fields where MR is frequently employed, as well as application domains that could benefit from further research.

5 Results

We leverage the proposed taxonomy to answer and contextualize the research questions, as well as identify gaps in the literature and discuss opportunities for future research. Before presenting our findings, we briefly discuss research activity in the field, including publication venue, year, and keyword distribution.

5.1 Publication activity

The 62 papers were published at 24 venues, detailed in the Appendix, Figure 3, with a majority of papers coming from ACM CHI (11), IEEE ISMAR (7), ACM VRST (5), IEEE VR (5), and IEEE TVCG (5). From the distribution of papers per year, a clear upwards trend from 2012 onward is visible, shown in Figure 4. In terms of keywords, we concluded that 57% of the articles were retrieved under the keyword “Collaboration”, which was the sole relevant keyword for 35 papers. Other useful combinations of keywords included "Assistance AND Collaboration" (7%), "Guidance" (7%), and "Collaboration AND Instruction(s)" (5%). None of the articles were retrieved under the keyword Tutorial(s).

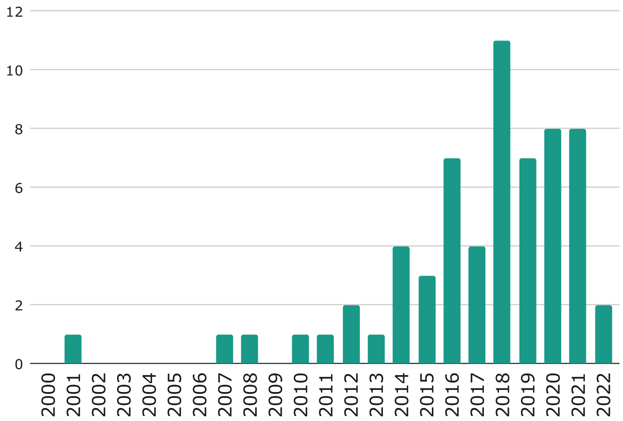


Fig. 4: Paper distribution by year of publication, which illustrates increased activity over the recent years.

		Collaboration Degree			
		1 expert to 1 trainee	1 expert to m trainee	n expert to 1 trainees	n experts to m trainees
Perspective Sharing	FPV	[39] [60] [77] [63] [18] [80] [10] [50] [30] [70] [95] [24] [4] [36] [35] [42] [83] [9] [94] [57] [52] [97] [96] [49] [93] [34] [44] [61] [25] [29] Σ=30	[32] [78] [37] [14] [84] Σ =5	Σ =0	[76] [74] Σ =2
	FPV ↕ TPV	[68] [62] [23] [3] [81] [88] [22] [24] [82] [45] [26] [27] [41] Σ=13	[16] [48] [5] Σ =3	[15] Σ =1	Σ =0
	TPV	[7] [85] [67] [6] [2] [43] [69] [79] Σ =8	Σ =0	Σ =0	Σ =0

Fig. 5: Distribution of articles by Degree of Collaboration and Perspective Sharing. FPV ↔ TPV refers to approaches that enable the transition between both modes.

5.2 User settings

To answer **RQ1** about how users communicate in MR-based remote assistance and training, we focus on the first two dimensions of the proposed taxonomy, *Degree of Collaboration* and *Perspective Sharing*, visualized in Figure 5.

For *Degree of Collaboration*, 82% of the articles adopted 1 ↔ 1 setting, meaning one expert was interacting with one trainee. 15% of articles employed a $1 \leftrightarrow m$ setting. Within the body of work we included in the survey, only one article dealt with providing assistance from many experts to one trainee ($n \leftrightarrow 1$) by Clergeaud et al. [15]. Two articles leveraged many-to-many interactions ($n \leftrightarrow m$). For $n \leftrightarrow 1$ scenario, Clergeaud et al. [15] enable a team of aerospace industrial design experts to assist one technician immersed in VR. The closest example related to $n \leftrightarrow m$ scenario is the work by Sra et al. [76]. They implemented a dance lesson experience where each trainee in one shared space had their own personal dance instructor, visible only to them and placed behind their dance partner. Expert instructors were identifiable by a spotlight on them, and trainees could follow their movements by facing their backs. Users could only look at their own instructor and thus were not able to take advantage of the expertise of the other dance instructors. Based on the dominance of 1 ↔ 1 settings in prior work, we believe that there is an open space for research related to assistance and training in MR that enables more than one expert and one trainee to interact. Enabling teams of multiple experts to help a small to a big group of trainees in their tasks, for example, constitutes an important case, since different expertise in different domains might be needed.

Regarding the *perspective of the workspace*, most articles enable a first-person view (FPV) for the expert (60%), while 13% of articles enable a third-person view (TPV); see Figure 6. As an example of FPV expert view, the ReMa system by Feick et al. [19] reproduces the orientation manipulations of an object so that the person manipulating the physical object and their remote collaborator seeing the object's proxy share the same perspective on the object. Works enabling FPV for the expert show that a shared point of view of FPV can be effective and is preferred compared to opposing point of view, due to decreased cognitive load in understanding spacial references [47] and avoiding occlusions [73]. We believe that this is one main factor for why the majority of works employ this perspective. It is worth noting, however, that when using TPV, experts can see partner's facial expressions and body posture [33, 91], which provides insights on task comprehension and can be advantageous for communication. For example, in Piumsomboon et al. [66], non-verbal communication and awareness cues such as gaze direction and body gestures were shared in TPV through embodiment into a miniature or giant avatar. Furthermore, sharing FPV requires to stabilize the view since the frequent changes caused by the trainee's head movements can hinder the observation by the expert [50].

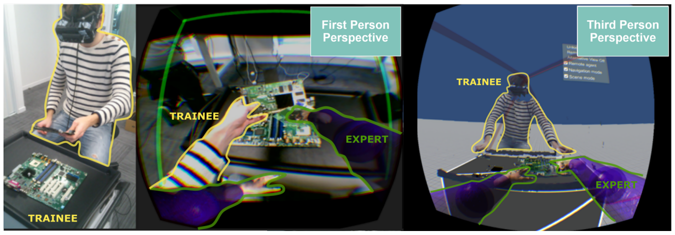


Fig. 6: An example of enabling a first-person view (FPV) or a third-person view (TPV) for the expert regarding the perspective of the trainee's workspace. Image from Le Chenecha et al. [47].

A number of articles (27%) enable experts to choose between FPV and TPV, depending on the situation at hand. For example, Speicher et al. [74] provide users with a third-person view by default but enable one collaborator at a time to gain control over everyone's 360° video. This enables all collaborators to view directions in a synchronized manner. Similarly, Teo et al. [81] enable experts to interact with trainees with a TPV by default, allowing them to move independently from the trainees in a reconstructed VR environment. Additionally, experts can decide to immerse themselves into the same point of view, live-streamed through a 360° camera attached to the trainee's head.

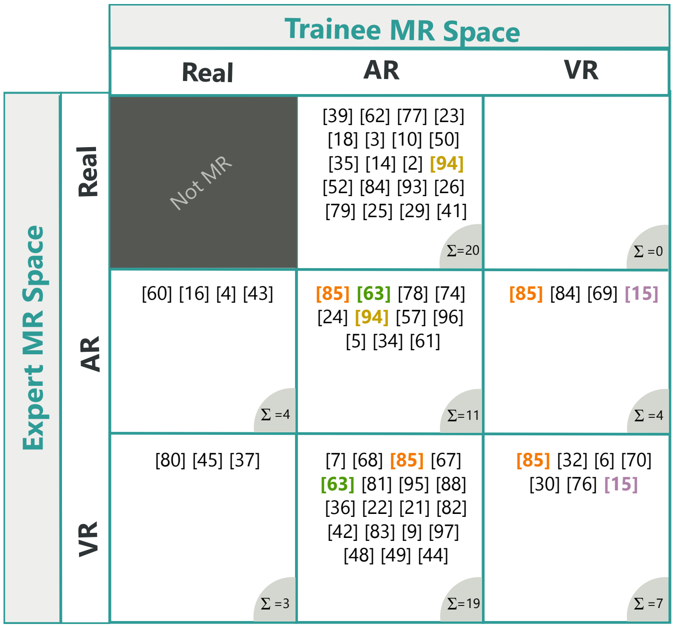
These results indicate that a large majority of works focus on FPV settings, enabling one expert to interact with one or multiple trainees. While a number of works enable users to transition between different perspectives, we believe that especially the multi-user setting affords more exploration in view representations beyond FPV. Giving one or more experts the ability to observe and interact with multiple trainees through TPV would enable a more flexible environment, for example.

5.3 Spatio-temporal configurations

We aim to answer the question of what types of configuration in terms of space and time prior work on remote assistance and training has explored (**RQ2**). To gather further insights, we leverage the dimensions of *time* and *MR space symmetry* from the proposed taxonomy. Figure 7 details into the symmetry of interaction space for experts and trainees.

In terms of *Time*, most papers adopt the full synchronous setup in terms of time configuration. Only three articles (5%) enable users to interact both synchronously and asynchronously. They allowed users to review previous performance or revisit missing details. As examples, Kumaravel et al. [85] recorded 3D and 2D videos of both the expert and the trainee so that they could review the real-time performances. Speicher et al. [74] allowed collaborators to view and annotate a 360° live stream, including the ability to rewind the live stream by 10 seconds in case they missed important information, saving all digital artifacts in a session at their exact positions (including all annotations and projections). Collaborators can also reload a session at a later point in time for asynchronous annotation. All other works focus on communication and interactions that are delivered without delay, or without giving user the ability to record or review prior actions. We believe that the focus on synchronous configurations results from the importance of direct communication and enabling experts to interact and correct trainees during a task. Considering the need of people across different time zones for remote assistance and training, however, we argue that asynchronous communication methods are a relevant topic of research that is currently under-explored. Enabling experts to interact in a direct yet asynchronous manner would enable overcoming some of the main barriers to remote assistance, training, and collaboration in general. This, however, is very challenging to achieve, as engagement oftentimes is connected to direct communication. While recent works [74, 85] start exploring short-time asynchronous communication, we believe that similar interaction could be useful to go beyond a replay of a few minutes.

For *MR Space Symmetry*, more than 80% of the papers employ asymmetrical interaction spaces in the MR Reality-Virtuality Continuum, illustrated in Figure 7. We observe that the most common setting for



Expert MR Space \ Trainee MR Space	Real	AR	VR
Real	Not MR	[39] [62] [77] [23] [18] [3] [10] [50] [35] [14] [2] [94] [52] [84] [93] [26] [79] [25] [29] [41] Σ=20	Σ =0
AR	[60] [16] [4] [43] Σ =4	[85] [63] [78] [74] [24] [94] [57] [96] [5] [34] [61] Σ=11	[85] [84] [69] [15] Σ =4
VR	[80] [45] [37] Σ =3	[7] [68] [85] [67] [63] [81] [95] [88] [36] [22] [21] [82] [42] [83] [9] [97] [48] [49] [44] Σ=19	[85] [32] [6] [70] [30] [76] [15] Σ =7

Fig. 7: Distribution of articles by MR Space Symmetry, i. e., in which part of the Reality-Virtuality Continuum **expert(s)** and **trainee(s)** are located during the task.

the trainee is an AR-based interface (80%). We believe it is under practical consideration that trainees need to directly see and interact with the physical environment where they need assistance or training. For training purposes, there exist extensions of such approaches that happen in fully virtual environments, especially when training in the physical environment is not possible. VR-based trainee interfaces were also common when the task did not require information in the physical environment, for example, in physical motion learning such as dance [76] and Taichi [13].

As for experts, most works enable them to interact from a desktop computer (32%), in AR (32%), or immersed in VR (45%). Some approaches enable experts to cross the space and choose between different points of the spectrum, alternating between the real world and AR [94], or AR and VR [15,63,85]. Additionally, Kumaravel et al. [85] enable the trainee to choose between AR and VR. We believe VR has great potential for experts to provide assistance and training since they can be immersed in a total reconstruction of the trainees' space, facilitating the understanding of the task at hand (e. g., Piumsomboon et al. [67]). We believe that there exists opportunity for future work to enable experts to communicate with trainees beyond 2D screens. Enabling immersive experiences with VR or AR would potentially hold benefits in terms of interaction and presence. As an example, Hoang et al. [30] developed OneBody, a VR system for physical activity training. Experts in VR can see an overlay of the trainee's body movements onto their own body, enabling them to check for mismatched body movements. A comparison between OneBody with training provided through a video screen indicated better posture accuracy in delivering movement instructions when in VR. This work highlights the potential of transitioning instructions from 2D into the 3D space.

5.4 Input and output modalities

We discuss how assistance and training approaches are implemented (**RQ3**) with the data on dimensions of *Input Modality*, *Output Modality*, and *Visual Display*.

We first analyze the input methods that experts and trainees used to convey their interaction intents, illustrated in Figure 8. We highlight that the most common input methods for the expert are controllers, mid-air gestures, and a keyboard and mouse. Several articles leverage other technologies such as touchscreens (10) or eye gaze (2). The expert also commonly relies on audio communication to deliver the instructions. For trainees, the most common setting is no explicit interaction with the expert, as they primarily interact with their physical environment. Other common input modalities for trainees include mid-air gestures, controllers and touchscreens. It is worth mentioning that gaze is rarely used as an input modality despite the emerging eye-tracking capabilities of devices (e. g., VR headsets).

As for the output modalities, the most commonly leveraged channels are visual and auditory, illustrated in Figure 9. Few other works applied haptic modalities [4,6,43,94], while none of the papers applied smell or taste as output. For example, Audaet al. [6] enhanced the communication between remote collaborators by using haptic props that make virtual objects graspable and investigated the ownership sharing mechanism of the virtual objects across users. As discussed earlier, however, our survey focused on articles dedicated to remote assistance and training, no general purpose methods that *could* be used for this context.

There exists several implementations of input and output modalities that aim at enhancing co-presence. For example, Kumaravel et al. [85] render the expert's relative position from which they observe the trainee's space in the trainee's view to enhance the awareness of each other's location. Speicher et al. [74] enhance gaze awareness display individual colored cones that indicate where each collaborator is looking. These implementations enable a higher sense of presence than the approaches where the trainee does not have access to information about the expert, or those only displaying visual annotations (e. g., [3,10,16,18,23,39,50,60]).

Figure 10 illustrates the visual display provided to experts and trainees. Most papers immersed the expert in VR (45%), followed by letting the expert interact with the trainee through a computer screen (31%), and through mobile or HMD AR (11% and 8% respectively). For trainees, the most common setup was either using an AR HMD (50%), mobile AR (21%), followed by a VR HMD (19%). The common use of an HMD is to free both hands to execute required tasks during the training or guided activities. We also hypothesize that a lot of works resort to a tablet or a smartphone (mobile AR) due to this type of display's availability to the general public and low cost. Other than relying on a single configuration, existing papers explore to enable users to switch between different devices. For example, Kumaravel et al. [85] allow users to connect, and resort to asymmetric technological configurations in various degrees of immersion, from different parts of the Reality-virtuality continuum, depending on what devices users have available.

In summary, a majority of works enable experts and trainees to interact through an audio channel, but focus largely on enabling experts to communicate with trainees. Few works provide trainees with advanced interaction methods such as gestures to explicitly communicate with experts. Most current systems focus on situations where experts are equipped with 2D screens, while trainees are in AR. We believe this highlights interesting opportunities for symmetric AR communication, or providing experts with immersive views of trainees and their environment through VR, for example, especially for scenarios involving multiple trainees.

5.5 Application domains

As final research question, we aimed to identify the application domains where MR has been applied in the context of remote assistance and training (**RQ4**), illustrated in Figure 11. The majority of articles (69%) are related to multidomain approaches that are applicable to multiple fields, which highlights the search for universal solutions. This is followed by MR solutions for remote maintenance and industrial tasks (18%). Within this space, there exists specific applications for performing repair and assembly tasks [10,15,18,24,26,27,52,63,78,88] and in production line planning [5]. Besides industrial assistance, there exists a body of work on medical training and assistance. For example, Ansar et al. [4] developed a prototype to enable the trainee surgeon to operate on a virtual patient while receiving real-time visual and haptic feedback on tissue properties from the remote site.

Other specific areas of application include the field of navigation [9,44,61], education [34,77,84], and sports [30]. For example, ObserVAR

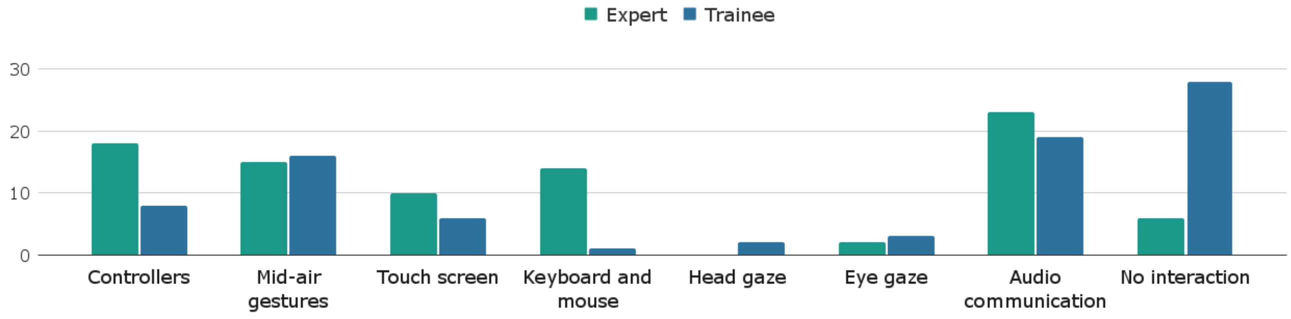


Fig. 8: Comparison between the types of input modality enabled for the expert to interact with the trainee's space (green) and for the trainee to interact with the expert's space (blue). The numbers of articles that enable each input modality are visualized.

system [84] allows the course instructor to guide visual focus of remote students to specific contents that they should notice. From these, we found that training approaches mostly relate to the health and sports domain, while assistive technologies are more related to the industry. Note that by analyzing the articles we collected through our systematic search, only 14% related to remote training alone, with other 61% of articles relating to remote assistance, and 25% relating to approaches that are both useful for assistance and training. We hypothesize that most remote training constitutes an educational factor. A research opportunity lies here due to the different requirements for training in education. Furthermore, there exists various examples of using AR for assistive surgery, or teleportation of robots. Our survey highlights opportunities for applying assistance to domains such as healthcare (e. g., remote physiological assistance) or industrial training.

6 Limitations

In our work, we aim to provide an overview of the field of remote assistance and training enabled through MR. Within the context of the presented survey, we focus mostly on professional settings, rather than educational settings such as the usage of MR in schools (e. g., Holstein et al. [31]). We believe that the professional setting is different as the requirements on technology and guidance between experts and trainees differ compared to teachers and students. Investigating the usage of MR in a classroom setting, however, is an interesting direction for further work. Additionally, we wanted to keep the survey self contained and focused, by providing a broader taxonomy for this topic. However, we acknowledge that our taxonomy could be extended to dig deeper into other interesting aspects such as the level of expertise of the users involved, the task's difficulty, the time frame of the task, or the intellectual abilities the task draws upon. Creating a taxonomy that expands on this extra layer of information remains as future work.

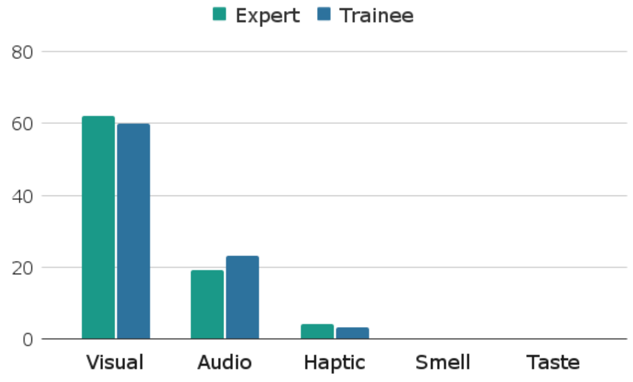


Fig. 9: Comparison between the type of output modality enabled for the expert (green) and the trainee (blue). The numbers of articles that enable each output modality are visualized.

We focus our search on the three main databases within the context of computer science research, as we are interested in innovative approaches within human-computer interaction, computer graphics, etc. This is similar to other research in the fields of cross-device interaction [11] or VR [1], for example. While those databases also cover specialized engineering disciplines, we cannot guarantee completeness in the articles that are covered. We believe, however, that our survey covers the majority of relevant articles in the field and provides insights and categorization of the benefits, challenges, and opportunities.

We focus our search on works that enable users to collaborate remotely though AR and VR, excluding approaches for co-located assistance and training. We acknowledge that we could have surveyed methods for assistance and training in AR and VR as a whole, including both remote and co-located methods as an additional dimension. However, we were interested in finding new directions for collaboration in this sub-area specifically remote, since the ability to perform these tasks remotely is essential for resilient, scalable, and effective collaboration. Effective ways of providing assistance and training remotely could reduce the environmental impact caused by air travel, and prove to be helpful in scenarios where it is not possible for users to be co-located, e. g., a pandemic situation.

We constructed the proposed taxonomy based on the articles we identified to be relevant. This process is influenced by the assumption that areas of research that have seen more work (i. e., a larger number of articles) were seen as promising directions to contribute significantly to increasing the quality of remote assistance and training. Conversely, we identified areas of research that were less well covered by prior work. Those areas are potentially promising but have not been well investigated and present opportunities for further exploration; or solutions within these spaces might not lead to significant improvements for remote assistance and training. Through qualitative analysis and discussion, we aimed to differentiate those two areas, for example, by taking the discussions of various articles into account. We believe that the presented taxonomy and discussion enable researchers to judge the value of under-explored areas, and showcase opportunities for future exploration.

7 Opportunities for Future Work

The main objective of our survey was to build a taxonomy for remote assistance and training in MR environments, identifying overall trends, under-explored research topics, and benefits and challenges of existing work. In the following, we discuss some opportunities for future work that we found relevant through our analysis. We set those opportunities into the context of the discussed taxonomy by considering which areas are less well explored, and highlight individual works that start to fill the gaps.

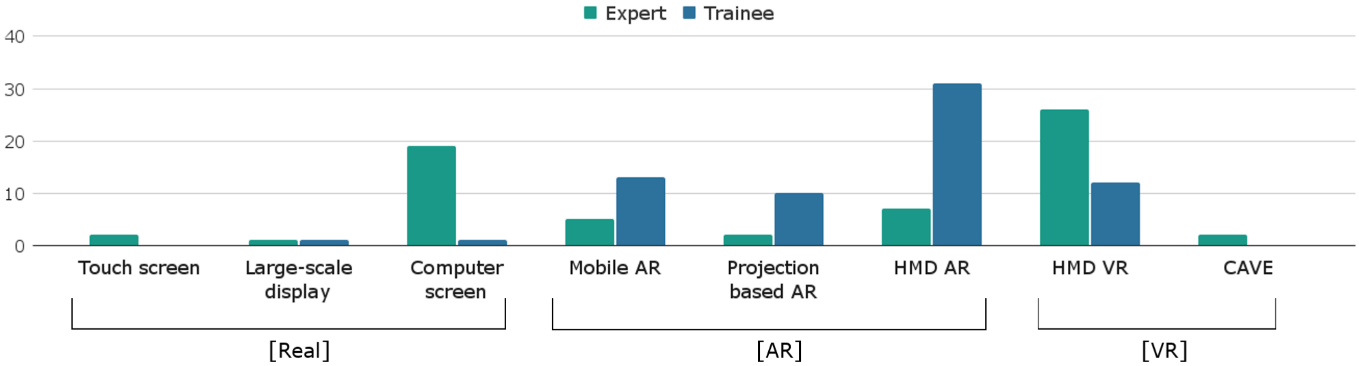
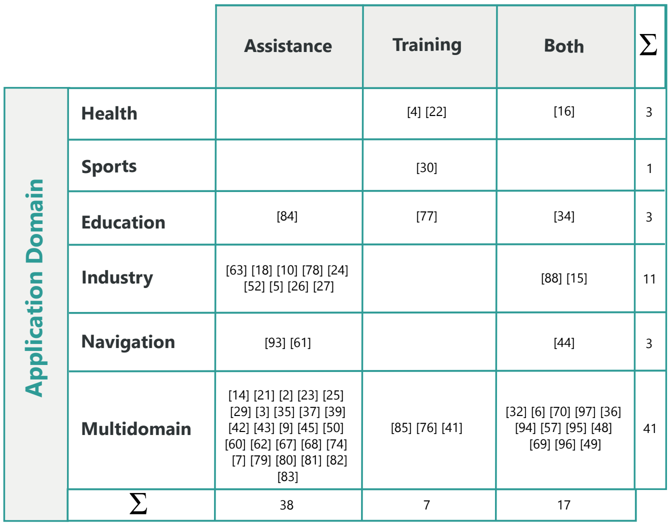


Fig. 10: Comparison between the types of visual displays that the expert (green) and the trainee (blue) use in different parts of Reality-Virtuality continuum. The numbers of articles that use each display are visualized.



		Assistance	Training	Both	Σ
Application Domain	Health		[4] [22]	[16]	3
	Sports		[30]		1
	Education	[84]	[77]	[34]	3
	Industry	[63] [18] [10] [78] [24] [52] [5] [26] [27]		[88] [15]	11
	Navigation	[93] [61]		[44]	3
	Multidomain	[14] [21] [2] [23] [25] [29] [3] [35] [37] [39] [42] [43] [9] [45] [50] [60] [62] [67] [68] [74] [7] [79] [80] [81] [82] [83]	[85] [76] [41]	[32] [6] [70] [97] [36] [94] [57] [95] [48] [69] [96] [49]	41
	Σ	38	7	17	

Fig. 11: Distribution of the articles by the application domains.

Degree of collaboration. The majority of current work focuses on 1 trainee ↔ 1 expert collaboration. We believe that it is important to explore degree of collaborations beyond this modality, especially for multiple experts to interact with multiple trainees as this scales to include the maximum of people. Works such as [76] showcase the potential of $m \leftrightarrow n$ collaboration for scenarios such as multi-trainee dance training. Enabling teams of various experts to help a small to a big group of trainees in their tasks constitutes an important case, since different expertise in different domains might be needed. Besides added benefits in terms of training, scaling training to beyond one trainee would have potential benefits in terms of availability and costs. For surgical training, for example, providing guidance for multiple trainees could increase efficiency and reduce training time for medical professionals.

Perspective sharing. Both first-person and third-person perspectives contribute in different ways to the overall remote experience (task comprehension *vs.* presence). While first-person can lead to increased body ownership and shared viewpoints [47], third-person can enable an improved overview of the surrounding space and enable experts to view trainees' posture and facial expression [91], for example. The benefit of such flexible approaches has been shown in work such as by Kumaravel et al. [85]. We thus believe that exploring mixed perspective approaches where the expert can transition between different points of view depending on which better fits the task at the moment would potentially enrich remote assistance and training scenarios.

MR space symmetry. Device availability and appropriateness is highly task-dependent. This is evident in the fact that a majority of research uses asymmetric approaches, i. e., experts and trainees use different modalities. Most approaches employ devices statically, meaning that users are only interacting with one device during one session. We believe that dynamic device assignments are an interesting direction for further research, effectively enabling users to move across the MR spectrum. Different devices lead to different degrees of immersion and few approaches embrace the dynamic nature of interaction. This approach, however, typically requires larger efforts in infrastructure and tailored interactions (e. g., as in Loki [85]), which might be a prohibitive factors for many researchers that can only be overcome through more mature and readily-available software and hardware platforms.

Time. We believe the integration of asynchronous communication mechanisms for assistance and training approaches to be worth exploring for future work. Although synchronous interaction enables effective and efficient communication, being able to collaborate from different time zones or even schedules is also important. Additionally, asynchronous mechanisms such as the ability to rewind or re-watch a collaborative session bring added value to the learning process of training tasks. This type of interaction would be particularly powerful for teams that are distributed across time zones. Finding a sweet-spot between replaying past actions, and engaging feedback and interactions, however, we believe is challenging to implement.

Input modality. Experts rely heavily on additional tools to perform the communication (e. g., pointers, sketches, arrows) through using controllers, a keyboard and mouse set-up, or mid-air gestures. We believe that current systems focus on those devices since they are readily available and offer stable and accurate performance. Exploring more advanced interaction techniques for communicating intent and actions to remote trainees, e. g., by taking advantage of hand tracking, eye tracking or speech recognition, will open rich additional information channels and is an interesting area for further work. While general purpose methods exist, tailoring those specifically to assistance and training scenarios will surface interesting benefits and challenges. Johnson et al. [35], for example, demonstrate that hand gesture-based referencing is beneficial for collaborative physical tasks.

Output modality. Visual and auditory stimuli dominate current interactive systems. While those are effective in delivering information to the users, we believe that utilizing other sensory stimuli, such as smell, taste, and haptics while interacting can increase the sense of presence for users and provide extra feedback to experts and trainees. Current research focuses on general-purpose methods, rather than the special requirements of training and simulation. We therefore believe that there is a large research space for this type of output that could be brought into the context of assistance and training, which in turn will lead to interesting challenges and opportunities to advance immersion, engagement, realism and fidelity of collaboration.

8 Conclusion

In today's world, there is an increasing need for systems that support remote assistance and training. MR technologies offer essential functionalities to improve remote assistance and training in various domains. We conducted a systematic search across different databases that resulted in the identification of 62 papers related to remote assistance and training in MR environments. Through the analysis of these papers, we created a taxonomy for the characterization along eight dimensions of degree of collaboration, perspective sharing, MR space symmetry, time, input and output modality, visual display, and application domain. Based on the paper distribution along these dimensions, we discuss the overall research trends, and identify under-explored research topics as opportunities for future research. Researchers could explore more advanced interaction techniques specific to this scenario by taking advantage of hand tracking for example, enable higher degrees of collaboration, or enable users to move across the reality virtuality continuum depending on what better fits the task. We hope this survey can help other researchers build stronger approaches for remote assistance and training by taking full advantage of current and future MR technology.

Supplemental Materials

All supplemental materials and documents are available at the URL `https://augmented-perception.org/publications/2023-training-survey.html`. In particular, they include (1) Excel files containing the raw survey data used for creating our Taxonomy, and (2) a full version of this paper with all appendices.

Acknowledgments

This work is co-financed by Fundação para a Ciência e a Tecnologia (Portuguese Foundation for Science and Technology) partially through grants 2022.09212.PTDC (XAVIER), UIDB/50021/2020 and the Carnegie Mellon/Portugal Program fellowship SFRH/BD/151465/2021, under the UNESCO Chair on AI&XR.

References

[1] P. Abtahi, S. Q. Hough, J. A. Landay, and S. Follmer. Beyond being real: A sensorimotor control perspective on interactions in virtual reality. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems*, CHI '22. Association for Computing Machinery, New York, NY, USA, 2022. doi: 10.1145/3491102.3517706 7

[2] M. Adcock, S. Anderson, and B. Thomas. Remotefusion: real time depth camera fusion for remote collaboration on physical tasks. In *Proceedings of the 12th ACM SIGGRAPH international conference on virtual-reality continuum and its applications in industry*, pp. 235–242, 2013. 14

[3] M. Adcock, D. Ranatunga, R. Smith, and B. H. Thomas. Object-based touch manipulation for remote guidance of physical tasks. In *Proceedings of the 2nd ACM symposium on Spatial user interaction*, pp. 113–122, 2014. 6, 14

[4] A. Ansar, D. Rodrigues, J. P. Desai, K. Daniilidis, V. Kumar, and M. F. Campos. Visual and haptic collaborative tele-presence. *Computers & Graphics*, 25(5):789–798, 2001. 6, 14

[5] D. Aschenbrenner, M. Li, R. Dukalski, J. Verlinden, and S. Lukosch. Collaborative production line planning with augmented fabrication. In *2018 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*, pp. 509–510. IEEE, 2018. 6, 14

[6] J. Auda, L. Busse, K. Pfeuffer, U. Gruenefeld, R. Rivu, F. Alt, and S. Schneegass. I'm in control! transferring object ownership between remote users with haptic props in virtual reality. In *Symposium on Spatial User Interaction*, pp. 1–10, 2021. 6, 14

[7] H. Bai, P. Sasikumar, J. Yang, and M. Billinghurst. A user study on mixed reality remote collaboration with eye gaze and hand gesture sharing. In *Proceedings of the 2020 CHI conference on human factors in computing systems*, pp. 1–13, 2020. 14

[8] Z. Bai and A. F. Blackwell. Analytic review of usability evaluation in ismar. *Interacting with Computers*, 24(6):450–460, 2012. 3

[9] T. Bednarz, C. James, C. Caris, K. Haustein, M. Adcock, and C. Gunn. Applications of networked virtual reality for tele-operation and tele-assistance systems in the mining industry. In *Proceedings of the 10th International Conference on Virtual Reality Continuum and Its Applications in Industry*, pp. 459–462, 2011. 6, 14

[10] S. Bottecchia, J.-M. Cieutat, and J.-P. Jessel. Tac: augmented reality system for collaborative tele-assistance in the field of maintenance through internet. In *Proceedings of the 1st augmented human international conference*, pp. 1–7, 2010. 6, 14

[11] F. Brudy, C. Holz, R. Rädle, C.-J. Wu, S. Houben, C. N. Klokmose, and N. Marquardt. Cross-device taxonomy: Survey, opportunities and challenges of interactions spanning across multiple devices. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, CHI '19, p. 1–28. Association for Computing Machinery, New York, NY, USA, 2019. doi: 10.1145/3290605.3300792 7

[12] J. Carmigniani, B. Furht, M. Anisetti, P. Ceravolo, E. Damiani, and M. Ivkovic. Augmented reality technologies, systems and applications. *Multimedia tools and applications*, 51(1):341–377, 2011. 2

[13] X. Chen, Z. Chen, Y. Li, T. He, J. Hou, S. Liu, and Y. He. Immertai: Immersive motion learning in vr environments. *Journal of Visual Communication and Image Representation*, 58:416–427, 2019. 6

[14] M. Cidota, S. Lukosch, D. Datcu, and H. Lukosch. Workspace awareness in collaborative ar using hmds: a user study comparing audio and visual notifications. In *Proceedings of the 7th Augmented Human International Conference 2016*, pp. 1–8, 2016. 14

[15] D. Clergeaud, J. S. Roo, M. Hachet, and P. Guitton. Towards seamless interaction between physical and virtual locations for asymmetric collaboration. In *Proceedings of the 23rd ACM Symposium on Virtual Reality Software and Technology*, pp. 1–4, 2017. 5, 6, 14

[16] M. C. Davis, D. D. Can, J. Pindrik, B. G. Rocque, and J. M. Johnston. Virtual interactive presence in global surgical education: international collaboration through augmented reality. *World neurosurgery*, 86:103–111, 2016. 6, 14

[17] S. De Freitas and T. Neumann. Pedagogic strategies supporting the use of synchronous audiographic conferencing: A review of the literature. *British Journal of Educational Technology*, 40(6):980–998, 2009. 2

[18] V. Domova, E. Vartiainen, and M. Englund. Designing a remote video collaboration system for industrial settings. In *Proceedings of the Ninth ACM International Conference on Interactive Tabletops and Surfaces*, pp. 229–238, 2014. 6, 14

[19] M. Feick, T. Mok, A. Tang, L. Oehlberg, and E. Sharlin. Perspective on and re-orientation of physical proxies in object-focused remote collaboration. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, p. 281. ACM, 2018. 5

[20] M. Funk, M. Kritzler, and F. Michahelles. Holocollab: a shared virtual platform for physical assembly training using spatially-aware head-mounted displays. pp. 1–7, 10 2017. doi: 10.1145/3131542.3131559 3

[21] L. Gao, H. Bai, M. Billinghurst, and R. W. Lindeman. User behaviour analysis of mixed reality remote collaboration with a hybrid view interface. In *32nd Australian Conference on Human-Computer Interaction*, pp. 629–638, 2020. 14

[22] D. Gasques, J. G. Johnson, T. Sharkey, Y. Feng, R. Wang, Z. R. Xu, E. Zavala, Y. Zhang, W. Xie, X. Zhang, et al. Artemis: A collaborative mixed-reality system for immersive surgical telementoring. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*, pp. 1–14, 2021. 14

[23] S. Gauglitz, B. Nuernberger, M. Turk, and T. Höllerer. In touch with the remote world: Remote collaboration with augmented reality drawings and virtual navigation. In *Proceedings of the 20th ACM Symposium on Virtual Reality Software and Technology*, pp. 197–205, 2014. 6, 14

[24] S. Gauglitz, B. Nuernberger, M. Turk, and T. Höllerer. World-stabilized annotations and virtual scene navigation for remote collaboration. In *Proceedings of the 27th annual ACM symposium on User interface software and technology*, pp. 449–459, 2014. 6, 14

[25] K. Gupta, G. A. Lee, and M. Billinghurst. Do you see what i see? the effect of gaze tracking on task space remote collaboration. *IEEE transactions on visualization and computer graphics*, 22(11):2413–2422, 2016. 14

[26] P. Gurevich, J. Lanir, and B. Cohen. Design and implementation of teleadvisor: a projection-based augmented reality system for remote collaboration. *Computer Supported Cooperative Work (CSCW)*, 24(6):527–562, 2015. 6, 14

[27] P. Gurevich, J. Lanir, B. Cohen, and R. Stone. Teleadvisor: a versatile augmented reality tool for remote assistance. In *Proceedings of the SIGCHI conference on human factors in computing systems*, pp. 619–622, 2012. 6, 14

[28] J. Herder, N. Brettschneider, J. de Mooij, and B. Ryskeldiev. Avatars for

co-located collaborations in hmd-based virtual environments. In *2019 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*, pp. 968–969, 2019. doi: 10.1109/VR.2019.8798132 2

[29] K. Higuch, R. Yonetani, and Y. Sato. Can eye help you? effects of visualizing eye fixations on remote collaboration scenarios for physical tasks. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*, pp. 5180–5190, 2016. 14

[30] T. N. Hoang, M. Reinoso, F. Vetere, and E. Tanin. Onebody: remote posture guidance system using first person view in virtual environment. In *Proceedings of the 9th Nordic Conference on Human-Computer Interaction*, pp. 1–10, 2016. 6, 14

[31] K. Holstein, B. M. McLaren, and V. Aleven. Intelligent tutors as teachers' aides: Exploring teacher needs for real-time analytics in blended classrooms. In *Proceedings of the Seventh International Learning Analytics & Knowledge Conference*, LAK '17, p. 257–266. Association for Computing Machinery, New York, NY, USA, 2017. doi: 10.1145/3027385.3027451 7

[32] A. H. Hoppe, F. van de Camp, and R. Stiefelhagen. Shisha: enabling shared perspective with face-to-face collaboration using redirected avatars in virtual reality. *Proceedings of the ACM on Human-Computer Interaction*, 4(CSCW3):1–22, 2021. 14

[33] H. Ishii, M. Kobayashi, and K. Arita. Iterative design of seamless collaboration media. *Communications of the ACM*, 37(8):83–97, 1994. 5

[34] D. Iwai, R. Matsukage, S. Aoyama, T. Kikukawa, and K. Sato. Geometrically consistent projection-based tabletop sharing for remote collaboration. *IEEE Access*, 6:6293–6302, 2017. 6, 14

[35] J. G. Johnson, D. Gasques, T. Sharkey, E. Schmitz, and N. Weibel. Do you really need to know where "that" is? enhancing support for referencing in collaborative mixed reality environments. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*, pp. 1–14, 2021. 8, 14

[36] B. Jones, Y. Zhang, P. N. Wong, and S. Rintel. Belonging there: Vrooming into the uncanny valley of xr telepresence. *Proceedings of the ACM on Human-Computer Interaction*, 5(CSCW1):1–31, 2021. 14

[37] S. Kasahara, S. Nagai, and J. Rekimoto. Jackin head: Immersive visual telepresence system with omnidirectional wearable camera. *IEEE transactions on visualization and computer graphics*, 23(3):1222–1234, 2016. 14

[38] M. Kenoui. Telemedicine meets augmented reality: Healthcare services delivery and distance training. In *2020 4th International Symposium on Informatics and its Applications (ISIA)*, pp. 1–5. IEEE, 2020. 2

[39] C. Kervegant, J. Castet, J. Vauchez, and C. Bailly. Distant assist cursor (dac): Designing an augmented reality system to facilitate remote collaboration for novice users. In *Interactive Surfaces and Spaces*, pp. 8–11. 2021. 6, 14

[40] K. Kim, M. Billinghurst, G. Bruder, H. B.-L. Duh, and G. F. Welch. Revisiting trends in augmented reality research: A review of the 2nd decade of ismar (2008–2017). *IEEE transactions on visualization and computer graphics*, 24(11):2947–2962, 2018. 3

[41] S. Kim, M. Billinghurst, and G. Lee. The effect of collaboration styles and view independence on video-mediated remote collaboration. *Computer Supported Cooperative Work (CSCW)*, 27(3):569–607, 2018. 14

[42] S. Kim, G. Lee, W. Huang, H. Kim, W. Woo, and M. Billinghurst. Evaluating the combination of visual communication cues for hmd-based mixed reality remote collaboration. In *Proceedings of the 2019 CHI conference on human factors in computing systems*, pp. 1–13, 2019. 14

[43] S. Kim and J. Park. Collaborative haptic exploration of dynamic remote environments. *IEEE Computer Graphics and Applications*, 38(5):84–99, 2018. 6, 14

[44] J. Kolkmeier, E. Harmsen, S. Giesselink, D. Reidsma, M. Theune, and D. Heylen. With a little help from a holographic friend: The openimpress mixed reality telepresence toolkit for remote collaboration systems. In *Proceedings of the 24th ACM Symposium on Virtual Reality Software and Technology*, pp. 1–11, 2018. 6, 14

[45] R. Komiyama, T. Miyaki, and J. Rekimoto. Jackin space: designing a seamless transition between first and third person view for effective telepresence collaborations. In *Proceedings of the 8th Augmented Human International Conference*, pp. 1–9, 2017. 14

[46] J.-F. Lapointe, H. Molyneaux, and M. S. Allili. A literature review of ar-based remote guidance tasks with user studies. In *International Conference on Human-Computer Interaction*, pp. 111–120. Springer, 2020. 2

[47] M. Le Chénéchal, T. Duval, V. Gouranton, J. Royan, and B. Arnaldi. Vishnu: virtual immersive support for helping users an interaction paradigm for collaborative remote guiding in mixed reality. In *2016 IEEE Third VR International Workshop on Collaborative Virtual Environments (3DCVE)*, pp. 9–12. IEEE, 2016. 5, 8

[48] G. Lee, H. Kang, J. Lee, and J. Han. A user study on view-sharing techniques for one-to-many mixed reality collaborations. In *2020 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*, pp. 343–352. IEEE, 2020. 14

[49] G. A. Lee, T. Teo, S. Kim, and M. Billinghurst. A user study on mr remote collaboration using live 360 video. In *2018 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*, pp. 153–164. IEEE, 2018. 14

[50] C. Lin, E. Rojas-Munoz, M. E. Cabrera, N. Sanchez-Tamayo, D. Andersen, V. Popescu, J. A. B. Noguera, B. Zarzaur, P. Murphy, K. Anderson, et al. How about the mentor? effective workspace visualization in ar telementoring. In *2020 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*, pp. 212–220. IEEE, 2020. 5, 6, 14

[51] T. J. Marion and S. K. Fixson. The transformation of the innovation process: How digital tools are changing work, collaboration, and organizations in new product development. *Journal of Product Innovation Management*, 38(1):192–215, 2021. 2

[52] B. Marques, S. Silva, P. Dias, and B. Sousa-Santos. Does size matter? exploring how standard and large-scale displays affect off-site experts during ar-remote collaboration. In *Proceedings of the 2022 International Conference on Advanced Visual Interfaces*, pp. 1–3, 2022. 6, 14

[53] K. Marriott, F. Schreiber, T. Dwyer, K. Klein, N. H. Riche, T. Itoh, W. Stuerzlinger, and B. H. Thomas, eds. *Immersive Analytics*, vol. 11190 of *Lecture Notes in Computer Science*. Springer, 2018. doi: 10.1007/978-3-030-01388-2 2

[54] L. Merino, M. Schwarzl, M. Kraus, M. Sedlmair, D. Schmalstieg, and D. Weiskopf. Evaluating mixed and augmented reality: A systematic literature review (2009-2019). In *2020 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*, pp. 438–451. IEEE, 2020. 3

[55] P. Milgram and F. Kishino. A taxonomy of mixed reality visual displays. *IEICE TRANSACTIONS on Information and Systems*, 77(12):1321–1329, 1994. 2, 4

[56] P. Milgram, H. Takemura, A. Utsumi, and F. Kishino. Augmented reality: A class of displays on the reality-virtuality continuum. In *Telemanipulator and telepresence technologies*, vol. 2351, pp. 282–292. International Society for Optics and Photonics, 1995. 2, 4

[57] S. Minatani, I. Kitahara, Y. Kameda, and Y. Ohta. Face-to-face tabletop remote collaboration in mixed reality. In *2007 6th IEEE and ACM International Symposium on Mixed and Augmented Reality*, pp. 43–46. IEEE, 2007. 14

[58] D. Moher, A. Liberati, J. Tetzlaff, D. G. Altman, et al. Preferred reporting items for systematic reviews and meta-analyses: the prisma statement. *Int J Surg*, 8(5):336–341, 2010. 2, 3, 12

[59] P. Mohr, S. Mori, T. Langlotz, B. H. Thomas, D. Schmalstieg, and D. Kalkofen. Mixed reality light fields for interactive remote assistance. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, pp. 1–12, 2020. 2

[60] P. Mohr, S. Mori, T. Langlotz, B. H. Thomas, D. Schmalstieg, and D. Kalkofen. Mixed reality light fields for interactive remote assistance. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, pp. 1–12, 2020. 6, 14

[61] J. Müller, R. Rädle, and H. Reiterer. Remote collaboration with mixed reality displays: How shared virtual landmarks facilitate spatial referencing. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*, pp. 6481–6486, 2017. 6, 14

[62] B. Nuernberger, K.-C. Lien, L. Grinta, C. Sweeney, M. Turk, and T. Höllerer. Multi-view gesture annotations in image-based 3d reconstructed scenes. In *Proceedings of the 22Nd ACM conference on virtual reality software and technology*, pp. 129–138, 2016. 14

[63] O. Oda, C. Elvezio, M. Sukan, S. Feiner, and B. Tversky. Virtual replicas for remote assistance in virtual and augmented reality. In *Proceedings of the 28th Annual ACM Symposium on User Interface Software & Technology*, pp. 405–415, 2015. 6, 14

[64] H. Phan, J. Rivas, and T. Song. Collaboration: a literature review research report. 2011. 1

[65] C. Pidel and P. Ackermann. Collaboration in virtual and augmented reality: a systematic overview. In *International Conference on Augmented Reality, Virtual Reality and Computer Graphics*, pp. 141–156. Springer, 2020. 2

[66] T. Piumsomboon, G. A. Lee, J. D. Hart, B. Ens, R. W. Lindeman, B. H. Thomas, and M. Billinghurst. Mini-me: An adaptive avatar for mixed reality remote collaboration. In *Proceedings of the 2018 CHI conference on human factors in computing systems*, pp. 1–13, 2018. 5

[67] T. Piumsomboon, G. A. Lee, J. D. Hart, B. Ens, R. W. Lindeman, B. H. Thomas, and M. Billinghurst. Mini-me: An adaptive avatar for mixed reality remote collaboration. In *Proceedings of the 2018 CHI conference on human factors in computing systems*, pp. 1–13, 2018. 6, 14

[68] T. Piumsomboon, G. A. Lee, A. Irlitti, B. Ens, B. H. Thomas, and M. Billinghurst. On the shoulder of the giant: A multi-scale mixed reality collaboration with 360 video sharing and tangible interaction. In *Proceedings of the 2019 CHI conference on human factors in computing systems*, pp. 1–17, 2019. 14

[69] T. Rhee, S. Thompson, D. Medeiros, R. Dos Anjos, and A. Chalmers. Augmented virtual teleportation for high-fidelity telecollaboration. *IEEE transactions on visualization and computer graphics*, 26(5):1923–1933, 2020. 14

[70] A. Santos-Torres, T. Zarraonandia, P. Díaz, and I. Aedo. Comparing visual representations of collaborative map interfaces for immersive virtual environments. *IEEE Access*, 2022. 14

[71] M. Sereno, X. Wang, L. Besançon, M. J. Mcguffin, and T. Isenberg. Collaborative work in augmented reality: A survey. *IEEE Transactions on Visualization and Computer Graphics*, 2020. 2

[72] R. Skarbez, M. Smith, and M. C. Whitton. Revisiting milgram and kishino's reality-virtuality continuum. *Frontiers in Virtual Reality*, 2:647997, 2021. 2

[73] M. Sousa, D. Mendes, R. K. d. Anjos, D. S. Lopes, and J. Jorge. Negative space: Workspace awareness in 3d face-to-face remote collaboration. In *The 17th International Conference on Virtual-Reality Continuum and its Applications in Industry*, pp. 1–2, 2019. 5

[74] M. Speicher, J. Cao, A. Yu, H. Zhang, and M. Nebeling. 360anywhere: Mobile ad-hoc collaboration in any environment using 360 video and augmented reality. *Proceedings of the ACM on Human-Computer Interaction*, 2(EICS):1–20, 2018. 5, 6, 14

[75] M. Speicher, B. D. Hall, and M. Nebeling. What is mixed reality? In *Proceedings of the 2019 CHI conference on human factors in computing systems*, pp. 1–15, 2019. 2

[76] M. Sra, A. Mottelson, and P. Maes. Your place and mine: Designing a shared vr experience for remotely located users. In *Proceedings of the 2018 Designing Interactive Systems Conference*, pp. 85–97, 2018. 5, 6, 8, 14

[77] H. Sun, Z. Zhang, Y. Liu, and H. B. Duh. Optobridge: assisting skill acquisition in the remote experimental collaboration. In *Proceedings of the 28th Australian Conference on Computer-Human Interaction*, pp. 195–199, 2016. 6, 14

[78] L. Sun, H. A. Osman, and J. Lang. An augmented reality online assistance platform for repair tasks. *ACM Transactions on Multimedia Computing, Communications, and Applications (TOMM)*, 17(2):1–23, 2021. 6, 14

[79] M. Tait and M. Billinghurst. The effect of view independence in a collaborative ar system. *Computer Supported Cooperative Work (CSCW)*, 24(6):563–589, 2015. 14

[80] F. Tecchia, L. Alem, and W. Huang. 3d helping hands: a gesture based mr system for remote collaboration. In *Proceedings of the 11th ACM SIGGRAPH international conference on virtual-reality continuum and its applications in industry*, pp. 323–328, 2012. 14

[81] T. Teo, A. F. Hayati, G. A. Lee, M. Billinghurst, and M. Adcock. A technique for mixed reality remote collaboration using 360 panoramas in 3d reconstructed scenes. In *25th ACM Symposium on Virtual Reality Software and Technology*, pp. 1–11, 2019. 5, 14

[82] T. Teo, L. Lawrence, G. A. Lee, M. Billinghurst, and M. Adcock. Mixed reality remote collaboration combining 360 video and 3d reconstruction. In *Proceedings of the 2019 CHI conference on human factors in computing systems*, pp. 1–14, 2019. 14

[83] T. Teo, G. A. Lee, M. Billinghurst, and M. Adcock. Hand gestures and visual annotation in live 360 panorama-based mixed reality remote collaboration. In *Proceedings of the 30th Australian Conference on Computer-Human Interaction*, pp. 406–410, 2018. 14

[84] S. Thanyadit, P. Punpongsanon, and T.-C. Pong. Observar: Visualization system for observing virtual reality users using augmented reality. In *2019 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*, pp. 258–268. IEEE, 2019. 6, 7, 14

[85] B. Thoravi Kumaravel, F. Anderson, G. Fitzmaurice, B. Hartmann, and T. Grossman. Loki: Facilitating remote instruction of physical tasks using bi-directional mixed-reality telepresence. In *Proceedings of the 32nd Annual ACM Symposium on User Interface Software and Technology*, pp. 161–174, 2019. 5, 6, 8, 14

[86] R. van der Kleij, J. Maarten Schraagen, P. Werkhoven, and C. K. De Dreu. How conversations change over time in face-to-face and video-mediated communication. *Small Group Research*, 40(4):355–381, 2009. 2

[87] N. Vaughan, B. Gabrys, and V. N. Dubey. An overview of self-adaptive technologies within virtual reality training. *Computer Science Review*, 22:65–87, 2016. 2

[88] P. Wang, X. Bai, M. Billinghurst, S. Zhang, W. He, D. Han, Y. Wang, H. Min, W. Lan, and S. Han. Using a head pointer or eye gaze: The effect of gaze on spatial ar remote collaboration for physical tasks. *Interacting with Computers*, 32(2):153–169, 2020. 6, 14

[89] P. Wang, X. Bai, M. Billinghurst, S. Zhang, X. Zhang, S. Wang, W. He, Y. Yan, and H. Ji. Ar/mr remote collaboration on physical tasks: a review. *Robotics and Computer-Integrated Manufacturing*, 72:102071, 2021. 2

[90] P. Wang, X. Bai, M. Billinghurst, S. Zhang, X. Zhang, S. Wang, W. He, Y. Yan, and H. Ji. Ar/mr remote collaboration on physical tasks: A review. *Robotics and Computer-Integrated Manufacturing*, 72:102071, 2021. 4

[91] S. Whittaker. Things to talk about when talking about things. *Human–Computer Interaction*, 18(1-2):149–170, 2003. 5, 8

[92] R. Wolff, D. J. Roberts, A. Steed, and O. Otto. A review of telecollaboration technologies with respect to closely coupled collaboration. *International Journal of Computer Applications in Technology*, 29(1):11–26, 2007. 2

[93] S. Yamada and N. P. Chandrasiri. Evaluation of hand gesture annotation in remote collaboration using augmented reality. In *2018 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*, pp. 727–728. IEEE, 2018. 14

[94] S. Yamamoto, H. Tamaki, Y. Okajima, K. Okada, and Y. Bannai. Symmetric model of remote collaborative mr using tangible replicas. In *2008 IEEE Virtual Reality Conference*, pp. 71–74, 2008. doi: 10.1109/VR.2008.4480753 6, 14

[95] B. Yoon, H.-i. Kim, S. Y. Oh, and W. Woo. Evaluating remote virtual hands models on social presence in hand-based 3d remote collaboration. In *2020 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*, pp. 520–532. IEEE, 2020. 14

[96] J. Young, T. Langlotz, M. Cook, S. Mills, and H. Regenbrecht. Immersive telepresence and remote collaboration using mobile and wearable devices. *IEEE transactions on visualization and computer graphics*, 25(5):1908–1918, 2019. 14

[97] K. Yu, G. Gorbachev, U. Eck, F. Pankratz, N. Navab, and D. Roth. Avatars for teleconsultation: effects of avatar embodiment techniques on user perception in 3d asymmetric telepresence. *IEEE Transactions on Visualization and Computer Graphics*, 27(11):4129–4139, 2021. 14

[98] F. Zhou, H. B.-L. Duh, and M. Billinghurst. Trends in augmented reality tracking, interaction and display: A review of ten years of ismar. In *2008 7th IEEE/ACM International Symposium on Mixed and Augmented Reality*, pp. 193–202. IEEE, 2008. 3

A APPENDIX

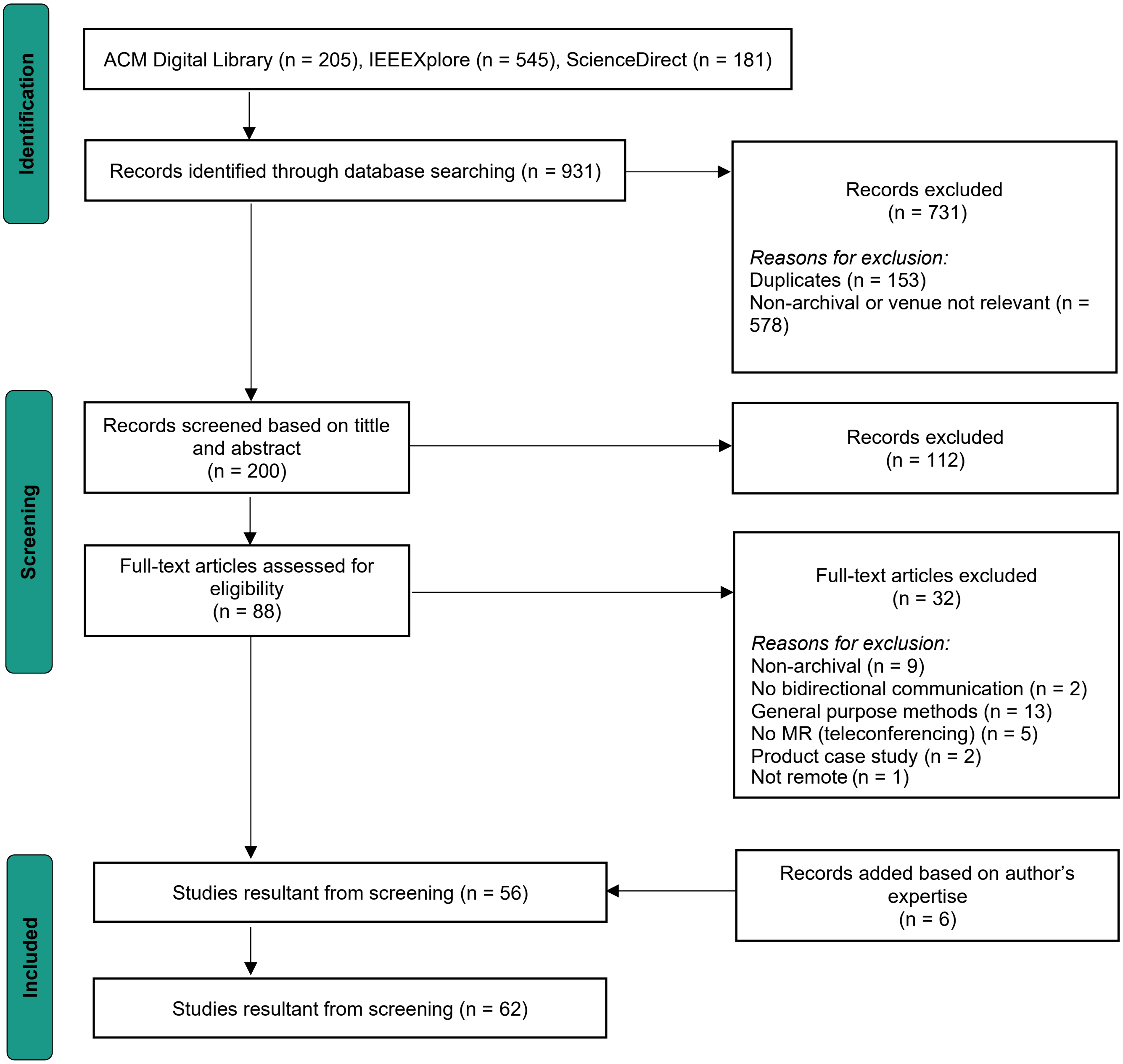

Identification
Screening
Included
ACM Digital Library (n = 205), IEEEXplore (n = 545), ScienceDirect (n = 181)
Records identified through database searching (n = 931)
Records excluded
(n = 731)
Reasons for exclusion:
Duplicates (n = 153)
Non-archival or venue not relevant (n = 578)
Records screened based on tittle and abstract
(n = 200)
Records excluded
(n = 112)
Full-text articles assessed for eligibility
(n = 88)
Full-text articles excluded
(n = 32)
Reasons for exclusion:
Non-archival (n = 9)
No bidirectional communication (n = 2)
General purpose methods (n = 13)
No MR (teleconferencing) (n = 5)
Product case study (n = 2)
Not remote (n = 1)
Studies resultant from screening (n = 56)
Records added based on author's expertise
(n = 6)
Studies resultant from screening (n = 62)

Figure A.1: PRISMA flow diagram for study selection process based on Moher et al. [58]

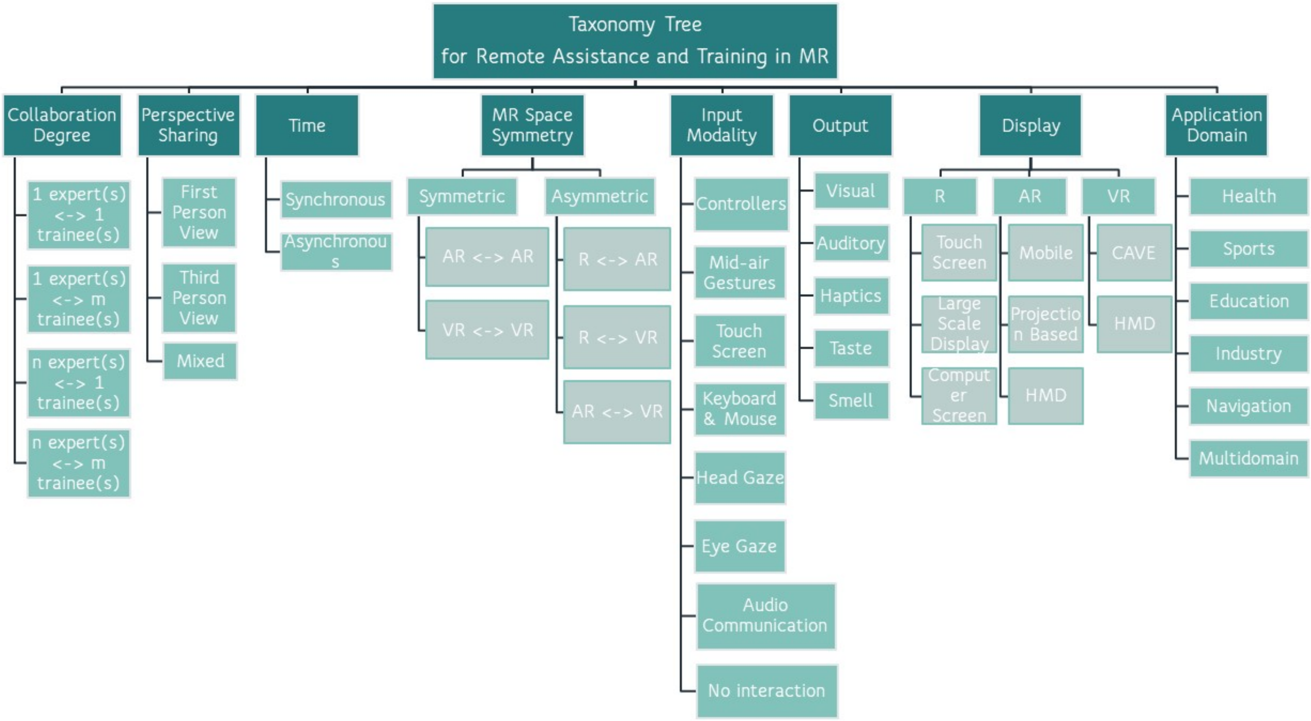


Figure A.2: Final dimensions of our Taxonomy for remote assistance and training in MR environments.

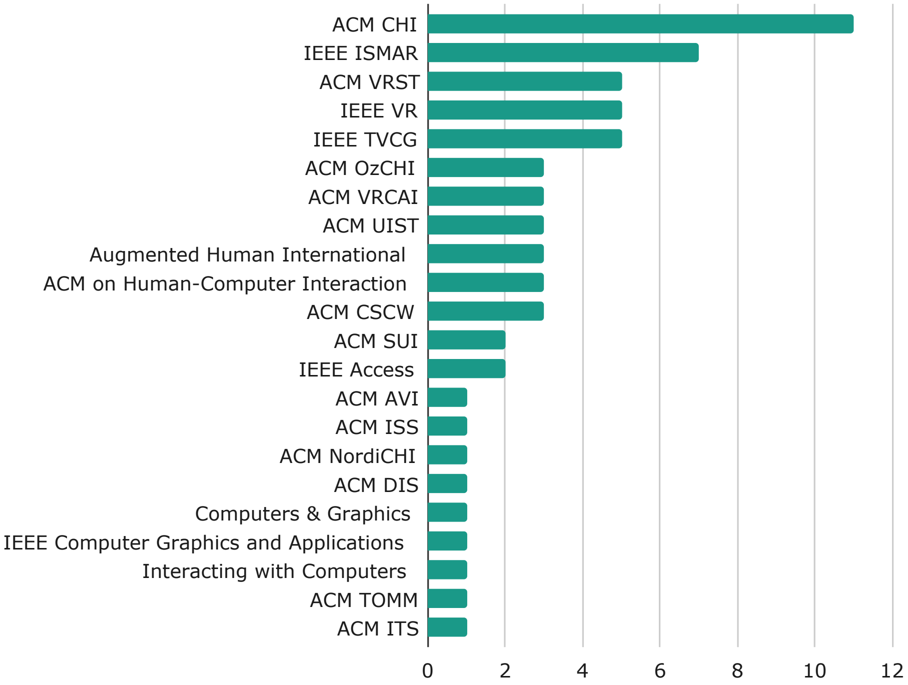


Figure A.3: Final publication distribution by venue.