Collaborative Work in Augmented Reality: A Survey

Mickael Sereno, Xiyao Wang, Lonni Besançon, Michael J. McGuffin, and Tobias Isenberg

Abstract—In Augmented Reality (AR), users perceive virtual content anchored in the real world. It is used in medicine, education, games, navigation, maintenance, product design, and visualization, in both single-user and multi-user scenarios. Multi-user AR has received limited attention from researchers, even though AR has been in development for more than two decades. We present the state of existing work at the intersection of AR and Computer-Supported Collaborative Work (AR-CSCW), by combining a systematic survey approach with an exploratory, opportunistic literature search. We categorize 65 papers along the dimensions of space, time, role symmetry (whether the roles of users are symmetric), technology symmetry (whether the hardware platforms of users are symmetric), and output and input modalities. We derive design considerations for collaborative AR environments, and identify under-explored research topics. These include the use of heterogeneous hardware considerations and 3D data exploration research areas. This survey is useful for newcomers to the field, readers interested in an overview of CSCW in AR applications, and domain experts seeking up-to-date information.

Index Terms—Introductory and Survey, Computer-Supported Cooperative Work, Virtual and Augmented Reality, Immersive Analytics

◆

1 INTRODUCTION

AUGMENTED Reality (AR) as a field can be traced back to the late 1960s (e. g., [160]) and the term was first used in the early 1990s by Caudell and Mizell [37]—former Boeing engineers seeking to display simple information (e. g., text) in 3D space to enhance manufacturing processes.

Since then, researchers have continued to improve AR hardware, algorithms, and user interfaces, with several communities showing interest in augmenting physical reality with virtual content. AR has been applied to medicine [14], [145], education [4], [130], archaeology [13], [17], games [131], remote expert guidance [54], [95], [121], industry [37], [41], [72], [125], crisis response [118], [139], and information visualization [15]. Since 2016, AR Head-Mounted Displays (AR-HMDs) have become much more affordable and closer to consumer markets with the release of Microsoft's HoloLens and HoloLens 2, the Meta 2, and Magic Leap One, all priced at US$ 3500 or less, and supporting Unity as an accessible SDK. Over the same period, smartphone applications (e. g., the Pokémon Go game) have increased the public's knowledge of, and interest in, augmented reality.

Meanwhile, the field of Computer-Supported Collaborative Work (CSCW)—whose main conference, ACM CSCW, started in 1986—is relatively mature and supported by a large body of literature. Researchers understand the psychological aspects and difficulties to consider when designing collaborative systems. CSCW is also important in industry, where multiple collaborators, remote or co-located, need to interact with each other both synchronously and asynchronously.

However, the main academic conference in AR, IEEE International Symposium on Mixed and Augmented Reality (ISMAR), has paid little attention to CSCW. Indeed, only ≈ 1.7% of papers published between 2008 and 2017 at ISMAR discussed collaborative work in Mixed Reality (MR) or AR [87]. The intersection of AR and CSCW, which we call AR-CSCW, possesses its own challenges to overcome, but has not been studied much by researchers. With the present survey, we aim to provide both software designers and researchers with an overview and a research agenda for the use of AR in CSCW contexts and a new taxonomy of AR-CSCW systems.

Our work extends previous surveys on subjects related to AR-CSCW. Some of these past surveys focused on more limited domains, such as education [4], [130], [175] or medical training [14]. Other surveys have discussed AR and CSCW in a more general way. In chronological order, these are Billinghurst and Kato [26], now 20 years old; Alem and Huang [6] and Billinghurst and Thomas [28], who focused on mobile collaborative work; Lukosch et al. [105], whose work still predates recent HMDs such as the HoloLens, and unlike the present survey, did not discuss psychological aspects of AR nor interaction techniques nor provide design guidelines; Irlitti et al. [78], who focused on asynchronous AR collaboration; Billinghurst et al. [25], who discussed the emerging field of Collaborative Immersive Analytics, but without a specific focus on AR; and Ens et al. [56] whose taxonomy focused more on the purpose of the systems while we focus more on their output and input interfaces (see Section 3). In contrast to their taxonomy, we analyze different dimensions, examine different application areas— we discuss general CSCW topics and, in particular, the field of visualization—, and propose a summary of design considerations that we extracted from the surveyed papers. Looking at research in AR in general and without a focus on CSCW, Zhou et al. [179] and later Kim et al. [87] surveyed works presented at the ISMAR conference, categorizing and analyzing papers by research area (e. g., tracking, displays).

- *Mickael Sereno, Xiyao Wang, and Tobias Isenberg are with Université Paris-Saclay, CNRS, Inria, LRI, France.*
  Email: {mickael.sereno,xiyao.wang,tobias.isenberg}@inria.fr
- *Lonni Besançon is with Linköping University, Sweden.*
  Email: lonni.besancon@gmail.com
- *Michael J. McGuffin is with École de Technologie Supérieure, Canada.*
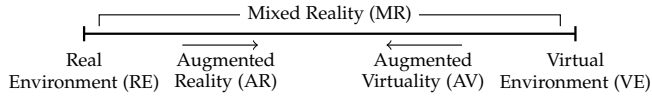 *Email: michael.mcguffin@etsmtl.ca*

Fig. 1: Milgram and Kishino's Reality-Virtuality Continuum [114].

The present survey is thus unique in surveying AR and CSCW, focusing in particular on user interfaces. We have systematically examined contributions from the ISMAR and CSCW conferences between 2008 and 2019, in addition to including work from other venues that we found to be relevant for a total of 65 papers. We organize these papers in a taxonomy based on the following dimensions: space, time, role symmetry (whether the roles of users are symmetric), technology symmetry (whether the hardware platforms of users are symmetric), and output and input devices.

Our survey has two major parts. Section 2 reviews fundamentals of both AR and CSCW and can be skipped by practitioners and experts. The subsequent sections discuss the papers we selected for review, present design considerations, identify research areas to further investigate, and propose a discussion on visual analytics which we added to increase the relevance of the survey to the visualization community.

2 AUGMENTED REALITY AND CSCW

Here we review relevant definitions as well as psychological and technological aspects related to AR.

2.1 Augmented Reality

AR and Mixed Reality (MR) can be explained in terms of a *Reality-Virtuality Continuum* (RVC, Figure 1) defined by Milgram and Kishino [114] which covers different ways of mixing computed and physical objects. At one extreme is the real (physical) environment, and at the other is a purely virtual environment (VE). MR applications are located between these two extremes. Users may receive feedback about objects through any sensory channel, most commonly visual, audio, and haptic. Roo and Hachet's [146] system allowed users to transition across all levels of a kind of RVC, supporting six levels: the purely physical world, augmented surfaces (i. e., digital information placed on physical objects), mid-air digital content, object decoupling (i. e., manipulating virtual objects representing real ones), body decoupling (i. e., virtually navigate in the real world), and a fully virtual world. Finally, systems that *remove* physical objects instead of *adding* virtual contents relate to the *Diminished Reality* [116] concept.

AR is a special case of MR and is closely connected to the real world [9]. Azuma [9] and Azuma et al. [8] suggested the following requirements for AR systems: (1) **the merging and alignment** of real and virtual information, (2) **real-time rendering** through all the sensory channels, and (3) a **real-time interactive** environment. When coupled with CSCW features, we speak of a *Mixed-Reality Collaborative Environment*, with *Studierstube* [149], [150], [162] being among the first systems that were created.

2.2 Psychological Aspects

In AR systems, users may feel that virtual objects are transported to them or that they are transported to a remote place. With users experiencing a 3D augmented world mimicking their real environments, designers can improve the users' sense of presence, and their performance [174], by considering how users behave in real life for a particular task. Still, past research [118] has shown that the choice of devices and interfaces influences the user's level of engagement, immersion, and presence, which in turn influence their performance. We therefore discuss the concepts of *presence and immersion* as well as *engagement*.

2.2.1 Presence and Immersion

Presence is a matter of focus that users experience when they feel involved and immersed in their tasks [61], [174]. Witmer and Singer [174] found a weak but consistent positive correlation between presence and performance. They theorized several factors contributing to the sense of presence:

Control Factors refer to ways in which users can control the system and environment. Interactivity should happen in real-time and users should be able to foresee system feedback. Yuill and Roger [178] similarly identified controllability as important for the success of multi-user systems.

Sensory Factors refer to the richness of feedback to the user's senses. Inconsistent stimuli engender illness.

Distraction Factors are determined by the user's isolation. Presence increases with greater isolation.

Realism Factors are determined by the realism and the consistency of the environment.

Witmer and Singer also defined immersion as a psychological state users experience in an environment that continuously streams stimuli. For MR environments, their concept of immersion relates to the technological features (e. g., field-of-view, FoV) [33] which positively influence (medium-size effect) the users' sense of presence [43]. Slater [152] defined *immersion* and *place illusion* as a set of *sensorimotor contingencies* that approximate those of physical reality, and the *valid actions* users can perform in their VE. He also introduced the concept of *plausibility illusion* as "the illusion that the scenario being depicted is actually occurring," which relates to the realism factor. Immersion is usually greater in MR environments than in regular workstations, as long as users do not experience *cybersickness* [94].

Finally, presence can be categorized into three categories:

Physical presence A cognitive state in which users experience the virtual stimuli as real [98]. Strong physical presence means the MR environment is expected to exhibit plausible physics, with the user's imagination often interpolating incomplete stimuli. In AR environments, physical presence is related to the *object presence* which refers to how realistically a virtual object is anchored in the real environment [84], [158].

Social presence *"The sense of being with another"* [30] is one of the main concepts in remote AR collaboration. According to Biocca et al. [30], it mainly influences the productivity of collaborative environments.

Self presence A cognitive state where the virtual self is experienced as the real self [98]. This is more related to virtual environments where virtual avatars represent users.

As it is for traditional relaxed What-You-See-Is-What-I-See CSCW systems, AR and VR CSCW applications can display custom content per user (albeit of their viewpoint) because users wear their own rendering system. These

TABLE 1: The presentation/placement awareness display techniques matrix [69]. *Literal* information is displayed as it is captured, whereas *Symbolic* information has been codified. *Situated* information is displayed in the workspace, whereas *Separate* information is displayed outside.

		Information Placement	
		Situated	Separate
Information Presentation	Literal		
	Symbolic		

subjective views [3] can then alter the users' social presence if the workspace loses its spatial coherency [128].

Finally, the use of AR, as compared to traditional tools, can improve user performance and learnability [130].

2.2.2 Engagement

Attfield et al. [7] defined engagement as *"the emotional, cognitive and behavioral connection that exists, at any point in time and possibly over time, between a user and a resource."* Potentially, this produces a state of *flow* [42] which is always enjoyable.

O'Brien and Toms [119] characterized the level of engagement of an application by its *aesthetic appeal, attention, challenge, endurability, feedback, interactivity, perceived user control, pleasure, sensory appeal*, and *novelty*. AR tends to engage users more than traditional tools and sometimes to pure VEs. The control and richness of sensory feedback (e. g., stereoscopic vision, spatial audio), and the ubiquitous interfaces increasing both perceived user control and novelty, tend to be higher in MR systems than in regular workstations. Moreover, the virtual and real-world objects/tools are blended together, making focus transitions seamless [26]. Finally, social presence is higher in AR systems by being close to real-world rules and by preserving social protocols as compared to regular workstations and VEs [31], [70].

Scenarios with users not sharing the same role exist. Some displays (e. g., AR-HMD, traditional 2D screens) may then be more suitable for some users than others, which relates to Isenberg et al.'s [80] CSCW visual applications categorization based on users' engagement:

Viewing occurs when a majority is spectator of a minority's actions. Traditional 2D screens might be considered for the spectators while the actors use AR devices.

Interacting/Exploring occurs when the group is exploring, discussing, and understanding the same data.

Sharing/Creating occurs when participants of the group can create, upload, and share new content.

2.2.3 Awareness, Embodiment

According to Gutwin and Greenberg [69], "previous researchers have defined awareness as knowledge created through interaction between an agent and its environment." They also define Workspace Awareness (WA) as *"the up-to-the-moment understanding of another person's interaction with the shared workspace,"* which they derived from Adams et al.'s [1] *"up-to-the minute cognizance required to operate or maintain a system."* Gutwin and Greenberg [69] also define three sources of WA: bodies and consequential communication, workspace artifacts, and conversational elements (e. g., speech, gestures).

They defined from these sources the *Presentation and Placement awareness display techniques* matrix (Table 1). The situated-literal approach is perhaps the most natural one and provides effective awareness, with embodiment and expressive artifacts as two critical elements. They defined

embodiment as *"the visual representation of a person's body in the workspace,"* such as a real body, video records, and 3D avatar model with more or less cues (e. g., gaze cues, limb cues). Embodiment is a strong way to provide both consequential communication and conversational resources and is important in mixed-space collaborative environments.

2.3 Mixed-Space Collaborative Work

Mixed-space collaborative setups include any collaborations where users are in multiple shared-spaces, defined as a space placed at one point on the Reality-Virtuality Continuum.

After defining shared-spaces, Benford et al. [16] introduced Mixed-Space Boundaries, i. e., how one can see, view, or interact from one shared-space to another *adjacent* one, which has common properties/purposes. Misunderstandings of two separate spaces (e. g., two remote users, each having their own AR space) can hinder performance. For example, in distributed collaborations, each user should be aware of the others. A strong embodiment can support this goal, but also more cues, e. g., augmented pointing actions or displaying a user's FoV to others. One common mixed-space boundary is between one *exocentric* space (i. e., view of the data from outside) and one *egocentric* space (i. e., view of the data from inside). Following Kiyokawa et al. [90], Shared Augmented Environments, which appear in an exocentric way, lead to richer awareness than Shared Virtual Environments (SVE) and are more suited for co-located work. In contrast, SVEs better support parallel activities, e. g., visualizing from different perspective (e. g., scales, positions) and can be used in an egocentric *private view*. However, if other participants can see you and do not understand what you are doing (e. g., manipulating your private view), they can be confused: how can one understand whether the other is manipulating the private or the public view? Benford et al. [16] characterized Shared-Space Technologies with

Transportation How much is the user transported to another place or vice-versa? While not feeling transported in co-located systems is usually expected, it may be a requirement for remote ones.

Artificiality What is the degree of artificiality of the shared-space? This can be related to the RVC [114].

Spatiality How are the user's shared-space technology spatially defined? Is its coordinate system shared? Can others determine this user's position and orientation relative to themselves?

The transportation and artificiality axes rely mostly on the system's output and input modalities, which we review in the next sections and include in our taxonomy.

2.4 Displays

Workspace Awareness (WA), an important research topic in AR-CSCW, can be impacted by the type (and positioning) of display. We consider Head-Mounted Displays (HMDs), Hand-Held Displays (HHDs), and spatial displays anchored in the physical environment. We also consider optical see-through (OST) displays, video see-through (VST) displays, and projectors. Finally, we consider stereoscopic 3D displays (S3Ds). McIntire et al. [113] summarized, across 184 experiments, where and when S3Ds outperform traditional

ones. They asserted that S3Ds improve performance when monocular depth cues are weak, the objects of interest are *close* for a significant binocular disparity, and when the tasks are *difficult* (see also Bowman and McMahan [33]). S3Ds better support novices than expert users.

S3Ds rely on depth cues which Cutting and Vishton [44] categorized as monocular or binocular. Monocular depth cues include *occlusion*, *aerial perspective*, the relative *size*, *density*, and *height* of objects in the user's retinal image, and *motion perspective*. Binocular depth cues include *binocular disparity*, *convergence* and *accommodation*. Depth cue conflicts hinder performance and comfort [103].

HMDs support all the technologies cited above and are mostly stereoscopy-enabled. VST displays usually have higher FoV than OST displays due to physical constraints but have lower refresh rate due to camera capture rate and latency, strongly influencing cyber-sickness. HHDs are generally based on a VST display as part of a smartphone or a multi-touch tablet. Raskar et al. [143] introduced Spatial Augmented Reality (SAR) as virtual content displayed within or on the user's physical space, which can contain complex colors and geometries [29]. They are based on VST stationary screens and/or room-space projector devices. In SAR settings, users do not wear any display devices, except for shutter glasses when stereoscopy is required. Without glasses, users can still have a sense of perspective [18] which can be coupled with stereoscopy with worn glasses [19]. Multiple users can be in the AR space using time-multiplexed shuttered glasses to display contents individually.

Each display device has its own benefits and limitations. A SAR is non-intrusive but lacks mobility. HHDs are commonly found in the consumer market, with smartphones and tablets being ubiquitous, but are not stereoscopy-enabled. Regarding depth cues, designers should focus first on monocular ones and should avoid depth cue conflicts.

We next consider input modalities.

2.5 Interaction Techniques

Data exploration often relies on interaction. Input modalities include touch, tangible, and mid-air gesture input. When collaboration is an important aspect, not every modality might be useful. For example, it may be hard to understand what one is doing on a private multi-touch tablet that is used as the input device.

2.5.1 Touch Devices

Touch devices like smartphones or tablets can be used in multiple ways. First, they provide a good way to interact with Graphical User Interface (GUI) components (e. g., buttons, menus) [99], [148]. They can be used as VST AR devices which provide seamless interaction in both the GUI and the AR views. However, the view is limited by finger occlusion during interaction and low FoV, especially for small devices like smartphones. Finally, they can be used as a complement to another AR display, such as an AR-HMD [159].

The *Studierstube* system [149], [150], [162] used a lightweight Personal Interaction Panel (PIP) to interact—via a WIMP (Window, Icon, Menu, Pointer) interface—with data visualized in AR-HMDs. The touch device can be used to draw graphical (e. g., [155]) or textual annotations, a key feature of their collaborative system. The 3D pen allows direct

manipulation in the 3D space. López et al. [104] showed how a multi-touch tablet facilitates 3D data transformations visualized in a S3D. Büschel et al. [35] explored and studied multiple aspects of a tracked multi-touch device as a way to pan and zoom 3D datasets visualized in an AR-HMD.

The benefits of touch over other forms of interaction have been deeply studied for a variety of settings. The literature explains [34] why touch can be considered to be a direct form of interaction, usually improving speed at the cost of accuracy [5], [62]. Its speed advantage in 2D interfaces, however, does not seem to translate to 3D scenarios [21].

2.5.2 Tangible Interface

Tangible User Interfaces (TUIs) [82] augment the physical world by linking digital information with real objects. TUIs thus extend the graspable interface [59] concept, which uses physical objects to directly control digital objects. Most VR games use two hand-held controllers, which can serve as both classical controllers and tangible interfaces that represent, e. g., light sabers.[1] The buttons can be used as *triggers* of the virtual object to which the controller is attached. Tangible interaction facilitates fast and precise interaction [21], [40], [74] that mimic the real world [58], [82], foster collaboration [111], [124], and can provide entertainment [21], [177].

SARs allow users to directly manipulate physical objects. Piper et al.'s [133] *Illuminating Clay* allows users to interact with landscape data using AR and depth cameras on 3D landscape models. Wilson and Benko's [173] *LightSpace* is a luminous room that uses several depth cameras and projectors to track multiple users and let them tangibly interact with immobile rectangular augmented surfaces. Roo et al.'s [145] *Inner Garden* allows users to meditate by building a mini-world using a SAR tangible sandbox and exploring it with a VR-HMD. The SAR content changes based on the sand shape (e. g., display snow), with some effects (e. g., cloud shadows) derived from the user's breathing and heart-rate.

In AR, users may need their hands free to, for instance, repair an aircraft engine. Henderson and Feiner [71] imagined a system where the AR device selects special physical elements to anchor Augmented Widgets (e. g., buttons). This design leaves the physical environment unmodified, while providing the user with passive haptic feedback and tangible interfaces. They called these interfaces *Opportunistic Control*.

2.5.3 Gestural Interfaces

Many forms of gestures exist; e. g., touch/pen gestures (i. e., drawing), mid-air gestures, eye gestures, and gestures using props (e. g., conducting). Gestures in general may not be correctly understood by the system, while mid-air gestures can also fatigue users (e. g., *Gorilla-arm* syndrome [73], [83]). In co-located collaborative setups where users use body language, mid-air gestures can be confusing: it is unclear if gestural actions are meant as system input or deictic communication with a collaborator [17]. In addition, techniques relying on visual arm extensions for remote object selection (e. g., the *Go-Go* technique [141]) may not be a good fit for AR because users can see both their real and their virtual hands, introducing a perceptual conflict. Mid-air gestures, however, may be a requirement to provide users with a way to keep

---

1. See the Beat Saber game, https://beatsaber.com/ .

their hands free as it could be crucial in, e. g., sterile medical environments [51], [123]. They can be used with HMDs without additional input devices and may improve visual data exploration. Keefe and Isenberg identified challenges faced by using gestures in 3D visualization [86].

2.5.4 Hybrid Interfaces

Hybrid Interfaces use multiple kind of inputs. Bolt's [32] system, e. g., allows users to draw shapes at specific locations using both voice and gesture commands. Speech input can also be used to set the system mode (e. g., scaling, translating, rotating), while a gesture performs the action [107]. Touch devices can be used as tangible objects (e. g., 2D cutting planes [20], [155]). A tablet can be used to seamlessly grasp and release virtual objects. Surale et al. [159], e. g., proposed to use a tablet in VR space both as a touch device and as a tangible one. Their design space also includes mid-air gestures to interact with 3D content in a coarser way. The PIP discussed in Section 2.5.1 supports 3D interactions with AR data using a 6 DoF pen. Dedual et al.'s [45] hybrid interface associated a 2D multi-touch tabletop with an AR-HMD. The tabletop displayed footprints (2D projections) of 3D augmented objects that are visible through the HMD. Other collaborators can see the footprints and interact with them in the same coordinate system.

2.6 Multiple Rendering Windows

2D and/or 3D content can be grouped as contexts: e. g., private contexts only visible to a single user and public ones visible to all. Effective context representations may increase WA about what is publicly shared and what is not.

2.6.1 Multiple 2D Windows

Designers can add 2D content information in AR such as charts [89], web pages [26], 2D windows [57], and virtual cameras. The latter render real-world content from a different point-of-view to show occluded or remote information.

Ens et al.'s [55] *Ethereal Planes* framework categorizes 2D (extendable to 3D) elements along seven dimensions:

Perspective Is the information *exocentric* (relative to the physical world) or *egocentric* (relative to the user)? The authors suggest that users prefer exocentric views for public content and egocentric views for private data.

Movability Is the content movable in its reference frame?

Proximity can be *on-body*, *near* (within arm's reach), and *far* (outside of arm's reach).

Input mode Is the input mode *direct* or *indirect*? Indirect mode is more precise and less tiring, while direct mode is more user-friendly.

Tangible Is the information bound to a tangible object?

Visibility Is the information fully visible, partially visible, or does it rely little or not at all on the visual channel?

Discretization Is the information continuous or discrete?

These dimensions are not orthogonal (i. e., independent), e. g., a distant object is difficult to directly manipulate. Some CSCW applications permit participants to have a private view where they manipulate private data, a public view visible by all, and a customized view merging all content. Using egocentric views for private elements allows users to not disturb the workflow of other collaborators, by not

TABLE 2: Paper selected for review. References in parentheses refer to discarded papers.

		Total
IEEE ISMAR 2008–2019	[68], [88], [97], [100], [102], [107], [108], [118], [120], [122], [142], [165] ([101])	12 + 1
ACM CSCW 2008–2019	[12], [47], [64], [129], [139], [140]	6
Additional 2008–2019	[2], [18], [23], [41], [48], [49], [65], [66], [67], [75], [77], [79], [85], [91], [95], [96], [115], [117], [121], [126], [134], [136], [137], [138], [147], [154], [163], [164], [167], [168], [169], [173], ([78], [135])	32 + 2
Additional before 2008	[16], [17], [24], [26], [27], [36], [90], [106], [127], [133], [144], [149], [156], [161], [162]	15
Total		**65 + 3**

manipulating the data that is displayed for all workers. Custom views, however, can be either egocentric or exocentric, depending on their nature, e. g., some pathlines for 3D flow dataset can be visible to all or only to some users. Other private layouts may also be considered following the nature of the collaborative environment, e. g., adaptive layouts [93].

2.6.2 Multiple 3D Windows

Schmalstieg et al. [149] defined a 3D window as a context that contains its own scene graph—limited in space and distinguishable from others using decorations. 3D windows can be resized, moved, rotated, minimized, reopened, and have multiple graphical effects. When a user interacts with a 3D window, their system highlights its boundaries, which provides additional awareness cues. In other systems, empty 3D windows could also highlight others' private contexts.

3 METHOD

We now describe how we selected the papers we reviewed and discuss the dimensions of our taxonomy.

3.1 Selection Process

Our survey covers 65 papers related to AR-CSCW that we chose in two phases. The first phase was more systematic, while the second one added additional papers through an exploratory, opportunistic search. Our criteria for selecting papers were that they should focus on collaborative systems with at least a partial AR module, and they should be at least the length of a short paper, excluding extended abstracts.

In the first phase, we surveyed the two main conferences focused on AR and CSCW, respectively: IEEE ISMAR and ACM CSCW. We started with Kim et al.'s [87] survey, which classifies all ISMAR papers from 2008 to 2017, including nine papers related to collaboration. After further investigation, we added one more paper related to collaboration [120]. We then checked the proceedings of IEEE ISMAR 2018 and 2019, and added three additional relevant papers. Next, we used the ACM digital library to find papers within the ACM CSCW conference published between 2008 and 2018 and matching the keywords "Augmented," "Mixed," "AR," or "MR," which yielded 1001 papers. We filtered these papers by looking at titles and keywords, narrowing the selection down to a subset of 62 possibly relevant papers. We further filtered out extended abstracts and demos. We carefully read the remaining papers, and finally decided to add three of them to our survey based on our criteria. We compared our set of papers to those in Ens et al.'s recent survey [56], adding one additional paper from ACM CSCW. In the course of our

TABLE 3: Traditional Time-Space Matrix [11].

Time \ Space	Co-located	Remote
Synchronous	Co-Located synchronous	Remote synchronous
ASynchronous	Co-Located asynchronous	Remote asynchronous
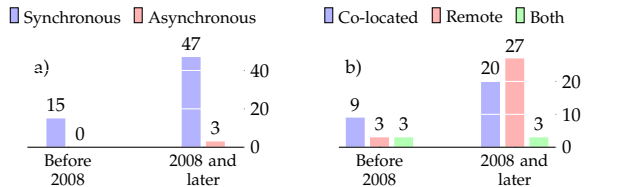
Fig. 2: Paper distribution along a) the time and b) the space dimensions.

Fig. 3: Surveyed papers by year of publication.

research, the CSCW 2019 proceedings were published, from which we added two more papers to our survey. Thus, our first phase resulted in selecting 19 papers on AR-CSCW, as shown in the first two rows of Table 2.

In the second, exploratory phase, we selected papers we already knew to be relevant, and we used other search tools such as Google Scholar, identifying 34 papers from the 2008–2019 period and 15 more published before 2008. We did not place the focus on a particular venue. We used the other surveys we mentioned in Section 1 to find highly relevant papers in the AR-CSCW field. Finally, to be up-to-date with the recent literature, we mostly focused on papers published during the last decade. We mostly used keywords including "collaborative AR/MR," "remote AR/MR," and "co-located AR/MR." In total, we found 68 papers related to AR-CSCW, out of which we discarded 3 papers, because they focused on algorithms or theory, without much detail about a concrete system. We list the results in Table 2.

3.2 Taxonomy

A categorization of collaborative AR techniques cannot rely only on AR or only on CSCW aspects. We thus base our taxonomy on a combination of past classifications from AR, CSCW, and AR-CSCW work. The most commonly used dimensions to classify CSCW systems are time and space (Table 3), which are fundamental for any CSCW system.

Time is split into **synchronous** and **asynchronous** work.

Space is subdivided into collaborators who are **co-located** (working in the same place) or **remote**.

We further assess aspects of symmetry considered by Ens et al. [56], which we split into two dimensions for a more fine-grained analysis. We also decided to exclude their dimensions of scenario and focus because they involve too many possible settings to be useful in our context.

Role Symmetry refers to the role users have: **symmetric** if users are performing the same kind of tasks; **asymmetric** if, for instance, one user is assisting another one.

Technology Symmetry is **symmetric** if users use the same hardware devices; **asymmetric** otherwise.

In these last two dimensions, a value of **both** indicates symmetry between some users and asymmetry between others.

In visualization applications, the Role Symmetry dimension relates to Isenberg et al.'s [80] *level of engagement* dimension: users may not have the same level of engagement
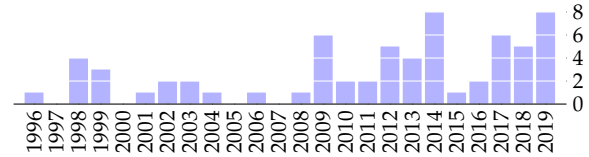
based on their roles. For example, symmetry occurs when all users are exploring their data, while asymmetry occurs when one user drives the analysis and others observe.

We also include two dimensions related to the used hardware devices, which are tightly linked to the technology symmetry dimension. As explained in Section 2.2, these influence the users' *immersion*, *presence*, and their *performance*.

Output devices refer to the type of visual output devices being used. In the surveyed papers, we found that **CAVEs**, **VR-HMDs**, **AR-HMDs**, **(AR) Hand-Held Displays**, **SAR devices**, and **traditional screens** were used.

Input devices refer to the type of input used to interact with the virtual contents. We found in the papers we surveyed the use of **hand tracking**, **tracked controllers**, **hand mid-air gestures** (which involves hand tracking), **touch**, **head gaze/orientation**, **eye gaze**, **tangible**, **non-tracked controller**, **speech**, and regular **keyboard and mouse** input modalities.

The dimension of Output devices also relates to Kiyokawa's [90] artificiality dimension and Gupta et al.'s [68] classification of wearable collaborative systems. Moreover, designers generally know in advance the type of output they want to use due to physical, marketing, or maintenance constraints. On the other hand, the Input devices dimension indicates how designers chose to empower users of AR, whether it is in a CSCW context or not. Note that the input and output devices are not independent dimensions, e.g., a touchscreen coupled with standalone AR-HMD would be an unusual and often technically awkward combination.

We did not consider other sensory output (haptic, audio) because the visual channel is typically the most important one, and because none of the works we surveyed discussed such alternative forms of output.

4 PAPER SURVEY

We begin by reviewing papers that do not implement a complete AR system in Section 4.1, before discussing the majority of papers with complete AR systems with respect to our taxonomy, as classified in Table 4. In our set, most papers were published after 2008 (see Figure 3) due to our semi-systematic selection and, as Ens et al. [56] stated, due to the technology becoming more affordable and easier to use (e.g., smartphones in the consumer market).

4.1 Quasi-AR Systems

Quasi-AR systems are those setups that do not respect Azuma et al.'s [8] conditions (see Section 2.1). Such setups, however, can still exemplify aspects relevant for full AR systems. For instance, Benford et al.'s [16] virtual theater was a Mixed-Space Collaborative Work. It demonstrated that *being aware* of another space is primordial if one wants two groups to pay attention to each other. Also, to better support

social presence, a common *spatial frame of reference* should be created, which allows users to better measure relative collaborators' positions and orientations and to better *feel together*. Benford et al.'s work [16] is thus a clear example of why *presence* should be considered in AR-CSCW systems.

Dunleavy et al.'s [49] spatially-aware educational system supports social presence and engagement from students who were solving riddles using location-aware smartphones. Their system, however, is not a fully AR one since it does not align 3D virtual contents in the 3D space. Maimone and Fuchs [108] proposed an algorithm and a proof-of-concept of 3D teleconferencing using several RGB-D cameras to capture a remote 3D scene. They aimed to display the remote scene on a 3D stereoscopic screen, which is rendered correctly by tracking the viewer's eyes. While they are not full AR systems, these systems still increase immersion for users.

4.2 Time–Space Collaborative Matrix

The time-space matrix is a well-known CSCW taxonomy [11]. We now discuss its implications for AR-CSCW approaches.

4.2.1 Asynchronous Collaboration

Little work explored asynchronous AR contexts [78], with only three surveyed papers [79], [85], [115] categorized as such that discuss the asynchronous creation and visualization of annotations. Irlitti et al. [79] studied how physical 3D tokens guide collaborators faster and with fewer errors to virtual anchored tags. The use of an AR-HHD with a projector instead of a whole SAR setup, however, may have biased the results due to its low FoV. Kasahara et al.'s [85] system propose to collaboratively tag an outdoor environment using AR-HHDs. Finally, Mora et al.'s *CroMAR* [115] allows users to visualize, in an outdoor environment, geolocalized tags (e. g., tweets), to share viewpoints via emails (which can be visualized with traditional tools), and to rate these tags.

Irlitti et al. [78] suggested to divide collaborators into (1) *producers* generating information to be visualized by (2) *consumers*. Because asynchronous collaboration often revolves around annotations, they proposed three research questions: (1) what are the time effects on collaborative work, (2) how to capture and later visualize AR annotations, and (3) what are the effects of the lack of communication cues.

The visualization of annotations may also depend on the output dimension. Indeed, SAR environments are generally bound to a specific room, which reduces their compatibility for asynchronous distributed contexts, while AR-HMDs may work everywhere. Moreover, systems relying on multiple output devices may also diverge a lot from standalone systems because some devices may not be AR-compatible (like the HHD used for input) or because they propose different AR views and functionalities, e. g., one output device used for annotations only. We can also differentiate AR applications displaying virtual objects tightly bound to real objects (e. g., AR situated analytics [52]) from those not proposing such bounding (e. g., climate visualization applications). The later can rely on traditional tools [170], [171] which should be considered as well.

Finally, as we explain in Section 4.3.1, AR-CSCW applications may use both an exocentric and an egocentric view, which may impact how the registered annotations should be rendered. Indeed, we can imagine an asynchronous application were users are always, for each work session, in pair in a synchronous manner. In these work sessions, users may need to get an overview in both view modes, which impacts the visualization of asynchronous annotations.

4.2.2 Synchronous Collaboration

Figure 2 highlights that only a few synchronous co-located setups were studied during the past decade, compared to distributed ones. We hypothesize that distributed setups require more research because awareness, one key functionality of CSCW, is more difficult to provide in a distributed than in a co-located setup, where co-workers directly see each other.

Co-located Collaboration: As embodiment is natural in co-located systems, awareness of others is better perceived since the *bodies and consequential communication* cues [69] follow social rules. For instance, in Rekimoto's [144] system, users could easily interpret others' deictic gestures. Rekimoto also found that users are mostly absorbed in the workspace. The cues, if any, should thus reach the workspace and not stay on users. This absorption may explain why Nilsson et al. [118] and Prytz et al. [142] did not notice a reduction of the communication quality by occluding the users' eyes. The communication was instead supported by other cues still visible without artifacts, such as head orientation and body gestures. Sometimes it can be problematic, however, to have physical bodies as two users may want to have physical access to the same locations. In such cases, one strategy is for one user to create virtual clones of the target object [132] and manipulate the data [176] or to create slightly different virtual coordinate systems in some specific cases [122].

Distributed Collaboration: Many fields are interested in AR-CSCW as discussed before. However, co-workers are not always physically able to be at the same place, hence the need to collaborate remotely. Billinghurst et al.'s tool [24], for instance, was one of the first teleconferencing systems to use AR. The main focus of such a system is to improve the sense of presence (e. g. [46], [88], [169]). They also rely mostly on asymmetric technologies (see Section 4.3.1), with remote guidance and video conferencing as common scenarios.

Some surveyed papers focused on remote guidance using annotations due to body language that may be partially or not-at-all conveyed. Annotations are thus a key component of CSCW systems which tend to improve the overall communication. For instance, Stafford et al. [156] showed a god-like interaction technique where outdoor users can see objects placed by an indoor user (e. g., cans). The indoor user has an exocentric god-like view and places objects using a table surface augmented with cameras, supporting navigational cues such as "go here" by annotating landmarks.

For teleconferencing, Kim et al. [88] showed that pointing cues and annotations created by a remote person allow all users to feel more *together*, *connected*, and facilitate a better *mutual understanding*. They argued that video streams are better from a user's eye perspective than hand-held views and that virtual pointers are better than virtual annotations when users can communicate vocally. Users could draw annotations directly into the moving video stream, and to freeze it to draw world-stabilized annotations. However, the authors argue that both modes may be inconvenient. Gauglitz et al.'s [65] world-stabilized technique allows a remote user to draw

in a 3D reconstructed world, first constructed by panorama mappers, and then by SLAM techniques [67]. The remote user can then navigate in this reconstructed world with a stabilized view, which is continuously updated by the local user's viewpoint. Both experiments showed that their technique performed better than video-only. The authors did not find evidence of a difference between video stream-stabilized and world-stabilized, which users, however, preferred. The authors later improved the remote user's user interface (UI) [66], introducing spray-like annotations (i. e., each point p_i is assigned a depth d_i) and planar (dominant or not) annotations (i. e., all p_i are on a plane). Users preferred the dominant plane method. Lien et al.'s algorithm [101] may improve this anchoring, hence the call for another study. Later, Huang et al. [75] proposed *HandsInTouch* to capture and display the remote user's hands on a multi-touch tablet and investigated sketching features. They found that the combination of both remote hand rendering and sketching may be better for some specific tasks, at the cost of a higher mental demand, which may be influenced by their UI design. Ou et al.'s system [127], finally, allowed remote users to draw normalized (i. e., recognized patterns) or free-hand strokes on a video stream. Their study showed that strokes improve the performance and that their auto erasures after a defined duration improves it even more. Normalized strokes can be understood once and then be cleared, as soon as the local user grasps the idea. The authors suggested it to be configurable.

In remote expert guidance scenarios, the cited systems implemented the annotation feature in another set of output and input devices accessible only by remote users, creating a technological asymmetry between local and remote users.

4.3 Technology and Role Symmetry

Table 4 shows a strong relationship between the role and technology dimensions, with 45 out of 65 papers ($\approx$ 69%) sharing the same values regarding these two dimensions. Out of all 36 papers that discuss remote collaboration, 28 ($\approx$ 78%) use partially or fully asymmetric technologies. This can be explained by expert guidance (i. e., an asymmetric role) being the most studied remote scenario (see also Ens et al.'s survey [56]), which typically relies on asymmetric technologies to suit both the local users performing the task and the remote experts giving instructions.

4.3.1 Technology Asymmetry

Technology asymmetry concerns systems where users benefit not from the same types of output devices simultaneously. Because the physical world is visible with AR devices, it is hard to keep spatial coherence between all remote users. In effect, all but two papers [2], [102] that discuss distributed asymmetrical setups rely on remote users using traditional screens or VR technologies. For instance, Komiyama et al.'s [91] *JackIn Space* allows seamless transitions between first-person views (i. e., remote users' viewpoint) and a third-person view (i. e., bird's view). Accessing both views was preferred by users over only having access to first-person views. Lehment et al.'s [100] merging algorithm may be appropriate for two remote users both using AR devices, yet their solution is limited to two separate rooms. Still, a study might be required to measure the efficiency of such consensus regarding the immersion and social presence metrics.

One of the main research questions in such setup concerns awareness conveyance between remote VR experts and local AR users. Oda et al.'s [121] system allows a remote expert, either via VR or video-streamed AR, to guide a local AR user with virtual replicas. They showed that direct manipulations were faster for assembly tasks than drawing annotations. Piumsomboon et al.'s [136] *Mini-Me* displays an adaptive remote VR user's avatar always visible to an AR user, either as a mini or as a life-size avatar. The avatar conveys pointing actions, is surrounded by a halo notifying the mini avatar entering of leaving the local user's FoV, has a blowing ring location cue for the actual life-size avatar, and has a graphical distinction between the mini and the life-size avatars. Piumsomboon et al.'s [138] *CoVAR* facilitates remote collaboration between AR and VR users. The AR environment was captured, reconstructed, and sent to the VR user, who can control his/her scale (i. e., magnified or minified). FoV and head, hand, and gaze orientations were captured and shared. They later studied these cues [134] and found that a 3D frustrum denoting the remote user's FoV improves performance, which increases even more when coupled with head or eye gaze cues. Surprisingly, they found that eye gaze alone is worse than head gaze, which can be explained by the eyes tendency to look everywhere. They also found that the balance of actions between the VR and AR users is not symmetrical, mostly due to the larger FoV the VR user has, improving his/her efficiency. Piumsomboon et al.'s [137] remote collaboration system allows a local AR user to manipulate, using a tangible device, the remote VR user's position. The study showed that the control a local user has over a remote user's orientation should depend on scenarios, and that displaying the FoV in addition to an avatar was preferred among participants. Adcock et al. [2] studied three techniques to track a remote collaborator's point-of-view. They found that, even if wedges were not valued in users' qualitative responses, they performed well in the quantitative results in spotting 3D locations via 2D cues. The authors suggested to couple wedges with shadows cast from the location, as local users felt more comfortable with them due to their everyday life similarity.

Lee et al. [96], [97] showed to a remote VR user a 360° stabilized live panorama view of a local AR user. The VR user saw the AR user's FoV, but controlled their own viewpoint to avoid sickness. The VR user's hand positions were shared to support non-verbal communication such as deictic cues. Surrounding halos allowed the local user to constantly track the guest's hands, which improved performance. They found that independent views improved social and spatial presence, require less mental effort, and were preferred by most participants, who also appreciated the visual cues.

In addition to awareness cues, annotations may improve further users' co-presence and performance. Teo et al. [164] showed that annotations, in addition to pointing action augmentations, improved even more the local AR user's understanding about what the remote VR user intends to do.

From these examples we can learn that awareness improves communication and co-presence, hence increases users' performance. Not all awareness cues, however, should be continuous, or the high amount of information can lead to distraction [38] and cognitive overload [178].

All papers discussed so far address one remote user collaborating with one local user. Thanyadit et al. [165] proposed an educational system where multiple students are in a VR environment, supervised by an AR professor. This is the only paper we found which focuses on one-to-many scenarios using AR. The authors listed awareness cues, views, and algorithms that place the students in the environment to maximize the professor's efficiency, depending on the scenario (helping individuals vs. supervising the group).

Finally, with AR increasingly being used in mobile scenarios, the question of modifying the environment where other people live [140] arises. This research area about ethics is, however, not well explored yet. Only two papers in our survey [85], [140] relate to it; see also Ens et al.'s survey [56].

Mixed Outputs in Co-Located Setup: Some co-located setups rely on multiple output technologies. Kiyokawa et al. [90] introduced one of the first mixed-space collaborative systems, in which users switch between AR and VR with one HMD. In VR mode, users can change their position and scaling without physically moving. The *MagicBook* [27] is a physical book that is augmented with 3D scenes to illustrate the text thanks to embedded markers. Users could either be in an exocentric AR view or in an egocentric VR view where they are in the book scene and can be viewed by AR exocentric users as 3D avatars. Several years later, Roo et al.'s [147] system allowed users to be either in an exocentric SAR or to wear a VR-HMD to get access to more complex virtual contents and being able to enter in the mock-up objects in an egocentric way. The exocentric users saw the egocentric ones as 2D arrows. Clergeaud et al.'s [41] system allowed a user to visualize and manipulate, in an egocentric viewpoint using a VR-HMD, an engineering prototype, while the others saw the prototype and the VR user in an exocentric SAR view.

Through these examples we see a distinction between a *collaborative mode* where each user shares the same space, which appears mostly in AR, and a *semi-private mode* where users are immersed in an egocentric view to get more insight regarding the object of interest, which mostly appear in VR. VR technologies are, however, not required. The *SHEEP* management game [106] allowed users to control sheep during their lifetime, rendered through several output devices: AR-HMDs, AR-HHDs, laptop screens, and SAR. While the UI was adapted to the display type, the authors did not demonstrate scenarios where such a number of displays would be useful. Butz et al.'s [36] *EMMIE* system incorporated multiple devices, which were associated to their own input paradigms, to exchange graphical object information. This environment accommodated HHDs, traditional computers, wall-sized displays, and AR-HMDs, all sharing the same database. When needed, a user could transfer an object from the AR-3D space to a laptop for further analysis.

4.3.2 Technology Symmetry

Compared to distributed systems, co-located ones rely more on symmetrical technology, with 21 out of 29 co-located papers ($\approx$ 72%) categorized as such. Technology symmetry leads to the use of fewer output device types. We saw only four papers with multiple output devices, including three which used traditional screens in addition to an AR output.

Role Asymmetry: Six of the papers we surveyed rely on asymmetric role and symmetric technology between users, while three of them concern awareness. Oda and Feiner [120] studied object selection in a 3D space captured by depth cameras. Sodhi et al. [154] used AR-HHDs and several depth cameras to allow remote users to point at, annotate, or manipulate virtual objects in the local user's environment. Le Chénéchal et al.'s [95] technique stretched a remote expert's arms from a local user's shoulders for guidance tasks. We clearly distinguish here between active and passive users.

Subjective views allow users to see the workspace in a way adapted to their roles. Nilsson et al. [118] and Prytz et al. [142], for example, studied crisis agents (e. g., police, military) using an AR-CSCW system to support discussion, with each agent having a customized view. Similarly, players see their opponent's tiles as blank tiles in Szalavari et al.'s [161] Mah-Jongg gaming system.

An expert user can also be the only one manipulating the AR space to empower the discussion this user has with, e. g., a customer or trainee [47].

4.4 Output and Input Devices

The output display usually drives how users interact with the system. For example, a mouse and a keyboard do not fit an AR-HMD well when users move in the 3D space. We thus group the different approaches by their main display.

The choice of the output device can significantly alter the users' performance. For instance, Müller et al. [117] showed that landmarks increase the sense of presence by creating a common ground between remote users. The use of HHDs, however, may have biased the speed results: Compared to HMDs, the HHDs' narrow FoV and lack encouragement for users to look around may have introduced the delay Müller et al. [117] noticed at the end of their study.

While some systems do not enforce the same technology to every user (see Section 4.3.1), others give users the same set of output and input devices. Butz et al.'s [36] system allowed users to benefit from different input paradigms (touch-based for HHD, tangible for AR-HMD, keyboard and mouse for PC), these outputs possibly being complimentary (e. g., using a tracked HHD to see complementary objects through an AR-HMD). Via these sets of output devices, users could use personal devices for private information (e. g., HHD). They also used a red spotlight as a metaphor to indicate private properties of a 3D object. Later, Benko et al.'s [17] archeology visualization software combined multiple output devices. Due to the low resolution of AR-HMD resolution at the time, they used a tabletop display and a tracked HHD to show high-resolution documents, while keeping AR-HMDs for 3D-stereoscopic views and 3D input.

These displays still have some features in common. They usually provide all users with an active device which can display private and subjective contents per user, the latter being reviewed in Section 4.3.2. Users found the distinction between public and private views useful in Billinghurst and Kato's [26] informal trials of an AR web browser.

4.4.1 AR Head-Mounted Display (AR-HMD)

AR-HMDs have been extensively used in the past, with 10 ($\approx$ 15%) papers relying only on, and 26 ($\approx$ 40%) partially on, AR-HMDs, out of all 65 AR-CSCW papers we surveyed.

AR-HMD users can share their viewpoint to others which might be useful in distributed setups (see Section 4.3.1).

Moreover, because AR-HMDs are still expensive, because not everyone feels comfortable with them, and because there may be too many passive users, Franz et al. [64] allowed these passive users to grasp, via a traditional screen set on a wall, the active AR user's experience who visualizes an augmented 2D landscape map rendered on a tabletop. Franz et al. [64] proposed both an *over the shoulder* view and an abstract data representation of the object of interest. They showed that the *over the shoulder* view forced the active user to constantly look at the object of interest for the passive users to see it as well. Displaying the intersection with the AR user's head gaze on the tabletop improved the communication, the sense of presence, and the usability of the system.

Display Characteristics: AR-HMDs tend to fit the users' anatomy, which might be useful in engineering and educational scenarios relying on 3D data. Dong et al.'s [48] *ARVita* used an AR-HMD to augment a tabletop displaying markers. The tool's goal was to improve students' learning of engineering processes. For the authors, using a tabletop allows users to have a better discussion, when compared to traditional tools. Wang and Dunston [167] demonstrated that users were faster in the civil engineering field using AR, compared to traditional perspective sketches. However, participants were uncomfortable with the HMD used at that time, which may change with modern AR-HMDs (e. g., Microsoft's HoloLens). They also found that navigation cues (e. g., a compass) are useful because multiple 3D transformations such as rotations and scaling can confuse users who are trying to orient themselves. Similarly, Wang and Dunston's study [168] showed that AR-HMDs drastically increase the speed in error detection tasks in civil engineering field, when compared to traditional tools.

Input Modalities: Not all input modalities are available by default on AR-HMDs (e. g., touch, keyboard and mouse) due to relative position and portability constraints. They often rely on voice, tangible, and gestural interfaces. Ong and Shen's [126] CAD system allowed co-located users to manipulate and model common 3D shapes using tangible devices. For consistency, only one user could modify the 3D model at a time. When finished, users committed their changes that others could then (in)validate. The *CoVar* system [138] combined VR and AR-HMDs to benefit from both technologies. Users were able to move objects closer or further away with mid-air gestures. The *GARDEN* system [120] studied, for the selection of 3D objects, how a user can refer to an object, highlight the selected object, and how others acknowledge it. This technique was more accurate than laser pointers, virtual arrows coming out from the controller, and video sharing techniques. But their technique was also slower, which can be explained by the higher number of steps required. Mahmood et al.'s system [107] allowed users to remotely collaborate and visualize geospatial data using AR-HMDs. Their user interface, allowing history recording, 3D transformations, and data filtering, relied on both voice and gestures. Their system, in addition to Lehment et al.'s [100] consensus algorithm, is the only surveyed system using solely AR-HMDs in distributed collaboration, which may be due to their application domain that does not require a strong relationship between virtual and real objects.

While AR-HMDs do not propose their own touch input, several systems use touch support by adding other input devices that provide 2D input [149] and/or tangible control [36]. Indeed, when mobility is less important, AR-HMDs can be coupled with a tabletop and 2D views [48], [64], [75]. Wang et al.'s [169] teleconferencing system regrouped, per local setup, one camera and one screen for each remote co-worker with whom this local user interacts. This improves workspace awareness and spatial faithfulness. Each user had a tabletop displaying markers augmented by AR, conveying remote collaborators' arms and objects of interest.

4.4.2 Hand-Held Display (HHD)

HHDs were also often used, with 20 ($\approx$ 31%) surveyed papers relying partially ($n = 11$; non-AR use-cases: $n = 2$), or fully ($n = 9$; non-AR use-cases: $n = 1$ [49]) on them.

HHD as a Complement: HHDs are often used to complement a larger system. Surale et al. [159], e. g., explored the role of a tablet in VR CAD contexts, which can be applied in AR. The *Studierstube* [149], discussed in Section 2.5.1, used a personal pad to add touch modalities, without using it as an output device. MacWilliams et al.'s system [106] used HHDs as spatially-aware tangible objects, and Butz et al.'s system [36] used HHDs as a magical lens for users not wearing an HMD, or a magical mirror to capture an object. Benko et al. [17] used an HHD as a magical lens to visualize documents in a higher resolution. AR-HHDs, if coupled, have mostly been used with AR-HMDs in co-located contexts. Only one paper in this survey [106] coupled an AR-HHD with SAR to add tangible input. Future work may investigate how the sensors of such devices, in addition to their high display resolution, improve the collaboration in SAR-CSCW contexts.

Ubiquitous AR-HHD: Touch devices such as smartphones are highly present in the consumer market, which can be used as affordable and widespread AR devices. Huynh et al. [77] claimed that AR-HHD collaborative games enable a kind of social play experience not present in non-AR counterparts. AR-HHD games can use the limited viewport as a feature, e. g., to intentionally hide information from the player. Their user study showed that, except for tracking issues, players enjoyed playing this kind of game. Occlusion from one user's device, however, may disturb the visual tracking of the other devices. Hence, multiple sources of tracking (e. g., SLAM and visual) may be needed. Bhattacharyya et al.'s [23] game design iterations showed that AR-HHD are so ubiquitous that users tend to use them in a specific way, e. g., tap-and-hold for selecting and moving objects, instead of gestures which are seen as awkward and cumbersome interactions. Due to the users' experience of 2D applications, some cues should be present to encourage users to physically walk in their 3D AR environment.

Awareness of AR-HHD: While we do not imagine to vertically align the users' heads, it may not be the same for hands holding AR-HHDs. Oda and Feiner [122] addressed this collision issue by applying small shifts between the virtual and the real environment for each user, which improved performance without disturbing the users. This should work only, however, when strong spatial coherency between the virtual and the real environments is not mandatory, or users may be confused. Moreover, AR-HHDs occupy the hands and force users to look at their screens, which may not be appropriate for all cases. Nilsson et al.'s [118] iterative study showed that users preferred to use an HMD over an

HHD to free the hands, allowing users to interact both with other people (social communication) and with the physical workspace. This may explain why Sohdi et al.'s [154] work is the only surveyed work which tracks users' hands with an AR-HHD, even though the hands are holding the device.

Despite these limitations, the movements of the tablet are visible by all, which everyone understands as manipulating the AR world in co-located contexts. In our survey, the only work focusing on awareness is Rekimoto et al.'s [144], which proposed to add an *eye gaze* equivalent as an awareness cue.

4.4.3 Spatial Augmented Reality (SAR)

Most of the SAR papers in our survey discussed how the whole environment becomes tangible. Therefore, we already mentioned some papers (e. g., [133]) in Section 2.5.2.

SAR systems, typically by adding several sensors in a room, can scan the room and enable mid-air gestures by distinctly tracking each user's body. SAR setups are used in distributed and co-located settings. In distributed settings, Pejsa et al. [129] brought a remote user into a local user's SAR space. This works well when the remote user is virtually close to their anchoring position. They compared their setting with *Skype* and *Face To Face* regarding *speed*, *presence*, and *communication efficiency* for assembly tasks. One further study could be to compare their settings with remote AR-HMD ones. Lincoln et al. [102] showed how SAR can support remote medical consultations by capturing the doctor's body at one location and using SAR at the other location. They proposed to capture the doctor's head and retransmit it onto a remote animatronic head using projectors.

In local settings, Benko et al. [18] allowed two face-to-face users to see in perspective without wearables. The personal tracking and the perspective facilitate collaboration and, even with depth cue conflicts, users were able to determine distances and pointing directions. Wilson and Benko [173] allowed a maximum of six users to move, in a tangible way, media objects between surfaces presented in their smartroom.

Finally, SAR environments can be non-intrusive, which might be useful in contexts where people expect and are used to some pre-defined behaviors (e. g., pencil selling [47].)

Most of the surveyed papers using SAR in all endpoints are symmetric regarding the role axis. The other work focuses on passive vs. active users [47], [102]. One interesting research question could be how SAR can permit parallel work, with users assigned to different roles. This setup would differ from other works using multiple types of output devices suiting the different roles between users.

5 DESIGN CONSIDERATIONS

Most of the papers we surveyed for remote settings focused on studying which workspace awareness (WA) cues improve user performance. Others provided insights about the benefits and limitations of output and input devices in local setups. Moreover, some early work showed the capabilities of AR systems applied to specific scenarios. These are usually linked to discussions about how AR influences mutual communication and the cognitive capabilities of users. Finally, a minority of papers discussed techniques to capture users and environments and to merge local and remote settings. Based on these contributions, we now provide high-level design considerations we extracted from these papers.

5.1 Private/Shared Views, and Awareness Cues

Embodiment in remote systems provides information about the users' locations. The general focus of a collaborator can be provided through FoV representation, and eye- and head-gaze cues [134], [136], [137], [138]. FoV cues work best and eye-gaze cues are not sufficient [134]. Embodiment can rely on hands-only [75], head-only [107], exocentric life-size avatars, or egocentric mini-avatars [136] which allow workers to keep the remote collaborators' expressions in sight. In co-located scenarios with a shared exocentric view, most of the listed cues come for free and can even be removed. A weaker eye contact due to wearing an HMD, e. g., can allow users to save their hands for 3D interactions and body gestures [118].

In both local and remote scenarios, WA cues are usually needed. Our survey showed that cues, such as compass or arrows [167], support navigational tasks [156]. To create a *common "virtual-real" space* [117] in distributed settings, systems can render additional virtual objects. Finally, annotations support remote discussion, see Sections 4.2.2 and 4.3.1.

Users may need to customize their views based on their role [118] or use private egocentric spaces [107]. Such a space, however, can complicate the communication with co-located collaborators due to a lack of shared information. If a user manipulates their private space, the physical appearance of their head and body with respect to another user's 3D scene can be misleading. One way to avoid such confusion is to enforce the switch between (1) an AR mode for when users want to share the same exocentric 3D scene and (2) a VR mode for *all* users when private, egocentric views are desired, so that one user's physical head and body will not give misleading cues to other users (e. g., [90]). In the specific case of distributed collaborative VR/AR scenarios, remote VR collaborators should control their own viewpoints for comfort, and FoV/head-gaze cues should be adapted [97].

Awareness cues can also be added in AR or VR to indicate each collaborator's virtual position and orientation [147]. By reflecting on all the surveyed papers, an interesting scenario to investigate another kind of awareness cue may be the following. If Alice and Bob are co-located, and Alice wishes to employ a private egocentric view, but Bob wishes to remain in AR to, e. g., share exocentric views with Carol, Bob's headset may display a virtual mask over Alice's to indicate that she no longer sees the same exocentric 3D scene.

5.2 Distributed Work Using Asymmetrical Technology

Most of distributed AR-CSCW systems we surveyed rely on either two similar rooms for both endpoints, or on one remote user using a VR-HMD while the local user relies on AR (Section 4.3). Relying on asymmetrical technology (e. g., AR-HMD + VR-HMD) makes sense when there is a strong relationship between the virtual and the real environments in the local space. When both users have the same role, we do not envision that using AR devices for both users should be considered if virtual objects are tightly related to real objects.

However, as Mahmood et al. [107] showed, using AR devices in both endpoints can be considered to have flexibility in both places when the virtual objects can float around.

5.3 Multimodal Interfaces Improve User Experience

Hybrid systems aim to support a heterogeneous mix of hardware for both output and input (e. g., HMD, HHD,

traditional screen; mouse, keyboard, mid-air gestures). The chosen AR hardware may depend on the available devices, the users' roles, and how much effort the users are willing to invest to participate [64]. Mouse + keyboard input, for example, is typically less tiring than mid-air gestures [81], and a 2D screen, projector, or tabletop can be casually used by multiple users [64], [139] if they are mostly "passive."

Through our survey, we found four roles for traditional screens. They, first, permit experts to follow what a remote AR user is doing and allow them to send annotations (e.g., [67]). Second, they allow non-AR users to follow a discussion about what a co-located AR user is visualizing (e.g., [36]). Third, they provide a common ground for 2D information (e.g., video stream [12]). Forth, they allow users to work with their usual tools and profit from high-resolution screens [17].

Hybrid systems using personal 2D screens can provide user-friendly personal 2D views and UIs [36], [118], [144].

5.4 Synchronous Workspace Consistency

The ownership of manipulated objects can support workspace consistency because it facilitates a more robust synchronization [126], [144]. If an object is at the center of the discussion, private proxies acting for public counterparts can be visualized with a correct viewpoint and allow users to direct manipulate them [121], e.g., using a private World-In-Miniature [157]. Private objects can also be shared and frozen by the owner of the object [107]. In an AR-CSCW context with more than two users, awareness cues can indicate which user is remotely manipulating or owning a public object, making it easier for others to track the users' actions. This can happen, e.g., through the decorations of the 2D/3D window-contexts [107], [149] introduced in Section 2.6.

6 REMAINING RESEARCH AREAS

In addition to these design considerations for established environments, we now discuss some remaining research areas based on the survey. We intentionally avoid overlap with Ens et al.'s [56] recent survey which already mentions (1) complex collaboration structures in time, space, and symmetry, (2) convergence and transitional interfaces, (3) empathic collaboration, (4) collaboration beyond the physical limits, and (5) social and ethical implications.

6.1 Co-Located Awareness in Interactive Environments

Few surveyed papers investigated synchronous co-located setups (Table 4 and Figure 2). Marai et al.'s [109] study showed, however, that a team is the most productive with such a platform. Questions about providing awareness to evaluate collaborators' internal state (e.g., privacy status), annotations' (e.g., who puts what?), workspace manipulation, and replaying previous actions and handling the undo/redo actions performed by multiple users (which relate to Mahmood et al.'s [107] interactions and designs) are areas that, to the best of our knowledge, are not yet well covered.

How viewpoints are shared so that each user is aware of others' views is not yet answered either. Proxies allow users to see the target objects without occlusion, but they may lose the correct orientation and spatial referencing. Users may need to move around to see the object. Otherwise, if

viewpoints are shared via a video stream (i.e., collaborator A directly sees B's point of view), it remains unclear how coherency issues can be handled, because collaborator A will, in this example, see their own body through B's view.

6.2 Annotation Rendering in Co-Located Settings

Few surveyed papers [67], [79], [85] supported 2D annotations (canvas), but they were rarely meant for multiple users. Should they be rendered in an egocentric way, exocentric way, or a mix of both? As a first approach, we hypothesize that it may be preferable to render annotations in an exocentric view for *workspace coherency*, while having an egocentric view bound to the exocentric one for better *legibility*. Still, more studies are needed to understand possible approaches in regard to these criteria. The on-the-fly creation of anchored annotations is also not well explored, as most studies we encountered focused on visualizing already registered annotations. For such tasks, the implications of the interaction techniques depending on the output dimensions remain unclear, i.e., how well these techniques support the collaboration and mutual understandings of all users.

6.3 Role and Technology Asymmetry in Remote Work

Most work on expert guidance (see Section 4.3.1) focused on one expert guiding one user, except for Poelman et al.'s system [139], which allows multiple experts to guide a remote user without further exploring that direction, and Thanyadit et al.'s work [165], which allows one professor to supervise multiple students. The management of an undetermined number of experts guiding an undetermined number of local users remains to be explored. Similarly, more work is needed to understand how awareness and annotations could be rendered so that local users understand who is doing or has done what. Moreover, all surveyed papers use an AR-HMD at one endpoint, no matter what devices (VR or AR) are used for the other endpoint. We summarize that, to follow a local user's position with a VR-HMD, the most common approach is to track the local user's head. We found only one paper that associated the remote VR user's position with a local AR user's tangible device [137]. Head tracking can also be achieved in SAR but reduces mobility. It remains unclear if a remote VR-HMD user can be directly coupled with an AR-HHD to keep the mobility, and what benefits and limitations such coupling entails.

6.4 Output and Input Devices in Asynchronous Work

In Section 4.2.1 we discussed asynchronous AR-CSCW systems, which mostly concern the creation and consumption of annotations [78], [79], [85]. Because multiple input modalities can benefit AR, it may be interesting to understand their implications in collaborative environments. Understanding how output devices influence the collaboration, and how coupling several devices changes the creation and consumption of annotations [78] are also research areas interesting to investigate. Moreover, because AR devices can provide both exocentric and egocentric views, it remains challenging to capture and reproduce recorded actions in both views.

6.5 Mixed Input Modalities

Mid-air gestures, although considered by many to be natural forms of input, are often not preferred for long-time use because they induce fatigue and have low position accuracy. In 3D selection tasks, for example, existing hand tracking techniques usually result in low precision. This issue is further amplified when one wants to point at a location beyond arm's reach. Oda and Feiner [120] tackled this problem using a special tracked device, but it could be cumbersome for existing systems. Moreover, their selection speed drops a lot. Under this circumstance, hybrid interfaces (e. g., AR-HMD+HHD [151], AR-HMD+PC [170], [171]) with AR are currently being studied for visualization, both for more effective input and as a combination with traditional work environments. However, the mapping between the different input and view modes (e. g., 2D input with the 3D AR view [104]) remains challenging. More research is also required to understand when and why users switch from one input device to another, possibly depending on the action and the needed accuracy. Finally, these private 2D interfaces provide only poor awareness cues, as the input does not reach the public workspace that is visible to others [144].

6.6 Comparison Between AR-HHD, AR-HMD and SAR

Although we are aware that some hardware is unavailable to some due to physical constraints, in relatively few of the papers we surveyed the authors specified why they selected their used output devices (e. g., Nilsson et al.'s [118] informal study). More research is required to understand when and where would a specific display be preferred, which may depend on the tasks and the users. Since several settings combine HHDs with other output devices (e. g., [17]), we wonder when an HMD combined with a multi-touch tablet for 2D input is more preferable than a standalone HHD.

One may argue that co-located gestural communication should only use well-established gestures. AR-HHDs may not support those, however, because users may struggle to look at both AR objects and their collaborators, as Rekimoto et al. [144] pointed out. Moreover, by holding an HHD, users cannot use their hands as they would otherwise do (e. g., body language), as we discuss in Section 4.4.2. Past work [17], [36], [106] suggested to couple an AR-HMD with an HHD, which then can be lowered and held into one hand. Only one paper [106] proposed to couple HHDs with SAR displays. As we state in Section 4.4.2, only more research can show how to combine HHDs and SAR devices w.r.t. the sensors and high display resolutions of typical HHDs.

7 COLLABORATIVE IMMERSIVE ANALYTICS (CIA)

Quite a few of the papers we surveyed tackle *analytics and visualization* tasks (our domain expertise) using AR. For this reason, we provide an overview of CIA next.

7.1 Immersive Analytics

Immersive Analytics (IA), first defined by Chandler et al. [39] as *"an emerging research thrust investigating how new interaction and display technologies can be used to support analytical reasoning and decision-making,"* is at the junction of MR Environments and Visual Analytics [166] (Figure 4).
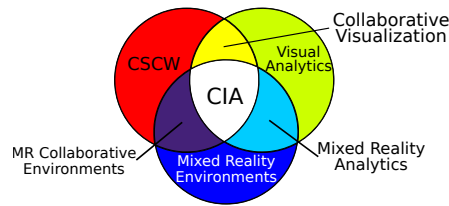

Fig. 4: CIA relationship, adapted from Billinghurst et al.'s [25] diagram.

AR can be a medium to support users in their analytical tasks [15], [53], [112], collaborative or not. Such systems need high computational power for both simulations and rendering to provide quick feedback. IA projects use VR platforms more often than AR ones, and only a few projects investigated IA using AR [60]. This lower interest for AR in IA may be explained by the fact that VR platforms are more mature and easier to develop for and by the higher constraints of AR devices, notably regarding tracking. We also hypothesize, regarding hardware platforms for IA, that

SAR is similar to the CAVEs used by VR researchers which, due to their size capacity, allow users to see each other and to use personal devices, e. g., laptops;

HHDs have limited value for 3D data visualization because they are non-stereoscopic and, compared to a desktop machine, are less convenient to use because they are often less powerful for rendering, have smaller screens, lack a keyboard and mouse (useful for scripting), and have less mature software tools for analytics; and

AR-HMDs have lower computing power than VR-HMDs by being embedded systems, while having more constraints than HHDs to bring immersion (e. g., higher resolutions and frequencies, lower weight). This may, however, change with the arrival of new powerful AR-HMDs, such as the Microsoft's HoloLens 2 with its Snapdragon 850. Compared to VR-HMDs, AR-HMDs might be considered for situated analytics (e. g., medical data).

7.2 AR Collaborative Immersive Analytics (AR-CIA)

Collaborative Immersive Analytics, which Billinghurst et al. [25] define as *"the shared use of new immersive interaction and display technologies by more than one person for supporting collaborative analytical reasoning and decision-making,"* is a field intersecting CSCW, Visual Analytics [166], and Immersive Environments (Figure 4). At the intersection of CSCW and visual analytics lies *collaborative visualization*, which Isenberg et al. [80] define as *"the shared use of computer-supported, (interactive) visual representations of data by more than one person with the common goal of contribution to joint information processing activities."* At the intersection of CSCW and MR Environment lies *MR Collaborative Environments*. *MR environments* intersect both visual analytics and MR environments.

Although researchers studied many scenarios in the past (e. g., security [118], archaeology [17], engineering [167]), visualization has thus far been largely neglected, with Mahmood et al.'s [107] work being the only one we found published during the last decade that tackled visualization.

7.3 Should 3D Visualization be Immersive?

Past research (e. g., [63], [76], [92]) showed that immersive technologies can significantly enhance the way people understand 3D data compared to traditional interfaces.

TABLE 4: The selected papers

Year	Reference	Space	Time	Role	Tech.	Output	Input
1996	Rekimoto [144]	co-loc.	sync.	sym.	sym.	AR-HHD	non-tracked controllers
1998	Benford et al. [16]	both	sync.	both	asym.	traditional screen	not described [1]
1998	Billinghurst et al. [24]	remote	sync.	sym.	asym.	AR-HMD / traditional screen	mouse (and trackball)
1998	Szalavari et al. [162]	co-loc.	sync.	sym.	sym.	AR-HMD	hands tracking / tangible / touch
1998	Szalavari et al. [161]	co-loc.	sync.	sym.	sym.	AR-HMD	hands tracking / tangible / touch
1999	Billinghurst and Kato [26]	both	sync.	sym.	both	AR-HMD / traditional screeen	head gaze / mouse (and trackball)
1999	Butz et al. [36]	co-loc.	sync.	both	both	AR-HMD / AR-HHD / traditional screen	tracked controllers /keyboard and mouse / touch / tangible
1999	Kiyokawa et al. [90]	co-loc.	sync.	sym.	both	AR-HMD / VR-HMD	hands tracking / head gaze
2001	Billinghurst et al. [27]	co-loc.	sync.	sym.	both	AR-HMD / VR-HMD	tangible / non-tracked controllers
2002	Piper et al. [133]	co-loc.	sync.	sym.	sym.	SAR	tangible / mouse
2002	Schmalstieg et al. [149]	co-loc.	sync.	sym.	sym.	AR-HMD	hands tracking / tangible / touch
2003	Ou et al. [127]	remote	sync.	sym.	asym.	traditional screen	touch / mouse
2003	MacWilliams et al. [106]	co-loc.	sync.	sym.	both	AR-HMD / AR-HHD / SAR / traditional screen	hands tracking / touch / speech / tangible
2004	Benko et al. [17]	both	sync.	sym.	both	AR-HMD / HHD / traditional screen	hands gestures / touch / tangible
2006	Stafford et al. [156]	remote	sync.	asym.	asym.	AR-HMD / traditional screen	hands gesture / tangible / mouse (trackball)
2008	Wang and Dunston [167]	co-loc.	sync.	sym.	sym.	AR-HMD	tangible
2009	Dunleavy et al. [49]	co-loc.	sync.	sym.	sym.	HHD	touch
2009	Huynh et al. [77]	co-loc.	sync.	sym.	sym.	AR-HHD	tangible
2009	Lincoln et al. [102]	remote	sync.	asym.	asym.	SAR	none
2009	Nilsson et al. [118]	co-loc.	sync.	asym.	sym.	AR-HMD / AR-HHD	tracked controllers
2009	Oda and Feiner [122]	co-loc.	sync.	sym.	sym.	AR-HHD	touch
2009	Ong and Shen [126]	co-loc.	sync.	sym.	sym.	AR-HMD	tangible
2010	Prytz et al. [142]	co-loc.	sync.	asym.	sym.	AR-HMD	tracked controllers
2010	Wilson and Benko [173]	co-loc.	sync.	sym.	sym.	SAR	hands tracking / touch / tangible
2011	Maimone and Fuchs [108]	remote	sync.	sym.	both	SAR / traditional screen	none
2011	Wang and Dunston [168]	both	sync.	sym.	sym.	AR-HMD	tangible
2012	Oda and Feiner [120]	co-loc.	sync.	asym.	sym.	AR-HMD	tracked controllers
2012	Poelman et al. [139]	remote	sync.	both	asym.	AR-HMD / traditional screen	hand gestures /keyboard and mouse
2012	Gauglitz et al. [65]	remote	sync.	asym.	asym.	AR-HHD / traditional screen	keyboard and mouse
2012	Kasahara et al. [85]	co-loc	async.	asym.	sym.	AR-HHD	touch
2012	Mora et al. [115]	both	async. [2]	both	both	AR-HHD / traditional screens [2]	touch /keyboard and mouse [2]
2013	Adcock et al. [2]	remote	sync.	asym.	asym.	SAR	tracked controllers
2013	Ballagas et al. [12]	co-loc.	sync.	sym.	sym.	SAR / traditional screen	touch
2013	Dong et al. [48]	co-loc.	sync.	sym.	sym.	AR-HMD / traditional screen	keyboard and mouse
2013	Sodhi et al. [154]	remote	sync.	asym.	sym.	AR-HHD	hands gestures / touch
2014	Benko et al. [18]	co-loc.	sync.	sym.	sym.	SAR	hands gestures
2014	Gauglitz et al. [67]	remote	sync.	asym.	asym.	AR-HHD / traditional screen	keyboard and mouse
2014	Gauglitz et al. [66]	remote	sync.	asym.	asym.	AR-HHD / traditional screen	touch
2014	Kim et al. [88]	remote	sync.	asym.	asym.	AR-HMD / AR-HHD / traditional screen	mouse
2014	LeChenechal et al. [95]	remote	sync.	asym.	sym.	VR-HMD	tracked controllers
2014	Lehment et al. [100]	remote	sync.	sym.	sym.	AR-HMD	none
2014	Wang et al. [169]	remote	sync.	sym.	sym.	AR-HMD / traditional screen	hands gestures / touch
2014	Irlitti et al. [79]	co-loc.	async.	both	sym.	AR-HHD	tangible
2015	Oda et al. [121]	remote	sync.	asym.	asym.	VR-HMD / AR-HMD / HHD	touch / tracked controllers
2016	Gupta et al. [68]	remote	sync.	asym.	asym.	AR-HMD / traditional screen	eye gaze /keyboard and mouse
2016	Pejsa et al. [129]	remote	sync.	sym.	sym.	SAR	none
2017	Clergeaud et al. [41]	co-loc.	sync.	both	both	VR-HMD / SAR	hands tracking / tangible
2017	Komiyama et al. [91]	both	sync.	both	both	CAVE	non-tracked controllers
2017	Lee et al. [96]	remote	sync.	asym.	asym.	VR-HMD / AR-HMD	hands tracking
2017	Muller et al. [117]	remote	sync.	sym.	sym.	AR-HHD	touch
2017	Piumsomboon et al. [138]	remote	sync.	sym.	asym.	VR-HMD / AR-HMD	hands gestures / head gaze / eye gaze
2017	Roo et al. [147]	co-loc.	sync.	both	both	VR-HMD / SAR	hands tracking / tracked controllers
2018	Lee et al. [97]	remote	sync.	asym.	asym.	VR-HMD / AR-HMD	hands tracking
2018	Piumsomboon et al. [136]	remote	sync.	both	asym.	VR-HMD / AR-HMD	hands tracking / hands gestures / head gaze
2018	Poretski et al. [140]	co-loc.	sync.	asym.	asym.	AR-HHD	touch
2018	Huang et al. [75]	remote	sync.	asym.	asym.	AR-HMD / traditional screen	hands tracking / touch
2018	Teo et al. [164]	remote	sync.	asym.	asym.	VR-HMD / AR-HMD	hands gestures
2019	Bhattacharyya et al. [23]	co-loc.	sync.	sym.	sym.	AR-HHD	touch / tangible
2019	Teo et al. [163]	remote	sync.	asym.	asym.	VR-HMD / AR-HMD	hands gestures
2019	Piumsomboon et al. [137]	remote	sync.	both	asym.	VR-HMD / AR-HMD	tracked controllers / hands gestures / tangible
2019	Piumsomboon et al. [134]	remote	sync.	asym.	asym.	VR-HMD /AR-HMD	hands gestures / head gaze / eye gaze
2019	Mahmood et al. [107]	remote	sync.	sym.	sym.	AR-HMD	hands gestures / speech / head gaze
2019	Thanyadit et al. [165]	remote	sync.	asym.	asym.	VR-HMD / AR-HMD	none
2019	Franz et al. [64]	co-loc.	sync.	sym.	asym.	AR-HMD / traditional screen	head gaze / non-tracked controller
2019	Dolata et al. [47]	co-loc.	sync.	asym.	sym.	SAR	tangible / touch

[1] The paper does not describe how they can interact with the virtual world. We suppose it is through hand tracking and mouse/keyboard but we are unsure.
[2] Traditional tools are included because their system allows the sending of emails visualizable "anywhere". We did not consider its synchronous case because it relies on video conferencing without AR support.

AR technologies seem to be promising for 3D data visualization [50], [110]. By merging the real and virtual worlds, users can use their common analytical tools, while visualizing in stereoscopy and using the "huge" (e. g., room-scale) workspace AR provides [170], [171]. As discussed in several education papers (e. g., [48]) and through this survey, students understand concepts better using stereoscopic 3D visualizations, which reduce their cognitive workload. Networks [172] and 3D vector fields [63] are better perceived using stereoscopy and motion cues, which AR relies on. AR could also be appropriate in 3D visual analytical tasks [167], [168], with AR reducing the time collaborators take to detect errors in 3D designs compared to traditional tools.

Immersive technologies, however, are not always better than traditional workstations [10]. One explanation is that users are more used to, and have more experience with, traditional tools. Wang et al. [170] envisioned that immersive technologies should enhance but not replace traditional tools for scientific workflows, mostly because some necessary tools (e. g., scripting tools) are not available in, or are hard to transpose to, AR 3D spaces [170], [171]. Stereoscopic 3D displays (S3Ds) may not be useful for every scenarios and user roles. McIntire et al. [113] examined 184 experimental comparisons of S3D to non-stereo displays, classified in 6 categories: *Judgment of position and/or distances*; *Finding/identifying/classifying objects*; *Real/virtual spatial manipulations of objects*; *Navigation*; *Spatial understanding, memory, recall*; and *Learning/training/planning* tasks. In each of the first three categories, 57% or more of the experiments found that S3D yields better performance. In each of the other three, between 36% to 52% found a definite benefit with S3D. Only rarely did S3Ds perform worse than non-stereo screens.

7.4 CIA Research Questions

Because the AR-CIA field is still new, we propose in this section some relevant research areas to be investigated.

7.4.1 AR Devices?

Because perspective projections have multiple parameters (e. g., far/near clipping plane, FoV), users often rely on orthographic views for some 3D tasks, e. g., drawing 2D lassos being extruded in 3D [22] or measuring distances in CAD software. AR-HMDs, however, lower this limitation by having perspective and stereoscopic parameters matching those of the human vision system. However, such wearable devices currently have lower resolutions and computing power than workstations and scientists still need to use workstation tools and software [170], [171]. A hybrid AR–workstation system, however, may remove the AR interaction benefits due to people following social rules and their limited mobility in their environment. Questions regarding which AR device or which hybrid interface to use for AR-CIA remain to be explored, and surely depend on the domain.

7.4.2 The 3D Coordinate System, AR Display + HHD

Lopez et al. [104] combined a S3D and a multi-touch tablet to propose user-friendly 9 DoF (3D translations, 3D rotations, 3D scaling) transformations. However, in CSCW contexts, there are multiple users, hence multiple local 3D coordinate systems. Should the system select a particular collaborator's

coordinate system for pointing actions? Should one visualize a dataset with the same relative position that another user has to understand the discussion? To track users, should the tablet represent the others' relative position? What features may be required when using this mobile device as both a rendering, touch input, and tangible input device?

7.4.3 Spatial Selections

Filtering Region-of-Interests (ROIs) and specific features are common interactions in 3D visualization software and were studied a lot; see Besançon et al.'s taxonomy [22]. They were not studied, however, in collaborative contexts, with users being either in AR or VR. How can we use these techniques in AR-CSCW where collaborators must understand the actions of others? Should a selected region be displayed only in private or subjective views? If yes, referencing can be challenging in scalar fields as discussed next. Moreover, compared to VR-based systems, users can easily move around, which can be a rich source of awareness.

7.4.4 Subjective Views and Immersive Analytics

Smith and Mariani [153] proposed two dimensions for subjective 3D: *appearance* and *modifier*. In our survey, Nilsson et al.'s [118] system proposed subjective object representations based on the users' roles (appearance dimension). The VR literature also contains work on subjective view [3], [153] and showed some potential issues regarding workspace coherency. However, we did not find studies related to 3D cloud points or volumetric datasets which are explored via a set of filters or transfer functions. These filters significantly alter the transparency (modifier dimension) of some part of the data and not the other, reducing the workspace coherency.

It is then unclear if such subjective filters can be efficient and with what visualization and interaction techniques; e. g., AR devices could switch to one of the collaborators' parameter sets when detecting communication cues (e. g., pointing). Similarly, should such a system use a UI that lists all registered parameters (e. g., one parameter per user plus one default) and select one before starting the discussion?

8 LIMITATIONS

The work we surveyed is scattered among several venues, explaining why only 18 ($\approx$ 28%) papers we discussed are from our systematic search. Our classification might thus not entirely reflect the real breadth and depth of the literature. Hence, the design considerations may be incomplete; and Section 6 discusses on-going research areas. We argue, however, that—while not exhaustive—we have surveyed a large part of the literature so that our discussion of research questions and design considerations are relevant. We also note that some of the surveyed papers may be subject to interpretation. For example, while Lehment et al.'s work [100] was tested with HMDs, the authors claimed that their algorithm is applicable to HHDs. Moreover, the difference between *tracked controller* and *tangible* that we used in our taxonomy is rather small and could be confusing at first. Similarly, *CroMAR* [115] allows users to share their viewpoints by emails which can be visualized in traditional screens, hence their inclusion as output devices.

9 CONCLUSION

In this article we first reviewed fundamentals regarding AR-CSCW, technologies that are used to provide a coherent and effective AR experience, and psychological aspects that designers should pay attention for. We then reviewed a total of 65 papers (18 through a systematic review and 47 additional ones) regarding collaborative works using AR, alone or not (e. g., combining AR and VR), and categorized them along our six dimensions. Based on this review, we extracted a current research agenda, discussed guidelines derived from the literature, and mentioned remaining research areas that yet have to be covered. In addition, we contemplated collaborative immersive analytics using AR technologies. With this survey we thus aim to give an overview of the field for newcomers, researchers and practitioners alike, and up-to-date information for domain experts.

REFERENCES

[1] M. J. Adams, Y. J. Tenney, and R. W. Pew, "Situation awareness and the cognitive management of complex systems," *Human Factors*, vol. 37, no. 1, pp. 85–104, Mar. 1995. doi: 10.1518/001872095779049462

[2] M. Adcock, D. Feng, and B. Thomas, "Visualization of off-surface 3D viewpoint locations in spatial augmented reality," in *Proc. SUI*. New York: ACM, 2013, pp. 1–8. doi: 10.1145/2491367.2491378

[3] M. Agrawala, A. C. Beers, I. McDowall, B. Fröhlich, M. Bolas, and P. Hanrahan, "The two-user responsive workbench: Support for collaboration through individual views of a shared space," in *Proc. SIGGRAPH*. New York: ACM, 1997, pp. 327—332. doi: 10.1145/258734.258875

[4] M. Akçayır and G. Akçayır, "Advantages and challenges associated with augmented reality for education: A systematic review of the literature," *Educational Research Review*, vol. 20, pp. 1–11, Feb. 2017. doi: 10.1016/j.edurev.2016.11.002

[5] P.-A. Albinsson and S. Zhai, "High precision touch screen interaction," in *Proc. CHI*. New York: ACM, 2003, pp. 105–112. doi: 10.1145/642611.642631

[6] L. Alem and W. Huang, *Recent trends of mobile collaborative augmented reality systems*. New York: Springer, 2011. doi: 10.1007/978-1-4419-9845-3

[7] S. Attfield, G. Kazai, M. Lalmas, and B. Piwowarski, "Towards a science of user engagement (position paper)," in *WSDM Workshop on User Modeling for Web Applications*, 2011.

[8] R. Azuma, Y. Baillot, R. Behringer, S. Feiner, S. Julier, and B. MacIntyre, "Recent advances in augmented reality," *IEEE Computer Graphics and Applications*, vol. 21, no. 6, pp. 34–47, Nov. 2001. doi: 10.1109/38.963459

[9] R. T. Azuma, "A survey of augmented reality," *Presence: Teleoperators and Virtual Environments*, vol. 6, no. 4, pp. 355–385, Aug. 1997. doi: 10.1162/pres.1997.6.4.355

[10] B. Bach, R. Sicat, J. Beyer, M. Cordeil, and H. Pfister, "The hologram in my hand: How effective is interactive exploration of 3D visualizations in immersive tangible augmented reality?" *IEEE Transactions on Visualization and Computer Graphics*, vol. 24, no. 1, pp. 457–467, Jan. 2018. doi: 10.1109/TVCG.2017.2745941

[11] R. M. Baecker, *Readings in groupware and computer-supported cooperative work*. San Francisco: Morgan Kaufmann Publishers, 1993.

[12] R. Ballagas, T. E. Dugan, G. Revelle, K. Mori, M. Sandberg, J. Go, E. Reardon, and M. Spasojevic, "Electric agents: Fostering sibling joint media engagement through interactive television and augmented reality," in *Proc. CSCW*. New York: ACM, 2013, pp. 225–236. doi: 10.1145/2441776.2441803

[13] J. Barbier, P. Kenny, J. Young, J.-M. Normand, M. T. Keane, M. O'Sullivan, and A. Ventresque, "MAAP annotate: When archaeology meets augmented reality for annotation of megalithic art," in *Proc. VSMM*. Los Alamitos: IEEE Computer Society, 2017, pp. 283–291. doi: 10.1109/VSMM.2017.8346282

[14] E. Barsom, M. Graafland, and M. Schijven, "Systematic review on the effectiveness of augmented reality applications in medical training," *Surgical endoscopy*, vol. 30, no. 10, pp. 4174–4183, Oct. 2016. doi: 10.1007/s00464-016-4800-6

[15] D. Belcher, M. Billinghurst, S. E. Hayes, and R. Stiles, "Using augmented reality for visualizing complex graphs in three dimensions," in *Proc. ISMAR*. Los Alamitos: IEEE Computer Society, 2003, pp. 84–93. doi: 10.1109/ISMAR.2003.1240691

[16] S. Benford, C. Greenhalgh, G. Reynard, C. Brown, and B. Koleva, "Understanding and constructing shared spaces with mixed-reality boundaries," *ACM Transactions on Computer-Human Interaction*, vol. 5, no. 3, pp. 185–223, Sep. 1998. doi: 10.1145/292834.292836

[17] H. Benko, E. W. Ishak, and S. Feiner, "Collaborative mixed reality visualization of an archaeological excavation," in *Proc. ISMAR*. Los Alamitos: IEEE Computer Society, 2004, pp. 132–140. doi: 10.1109/ISMAR.2004.23

[18] H. Benko, A. D. Wilson, and F. Zannier, "Dyadic projected spatial augmented reality," in *Proc. UIST*. New York: ACM, 2014, pp. 645–655. doi: 10.1145/2642918.2647402

[19] H. Benko, A. Wilson, and R. Jota, "MirageTable: Freehand interaction on a projected augmented reality tabletop," in *Proc. CHI*. New York: ACM, 2012, pp. 199–208. doi: 10.1145/2207676.2207704

[20] L. Besancon, P. Issartel, M. Ammi, and T. Isenberg, "Hybrid tactile/tangible interaction for 3D data exploration," *IEEE Transactions on Visualization and Computer Graphics*, vol. 23, no. 1, pp. 881–890, Jan. 2017. doi: 10.1109/TVCG.2016.2599217

[21] L. Besançon, P. Issartel, M. Ammi, and T. Isenberg, "Mouse, tactile, and tangible input for 3D manipulation," in *Proc. CHI*. New York: ACM, 2017, pp. 4727–4740. doi: 10.1145/3025453.3025863

[22] L. Besançon, M. Sereno, M. Ammi, L. Yu, and T. Isenberg, "Hybrid touch/tangible spatial 3D data selection," *Computer Graphics Forum*, vol. 38, no. 3, pp. 553–567, Jun. 2019. doi: 10.1111/cgf.13710

[23] P. Bhattacharyya, R. Nath, Y. Jo, K. Jadhav, and J. Hammer, "Brick: Toward a model for designing synchronous colocated augmented reality games," in *Proc. CHI*. New York: ACM, 2019, pp. 323:1–323:9. doi: 10.1145/3290605.3300553

[24] M. Billinghurst, J. Bowskill, and J. Morphett, "WearCom: A wearable communication space," in *Proc. CVE*, vol. 98, 1998.

[25] M. Billinghurst, M. Cordeil, A. Bezerianos, and T. Margolis, "Collaborative immersive analytics," in *Immersive Analytics*. Springer, 2018, pp. 221–257. doi: 10.1007/978-3-030-01388-2_8

[26] M. Billinghurst and H. Kato, "Collaborative mixed reality," in *Proceedings of the First International Symposium on Mixed Reality*, 1999, pp. 261–284.

[27] M. Billinghurst, H. Kato, and I. Poupyrev, "The MagicBook: A transitional AR interface," *Computers & Graphics*, vol. 25, no. 5, pp. 745–753, Oct. 2001. doi: 10.1016/S0097-8493(01)00117-0

[28] M. Billinghurst and B. H. Thomas, "Mobile collaborative augmented reality," in *Recent trends of mobile collaborative augmented reality systems*. New York: Springer, 2011, pp. 1–19. doi: 10.1007/978-1-4419-9845-3_1

[29] O. Bimber and R. Raskar, *Spatial Augmented Reality: Merging Real and Virtual Worlds*. Massachusetts, USA: A. K. Peters, Ltd., 2005.

[30] F. Biocca, C. Harms, and J. K. Burgoon, "Toward a more robust theory and measure of social presence: Review and suggested criteria," *Presence: Teleoperators & Virtual Environments*, vol. 12, no. 5, pp. 456–480, Oct. 2003. doi: 10.1162/105474603322761270

[31] J. Blascovich, J. Loomis, A. C. Beall, K. R. Swinth, C. L. Hoyt, and J. N. Bailenson, "Immersive virtual environment technology as a methodological tool for social psychology," *Psychological Inquiry*, vol. 13, no. 2, pp. 103–124, 2002. doi: 10.1207/S15327965PLI1302_01

[32] R. A. Bolt, ""put-that-there": Voice and gesture at the graphics interface," *ACM SIGGRAPH Computer Graphics*, vol. 14, no. 3, pp. 262–270, Jul. 1980. doi: 10.1145/965105.807503

[33] D. A. Bowman and R. P. McMahan, "Virtual reality: How much immersion is enough?" *Computer*, vol. 40, no. 7, pp. 36–43, Jul. 2007. doi: 10.1109/MC.2007.257

[34] S. Bruckner, T. Isenberg, T. Ropinski, and A. Wiebel, "A model of spatial directness in interactive visualization," *IEEE Transactions on Visualization and Computer Graphics*, vol. 25, no. 8, pp. 2514–2528, Aug. 2019. doi: 10.1109/TVCG.2018.2848906

[35] W. Büschel, A. Mitschick, T. Meyer, and R. Dachselt, "Investigating smartphone-based pan and zoom in 3D data spaces in augmented reality," in *Proc. MobileHCI*. New York: ACM, 2019, pp. 2:1–2:13. doi: 10.1145/3338286.3340113

[36] A. Butz, T. Höllerer, S. Feiner, B. MacIntyre, and C. Beshers, "Enveloping users and computers in a collaborative 3D augmented reality," in *Proc. IWAR*. Los Alamitos: IEEE Computer Society, 1999, pp. 35–44. doi: 10.1109/IWAR.1999.803804

[37] T. P. Caudell and D. W. Mizell, "Augmented reality: an application of heads-up display technology to manual manufacturing processes," in *Proc. HICSS*. Los Alamitos: IEEE Computer Society, 1992, pp. 659–669. doi: 10.1109/HICSS.1992.183317

[38] L.-W. Chan, H.-S. Kao, M. Y. Chen, M.-S. Lee, J. Hsu, and Y.-P. Hung, "Touching the void: Direct-touch interaction for intangible displays," in *Proc. CHI*. New York: ACM, 2010, pp. 2625–2634. doi: 10.1145/1753326.1753725

[39] T. Chandler, M. Cordeil, T. Czauderna, T. Dwyer, J. Glowacki, C. Goncu, M. Klapperstueck, K. Klein, K. Marriott, F. Schreiber, and E. Wilson, "Immersive analytics," in *Proc. BDVA*. Los Alamitos: IEEE Computer Society, 2015, pp. 73–80. doi: 10.1109/BDVA.2015.7314296

[40] M. Chen, S. J. Mountford, and A. Sellen, "A study in interactive 3-D rotation using 2-D control devices," *ACM SIGGRAPH Computer Graphics*, vol. 22, no. 4, pp. 121–129, Aug. 1988. doi: 10.1145/378456.378497

[41] D. Clergeaud, J. S. Roo, M. Hachet, and P. Guitton, "Towards seamless interaction between physical and virtual locations for asymmetric collaboration," in *Proc. VRST*. New York: ACM, 2017, pp. 17:1–17:4. doi: 10.1145/3139131.3139165

[42] M. Csikszentmihalyi and I. Csikszentmihalyi, *Beyond boredom and anxiety*. San Francisco: Jossey-Bass, 1975, vol. 721.

[43] J. J. Cummings and J. N. Bailenson, "How immersive is enough? a meta-analysis of the effect of immersive technology on user presence," *Media Psychology*, vol. 19, no. 2, pp. 272–309, 2016. doi: 10.1080/15213269.2015.1015740

[44] J. E. Cutting and P. M. Vishton, "Perceiving layout and knowing distances: The integration, relative potency, and contextual use of different information about depth," in *Perception of space and motion*. Elsevier, 1995, pp. 69–117. doi: 10.1016/B978-012240530-3/50005-5

[45] N. J. Dedual, O. Oda, and S. K. Feiner, "Creating hybrid user interfaces with a 2D multi-touch tabletop and a 3D see-through head-worn display," in *Proc. ISMAR*. Los Alamitos: IEEE Computer Society, 2011, pp. 231–232. doi: 10.1109/ISMAR.2011.6092391

[46] A. DeVincenzi, L. Yao, H. Ishii, and R. Raskar, "Kinected conference: Augmenting video imaging with calibrated depth and audio," in *Proc. CSCW*. New York: ACM, 2011, pp. 621–624. doi: 10.1145/1958824.1958929

[47] M. Dolata, D. Agotai, S. Schubiger, and G. Schwabe, "Pen-and-paper rituals in service interaction: Combining high-touch and high-tech in financial advisory encounters," *ACM Transactions on Computer-Human Interaction*, vol. 3, no. CSCW, pp. 224:1–224:24, Nov. 2019. doi: 10.1145/3359326

[48] S. Dong, A. H. Behzadan, F. Chen, and V. R. Kamat, "Collaborative visualization of engineering processes using tabletop augmented reality," *Advances in Engineering Software*, vol. 55, pp. 45–55, Jan. 2013. doi: 10.1016/j.advengsoft.2012.09.001

[49] M. Dunleavy, C. Dede, and R. Mitchell, "Affordances and limitations of immersive participatory augmented reality simulations for teaching and learning," *Journal of Science Education and Technology*, vol. 18, no. 1, pp. 7–22, Feb. 2009. doi: 10.1007/s10956-008-9119-1

[50] T. Dwyer, K. Marriott, T. Isenberg, K. Klein, N. Riche, F. Schreiber, W. Stuerzlinger, and B. Thomas, "Immersive analytics: An introduction," in *Immersive Analytics*. Springer, 2018, pp. 1–23. doi: 10.1007/978-3-030-01388-2_1

[51] S. Eivazi, H. Afkari, R. Bednarik, V. Leinonen, M. Tukiainen, and J. E. Jääskeläinen, "Analysis of disruptive events and precarious situations caused by interaction with neurosurgical microscope," *Acta Neurochirurgica*, vol. 157, no. 7, pp. 1147–1154, Jul. 2015. doi: 10.1007/s00701-015-2433-5

[52] N. ElSayed, B. Thomas, K. Marriott, J. Piantadosi, and R. Smith, "Situated analytics," in *Proc. BDVA*. Los Alamitos: IEEE Computer Society, 2015, pp. 96–103. doi: 10.1109/BDVA.2015.7314302

[53] N. A. ElSayed, B. H. Thomas, K. Marriott, J. Piantadosi, and R. T. Smith, "Situated analytics: Demonstrating immersive analytical tools with augmented reality," *Journal of Visual Languages & Computing*, vol. 36, pp. 13–23, Oct. 2016. doi: 10.1016/j.jvlc.2016.07.006

[54] C. Elvezio, M. Sukan, O. Oda, S. Feiner, and B. Tversky, "Remote collaboration in AR and VR using virtual replicas," in *Proc. SIGGRAPH*. New York: ACM, 2017, pp. 13:1–13:2. doi: 10.1145/3089269.3089281

[55] B. Ens, J. D. Hincapié-Ramos, and P. Irani, "Ethereal planes: A design framework for 2D information space in 3D mixed reality environments," in *Proc. SUI*. New York: ACM, 2014, pp. 2–12. doi: 10.1145/2659766.2659769

[56] B. Ens, J. Lanir, A. Tang, S. Bateman, G. Lee, T. Piumsomboon, and M. Billinghurst, "Revisiting collaboration through mixed reality: The evolution of groupware," *International Journal of Human-Computer Studies*, vol. 131, pp. 81–98, Nov. 2019. doi: 10.1016/j.ijhcs.2019.05.011

[57] S. Feiner, B. MacIntyre, M. Haupt, and E. Solomon, "Windows on the world: 2D windows for 3D augmented reality," in *Proc. UIST*. New York: ACM, 1993, pp. 145–155. doi: 10.1145/168642.168657

[58] G. W. Fitzmaurice, "Graspable user interfaces," Ph.D. dissertation, University of Toronto, Canada, 1996.

[59] G. W. Fitzmaurice, H. Ishii, and W. A. S. Buxton, "Bricks: Laying the foundations for graspable user interfaces," in *Proc. CHI*. New York: ACM, 1995, pp. 442–449. doi: 10.1145/223904.223964

[60] A. Fonnet and Y. Prié, "Survey of immersive analytics," *IEEE Transactions on Visualization and Computer Graphics*, 2020, to appear. doi: 10.1109/TVCG.2019.2929033

[61] G. Fontaine, "The experience of a sense of presence in intercultural and international encounters," *Presence: Teleoperators and Virtual Environments*, vol. 1, no. 4, pp. 482–490, Fall 1992. doi: 10.1162/pres.1992.1.4.482

[62] C. Forlines, D. Wigdor, C. Shen, and R. Balakrishnan, "Direct-touch vs. mouse input for tabletop displays," in *Proc. CHI*. New York: ACM, 2007, pp. 647–656. doi: 10.1145/1240624.1240726

[63] A. Forsberg, J. Chen, and D. Laidlaw, "Comparing 3D vector field visualization methods: A user study," *IEEE Transactions on Visualization and Computer Graphics*, vol. 15, no. 6, pp. 1219–1226, Nov. 2009. doi: 10.1109/TVCG.2009.126

[64] J. Franz, M. Alnusayri, J. Malloch, and D. Reilly, "A comparative evaluation of techniques for sharing AR experiences in museums," *ACM Transactions on Computer-Human Interaction*, vol. 3, no. CSCW, Nov. 2019. doi: 10.1145/3359226

[65] S. Gauglitz, C. Lee, M. Turk, and T. Höllerer, "Integrating the physical environment into mobile remote collaboration," in *Proc. MobileHCI*. New York: ACM, 2012, pp. 241–250. doi: 10.1145/2371574.2371610

[66] 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 *Proc. VRST*. New York: ACM, 2014, pp. 197–205. doi: 10.1145/2671015.2671016

[67] S. Gauglitz, B. Nuernberger, M. Turk, and T. Höllerer, "World-stabilized annotations and virtual scene navigation for remote collaboration," in *Proc. UIST*. New York: ACM, 2014, pp. 449–459. doi: 10.1145/2642918.2647372

[68] 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*, vol. 22, no. 11, pp. 2413–2422, Nov. 2016. doi: 10.1109/TVCG.2016.2593778

[69] C. Gutwin and S. Greenberg, "A descriptive framework of workspace awareness for real-time groupware," *Computer Supported Cooperative Work*, vol. 11, no. 3, pp. 411–446, Sep. 2002. doi: 10.1023/A:1021271517844

[70] C. Heeter, "Being there: The subjective experience of presence," *Presence: Teleoperators and Virtual Environments*, vol. 1, no. 2, pp. 262–271, Spring 1992. doi: 10.1162/pres.1992.1.2.262

[71] S. J. Henderson and S. Feiner, "Opportunistic controls: Leveraging natural affordances as tangible user interfaces for augmented reality," in *Proc. VRST*. New York: ACM, 2008, pp. 211–218. doi: 10.1145/1450579.1450625

[72] S. J. Henderson and S. K. Feiner, "Augmented reality in the psychomotor phase of a procedural task," in *Proc. ISMAR*. Los Alamitos: IEEE Computer Society, 2011, pp. 191–200. doi: 10.1109/ISMAR.2011.6092386

[73] J. D. Hincapié-Ramos, X. Guo, P. Moghadasian, and P. Irani, "Consumed endurance: A metric to quantify arm fatigue of mid-air interactions," in *Proc. CHI*. New York: ACM, 2014, pp. 1063–1072. doi: 10.1145/2556288.2557130

[74] K. Hinckley, J. Tullio, R. Pausch, D. Proffitt, and N. Kassell, "Usability analysis of 3D rotation techniques," in *Proc. UIST*. New York: ACM, 1997, pp. 1–10. doi: 10.1145/263407.263408

[75] W. Huang, M. Billinghurst, L. Alem, and S. Kim, "HandsInTouch: Sharing gestures in remote collaboration," in *Proc. OzCHI*. New York: ACM, 2018, pp. 396–400. doi: 10.1145/3292147.3292177

[76] C. Hurter, N. H. Riche, S. M. Drucker, M. Cordeil, R. Alligier, and R. Vuillemot, "FiberClay: Sculpting three dimensional trajectories to reveal structural insights," *IEEE Transactions on Visualization*

and Computer Graphics, vol. 25, no. 1, pp. 704–714, Jan 2019. doi: 10 .1109/TVCG.2018.2865191

[77] D.-N. T. Huynh, K. Raveendran, Y. Xu, K. Spreen, and B. MacIntyre, "Art of defense: A collaborative handheld augmented reality board game," in Proc. SIGGRAPH. New York: ACM, 2009, pp. 135–142. doi: 10.1145/1581073.1581095

[78] A. Irlitti, R. T. Smith, S. Von Itzstein, M. Billinghurst, and B. H. Thomas, "Challenges for asynchronous collaboration in augmented reality," in Proc. ISMAR-Adjunct. Los Alamitos: IEEE CS, 2016, pp. 31–35. doi: 10.1109/ISMAR-Adjunct.2016.0032

[79] A. Irlitti, S. Von Itzstein, R. T. Smith, and B. H. Thomas, "Performance improvement using data tags for handheld spatial augmented reality," in Proc. VRST. New York: ACM, 2014, p. 161–165. doi: 10.1145/2671015.2671116

[80] P. Isenberg, N. Elmqvist, J. Scholtz, D. Cernea, K.-L. Ma, and H. Hagen, "Collaborative visualization: Definition, challenges, and research agenda," Information Visualization, vol. 10, no. 4, pp. 310–326, Oct. 2011. doi: 10.1177/1473871611412817

[81] T. Isenberg and M. Hancock, "Gestures vs. postures: 'Gestural' touch interaction in 3D environments," in Proc. 3DCHI, 2012, pp. 53–61.

[82] H. Ishii and B. Ullmer, "Tangible bits: Towards seamless interfaces between people, bits and atoms," in Proc. CHI. New York: ACM, 1997, pp. 234–241. doi: 10.1145/258549.258715

[83] S. Jang, W. Stuerzlinger, S. Ambike, and K. Ramani, "Modeling cumulative arm fatigue in mid-air interaction based on perceived exertion and kinetics of arm motion," in Proc. CHI. New York: ACM, 2017, pp. 3328–3339. doi: 10.1145/3025453.3025523

[84] P. Kán, A. Dünser, M. Billinghurst, C. Schönauer, and H. Kaufmann, "The effects of direct and global illumination on presence in augmented reality," in Proc. ISPR. Vienna: Facultas Verlags- und Buchhandels AG, 2014, pp. 223–230.

[85] S. Kasahara, V. Heun, A. S. Lee, and H. Ishii, "Second surface: Multi-user spatial collaboration system based on augmented reality," in Proc. SIGGRAPH. New York: ACM, 2012, p. 1–4. doi: 10.1145/2407707.2407727

[86] D. F. Keefe and T. Isenberg, "Reimagining the scientific visualization interaction paradigm," Computer, vol. 46, no. 5, pp. 51–57, May 2013. doi: 10.1109/MC.2013.178

[87] 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, vol. 24, no. 11, pp. 2947–2962, Nov. 2018. doi: 10.1109/TVCG.2018.2868591

[88] S. Kim, G. Lee, N. Sakata, and M. Billinghurst, "Improving co-presence with augmented visual communication cues for sharing experience through video conference," in Proc. ISMAR. Los Alamitos: IEEE Computer Society, 2014, pp. 83–92. doi: 10.1109/ISMAR.2014.6948412

[89] C. Kirner and T. G. Kirner, "A data visualization virtual environment supported by augmented reality," in Proc. SMC. Los Alamitos: IEEE Computer Society, 2006, pp. 97–102. doi: 10.1109/ICSMC.2006.384365

[90] K. Kiyokawa, H. Takemura, and N. Yokoya, "A collaboration support technique by integrating a shared virtual reality and a shared augmented reality," in Proc. SMC. Los Alamitos: IEEE Computer Society, 1999, pp. 48–53. doi: 10.1109/ICSMC.1999.816444

[91] R. Komiyama, T. Miyaki, and J. Rekimoto, "JackIn space: Designing a seamless transition between first and third person view for effective telepresence collaborations," in Proc. AH. New York: ACM, 2017, pp. 14:1–14:9. doi: 10.1145/3041164.3041183

[92] O.-H. Kwon, C. Muelder, K. Lee, and K.-L. Ma, "A study of layout, rendering, and interaction methods for immersive graph visualization," IEEE Transactions on Visualization and Computer Graphics, vol. 22, no. 7, pp. 1802–1815, Jul. 2016. doi: 10.1109/TVCG.2016.2520921

[93] W. S. Lages and D. A. Bowman, "Walking with adaptive augmented reality workspaces: Design and usage patterns," in Proc. IUI. ACM, 2019, pp. 356–366. doi: 10.1145/3301275.3302278

[94] J. J. LaViola, Jr., "A discussion of cybersickness in virtual environments," SIGCHI Bulletin, vol. 32, no. 1, pp. 47–56, Jan. 2000. doi: 10.1145/333329.333344

[95] M. Le Chénéchal, T. Duval, V. Gouranton, J. Royan, and B. Arnaldi, "The stretchable arms for collaborative remote guiding," in Proc. ICAT – EGVE. Goslar, Germany: Eurographics Association, 2015, pp. 147–150. doi: 10.2312/egve.20151322

[96] G. A. Lee, T. Teo, S. Kim, and M. Billinghurst, "Mixed reality collaboration through sharing a live panorama," in Proc. SA. New York: ACM, 2017, pp. 14:1–14:4. doi: 10.1145/3132787.3139203

[97] G. A. Lee, T. Teo, S. Kim, and M. Billinghurst, "A user study on MR remote collaboration using live 360 video," in Proc. ISMAR. Los Alamitos: IEEE Computer Society, 2018, pp. 153–164. doi: 10.1109/ISMAR.2018.00051

[98] K. M. Lee, "Presence, explicated," Communication Theory, vol. 14, no. 1, pp. 27–50, 2004. doi: 10.1111/j.1468-2885.2004.tb00302.x

[99] S. Lee and S. Zhai, "The performance of touch screen soft buttons," in Proc. CHI. New York: ACM, 2009, pp. 309–318. doi: 10.1145/1518701.1518750

[100] N. H. Lehment, D. Merget, and G. Rigoll, "Creating automatically aligned consensus realities for AR videoconferencing," in Proc. ISMAR. Los Alamitos: IEEE Computer Society, 2014, pp. 201–206. doi: 10.1109/ISMAR.2014.6948428

[101] K.-C. Lien, B. Nuernberger, T. Höllerer, and M. Turk, "PPV: Pixel-point-volume segmentation for object referencing in collaborative augmented reality," in Proc. ISMAR. IEEE, 2016, pp. 77–83. doi: 10.1109/ISMAR.2016.21

[102] P. Lincoln, G. Welch, A. Nashel, A. Ilie, A. State, and H. Fuchs, "Animatronic shader lamps avatars," in Proc. ISMAR. Los Alamitos: IEEE Computer Society, 2009, pp. 27–33. doi: 10.1109/ISMAR.2009.5336503

[103] S. Liu, D. Cheng, and H. Hua, "An optical see-through head mounted display with addressable focal planes," in Proc. ISMAR. Los Alamitos: IEEE Computer Society, 2008, pp. 33–42. doi: 10.1109/ISMAR.2008.4637321

[104] D. López, L. Oehlberg, C. Doger, and T. Isenberg, "Towards an understanding of mobile touch navigation in a stereoscopic viewing environment for 3D data exploration," IEEE Transactions on Visualization and Computer Graphics, vol. 22, no. 5, pp. 1616–1629, May 2016. doi: 10.1109/TVCG.2015.2440233

[105] S. Lukosch, M. Billinghurst, L. Alem, and K. Kiyokawa, "Collaboration in augmented reality," Computer Supported Cooperative Work, vol. 24, no. 6, pp. 515–525, 2015. doi: 10.1007/s10606-015-9239-0

[106] A. MacWilliams, C. Sandor, M. Wagner, M. Bauer, G. Klinker, and B. Bruegge, "Herding sheep: Live system development for distributed augmented reality," in Proc. ISMAR. Los Alamitos: IEEE Computer Society, 2003, pp. 123–132. doi: 10.1109/ISMAR.2003.1240695

[107] T. Mahmood, W. Fulmer, N. Mungoli, J. Huang, and A. Lu, "Improving information sharing and collaborative analysis for remote GeoSpatial visualization using mixed reality," in Proc. ISMAR. Los Alamitos: IEEE Computer Society, 2019, pp. 236–247. doi: 10.1109/ISMAR.2019.00021

[108] A. Maimone and H. Fuchs, "Encumbrance-free telepresence system with real-time 3D capture and display using commodity depth cameras," in Proc. ISMAR. Los Alamitos: IEEE Computer Society, 2011, pp. 137–146. doi: 10.1109/ISMAR.2011.6092379

[109] G. E. Marai, A. G. Forbes, and A. Johnson, "Interdisciplinary immersive analytics at the electronic visualization laboratory: Lessons learned and upcoming challenges," in Workshop on Immersive Analytics. Los Alamitos: IEEE Computer Society, 2016, pp. 54–59. doi: 10.1109/IMMERSIVE.2016.7932384

[110] K. Marriott, F. Schreiber, T. Dwyer, N. H. R. Karsten Klein, T. Itoh, W. Stuerzlinger, and B. H. Thomas, Immersive Analytics. Springer, 2018. doi: 10.1007/978-3-030-01388-2

[111] P. Marshall, R. Fleck, A. Harris, J. Rick, E. Hornecker, Y. Rogers, N. Yuill, and N. S. Dalton, "Fighting for control: Children's embodied interactions when using physical and digital representations," in Proc. CHI. New York: ACM, 2009, pp. 2149–2152. doi: 10.1145/1518701.1519027

[112] S. Matsutomo, T. Miyauchi, S. Noguchi, and H. Yamashita, "Real-time visualization system of magnetic field utilizing augmented reality technology for education," IEEE Transactions on Magnetics, vol. 48, no. 2, pp. 531–534, Feb. 2012. doi: 10.1109/TMAG.2011.2174208

[113] J. P. McIntire, P. R. Havig, and E. E. Geiselman, "Stereoscopic 3D displays and human performance: A comprehensive review," Displays, vol. 35, no. 1, pp. 18–26, Jan. 2014. doi: 10.1016/j.displa.2013.10.004

[114] P. Milgram and F. Kishino, "A taxonomy of mixed reality visual displays," IEICE Transactions on Information and Systems, vol. 77, no. 12, pp. 1321–1329, Dec. 1994.

[115] S. Mora, A. Boron, and M. Divitini, "CroMAR: Mobile augmented reality for supporting reflection on crowd management," Interna-

tional Journal of Mobile Human Computer Interaction, vol. 4, no. 2, pp. 88–101, Apr. 2012. doi: 10.4018/jmhci.2012040107

[116] S. Mori, S. Ikeda, and H. Saito, "A survey of diminished reality: Techniques for visually concealing, eliminating, and seeing through real objects," *IPSJ Transactions on Computer Vision and Applications*, vol. 9, no. 1, pp. 17:1–17:14, Jun. 2017. doi: 10.1186/s41074-017-0028-1

[117] J. Müller, R. Rädle, and H. Reiterer, "Remote collaboration with mixed reality displays: How shared virtual landmarks facilitate spatial referencing," in *Proc. CHI*. New York: ACM, 2017, pp. 6481–6486. doi: 10.1145/3025453.3025717

[118] S. Nilsson, B. Johansson, and A. Jonsson, "Using AR to support cross-organisational collaboration in dynamic tasks," in *Proc. ISMAR*. Los Alamitos: IEEE Computer Society, 2009, pp. 3–12. doi: 10.1109/ISMAR.2009.5336522

[119] H. L. O'Brien and E. G. Toms, "What is user engagement? a conceptual framework for defining user engagement with technology," *Journal of the American Society for Information Science and Technology*, vol. 59, no. 6, pp. 938–955, Apr. 2008. doi: 10.1002/asi.20801

[120] O. Oda and S. Feiner, "3D referencing techniques for physical objects in shared augmented reality," in *Proc. ISMAR*. Los Alamitos: IEEE Computer Society, 2012, pp. 207–215. doi: 10.1109/ISMAR.2012.6402558

[121] O. Oda, C. Elvezio, M. Sukan, S. Feiner, and B. Tversky, "Virtual replicas for remote assistance in virtual and augmented reality," in *Proc. UIST*. New York: ACM, 2015, pp. 405–415. doi: 10.1145/2807442.2807497

[122] O. Oda and S. Feiner, "Interference avoidance in multi-user handheld augmented reality," in *Proc. ISMAR*. Los Alamitos: IEEE Computer Society, 2009, pp. 13–22. doi: 10.1109/ISMAR.2009.5336507

[123] K. O'Hara, G. Gonzalez, A. Sellen, G. Penney, A. Varnavas, H. Mentis, A. Criminisi, R. Corish, M. Rouncefield, N. Dastur, and T. Carrell, "Touchless interaction in surgery," *Communications of the ACM*, vol. 57, no. 1, pp. 70–77, Jan. 2014. doi: 10.1145/2541883.2541899

[124] I. C. Olson, Z. Atrash Leong, U. Wilensky, and M. S. Horn, ""It's just a toolbar!" Using tangibles to help children manage conflict around a multi-touch tabletop," in *Proc. TEI*. New York: ACM, 2011, pp. 29–36. doi: 10.1145/1935701.1935709

[125] S. K. Ong, M. L. Yuan, and A. Y. C. Nee, "Augmented reality applications in manufacturing: a survey," *International Journal of Production Research*, vol. 46, no. 10, pp. 2707–2742, Mar. 2008. doi: 10.1080/00207540601064773

[126] S. Ong and Y. Shen, "A mixed reality environment for collaborative product design and development," *CIRP Annals*, vol. 58, no. 1, pp. 139–142, 2009. doi: 10.1016/j.cirp.2009.03.020

[127] J. Ou, S. R. Fussell, X. Chen, L. D. Setlock, and J. Yang, "Gestural communication over video stream: Supporting multimodal interaction for remote collaborative physical tasks," in *Proc. ICMI*. New York: ACM, 2003, pp. 242–249. doi: 10.1145/958432.958477

[128] K. S. Park, A. Kapoor, and J. Leigh, "Lessons learned from employing multiple perspectives in a collaborative virtual environment for visualizing scientific data," in *Proc. CVE*. New York: ACM, 2000, pp. 73–82. doi: 10.1145/351006.351015

[129] T. Pejsa, J. Kantor, H. Benko, E. Ofek, and A. Wilson, "Room2Room: Enabling life-size telepresence in a projected augmented reality environment," in *Proc. CSCW*. New York: ACM, 2016, pp. 1716–1725. doi: 10.1145/2818048.2819965

[130] D. N. E. Phon, M. B. Ali, and N. D. A. Halim, "Collaborative augmented reality in education: A review," in *Proc. LaTiCE*. Los Alamitos: IEEE Computer Society, 2014, pp. 78–83. doi: 10.1109/LaTiCE.2014.23

[131] W. Piekarski and B. Thomas, "ARQuake: The outdoor augmented reality gaming system," *Communications of the ACM*, vol. 45, no. 1, pp. 36–38, Jan. 2002. doi: 10.1145/502269.502291

[132] J. S. Pierce, B. C. Stearns, and R. Pausch, "Voodoo dolls: Seamless interaction at multiple scales in virtual environments," in *Proc. I3D*. New York: ACM, 1999, pp. 141–145. doi: 10.1145/300523.300540

[133] B. Piper, C. Ratti, and H. Ishii, "Illuminating clay: a 3D tangible interface for landscape analysis," in *Proc. CHI*. New York: ACM, 2002, pp. 355–362. doi: 10.1145/503376.503439

[134] T. Piumsomboon, A. Dey, B. Ens, G. Lee, and M. Billinghurst, "The effects of sharing awareness cues in collaborative mixed reality,"

Frontiers in Robotics and AI, vol. 6, pp. 5:1–5:18, Feb. 2019. doi: 10.3389/frobt.2019.00005

[135] T. Piumsomboon, G. A. Lee, B. Ens, B. H. Thomas, and M. Billinghurst, "Superman vs Giant: A study on spatial perception for a multi-scale mixed reality flying telepresence interface," *IEEE Transactions on Visualization and Computer Graphics*, vol. 24, no. 11, pp. 2974–2982, Nov. 2018. doi: 10.1109/TVCG.2018.2868594

[136] 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 *Proc. CHI*. New York: ACM, 2018, pp. 46:1–46:13. doi: 10.1145/3173574.3173620

[137] 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 *Proc. CHI*. New York: ACM, 2019, pp. 228:1–228:17. doi: 10.1145/3290605.3300458

[138] T. Piumsomboon, Y. Lee, G. Lee, and M. Billinghurst, "CoVAR: A collaborative virtual and augmented reality system for remote collaboration," in *SIGGRAPH Asia Emerging Technologies*. New York: ACM, 2017, pp. 3:1–3:2. doi: 10.1145/3132818.3132822

[139] R. Poelman, O. Akman, S. Lukosch, and P. Jonker, "As if being there: Mediated reality for crime scene investigation," in *Proc. CSCW*. New York: ACM, 2012, pp. 1267–1276. doi: 10.1145/2145204.2145394

[140] L. Poretski, J. Lanir, and O. Arazy, "Normative tensions in shared augmented reality," *Proceedings of the ACM on Human-Computer Interaction*, vol. 2, no. CSCW, pp. 142:1–142:22, Nov. 2018. doi: 10.1145/3274411

[141] I. Poupyrev, M. Billinghurst, S. Weghorst, and T. Ichikawa, "The go-go interaction technique: Non-linear mapping for direct manipulation in VR," in *Proc. UIST*. New York: ACM, 1996, pp. 79–80. doi: 10.1145/237091.237102

[142] E. Prytz, S. Nilsson, and A. Jönsson, "The importance of eye-contact for collaboration in AR systems," in *Proc. ISMAR*. Los Alamitos: IEEE Computer Society, 2010, pp. 119–126. doi: 10.1109/ISMAR.2010.5643559

[143] R. Raskar, G. Welch, and H. Fuchs, "Spatially augmented reality," in *Proc. IWAR*. Natick, MA, USA: A. K. Peters, Ltd., 1999, pp. 63–72.

[144] J. Rekimoto, "Transvision: A hand-held augmented reality system for collaborative design," in *Proc. VSMM*, 1996, pp. 85–90.

[145] J. S. Roo, R. Gervais, and M. Hachet, "Inner garden: An augmented sandbox designed for self-reflection," in *Proc. TEI*. New York: ACM, 2016, pp. 570–576. doi: 10.1145/2839462.2856532

[146] J. S. Roo and M. Hachet, "One reality: Augmenting how the physical world is experienced by combining multiple mixed reality modalities," in *Proc. UIST*. New York: ACM, 2017, pp. 787–795. doi: 10.1145/3126594.3126638

[147] J. S. Roo and M. Hachet, "Towards a hybrid space combining spatial augmented reality and virtual reality," in *Proc. 3DUI*. Los Alamitos: IEEE Computer Society, 2017, pp. 195–198. doi: 10.1109/3DUI.2017.7893339

[148] L. Sambrooks and B. Wilkinson, "Comparison of gestural, touch, and mouse interaction with fitts' law," in *Proc. OzCHI*. New York: ACM, 2013, pp. 119–122. doi: 10.1145/2541016.2541066

[149] D. Schmalstieg, A. Fuhrmann, G. Hesina, Z. Szalavári, L. M. Encarnação, M. Gervautz, and W. Purgathofer, "The Studierstube augmented reality project," *Presence: Teleoperators and Virtual Environments*, vol. 11, no. 1, pp. 33–54, Feb. 2002. doi: 10.1162/105474602317343640

[150] D. Schmalstieg, A. Fuhrmann, Z. Szalavari, and M. Gervautz, "Studierstube—An environment for collaboration in augmented reality," in *Proc. CVE*, 1996.

[151] M. Sereno, L. Besançon, and T. Isenberg, "Supporting volumetric data visualization and analysis by combining augmented reality visuals with multi-touch input," in *Posters at EuroVis*, 2019, pp. 21–23. doi: 10.2312/eurp.20191136

[152] M. Slater, "Place illusion and plausibility can lead to realistic behaviour in immersive virtual environments," *Philosophical Transactions of the Royal Society B: Biological Sciences*, vol. 364, no. 1535, pp. 3549–3557, 2009. doi: 10.1098/rstb.2009.0138

[153] G. Smith and J. Mariani, "Using subjective views to enhance 3D applications," in *Proc. VRST*. New York: ACM, 1997, p. 139–146. doi: 10.1145/261135.261161

[154] R. S. Sodhi, B. R. Jones, D. Forsyth, B. P. Bailey, and G. Maciocci, "BeThere: 3D mobile collaboration with spatial input," in *Proc.*

CHI.   New York: ACM, 2013, pp. 179–188. doi: 10.1145/2470654.2470679

[155] P. Song, W. B. Goh, C.-W. Fu, Q. Meng, and P.-A. Heng, "WYSI-WYF: Exploring and annotating volume data with a tangible handheld device," in *Proc. CHI*.   New York: ACM, 2011, pp. 1333–1342. doi: 10.1145/1978942.1979140

[156] A. Stafford, W. Piekarski, and B. Thomas, "Implementation of god-like interaction techniques for supporting collaboration between outdoor ar and indoor tabletop users," in *Proc. ISMAR*.   Los Alamitos: IEEE Computer Society, 2006, pp. 165–172. doi: 10.1109/ISMAR.2006.297809

[157] R. Stoakley, M. J. Conway, and R. Pausch, "Virtual reality on a WIM: Interactive worlds in miniature," in *Proc. CHI*.   New York: ACM, 1995, pp. 265–272. doi: 10.1145/223904.223938

[158] N. Sugano, H. Kato, and K. Tachibana, "The effects of shadow representation of virtual objects in augmented reality," in *Proc. ISMAR*.   Los Alamitos: IEEE Computer Society, 2003, pp. 76–83. doi: 10.1109/ISMAR.2003.1240690

[159] H. B. Surale, A. Gupta, M. Hancock, and D. Vogel, "TabletInVR: Exploring the design space for using a multi-touch tablet in virtual reality," in *Proc. CHI*.   New York: ACM, 2019, pp. 13:1–13:13. doi: 10.1145/3290605.3300243

[160] I. E. Sutherland, "A head-mounted three dimensional display," in *Proc. AFIPS*.   New York: ACM, 1968, pp. 757–764. doi: 10.1145/1476589.1476686

[161] Z. Szalavári, E. Eckstein, and M. Gervautz, "Collaborative gaming in augmented reality," in *Proc. VRST*.   New York: ACM, 1998, pp. 195–204. doi: 10.1145/293701.293740

[162] Z. Szalavári, D. Schmalstieg, A. Fuhrmann, and M. Gervautz, ""Studierstube": An environment for collaboration in augmented reality," *Virtual Reality*, vol. 3, no. 1, pp. 37–48, Mar. 1998. doi: 10.1007/BF01409796

[163] T. Teo, L. Lawrence, G. A. Lee, M. Billinghurst, and M. Adcock, "Mixed reality remote collaboration combining 360 video and 3D reconstruction," in *Proc. CHI*.   New York: ACM, 2019, pp. 201:1–201:14. doi: 10.1145/3290605.3300431

[164] 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 *Proc. OzCHI*.   New York: ACM, 2018, pp. 406–410. doi: 10.1145/3292147.3292200

[165] S. Thanyadit, P. Punpongsanon, and T. Pong, "ObserVAR: Visualization system for observing virtual reality users using augmented reality," in *Proc. ISMAR*.   Los Alamitos: IEEE Computer Society, 2019, pp. 258–268. doi: 10.1109/ISMAR.2019.00023

[166] J. J. Thomas and K. A. Cook, *Illuminating the path: The research and development agenda for visual analytics*.   United States. Department of Homeland Security, 2005.

[167] X. Wang and P. S. Dunston, "User perspectives on mixed reality tabletop visualization for face-to-face collaborative design review," *Automation in Construction*, vol. 17, no. 4, pp. 399–412, May 2008. doi: 10.1016/j.autcon.2007.07.002

[168] X. Wang and P. S. Dunston, "Comparative effectiveness of mixed reality-based virtual environments in collaborative design," *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)*, vol. 41, no. 3, pp. 284–296, May 2011. doi: 10.1109/TSMCC.2010.2093573

[169] X. Wang, P. E. Love, M. J. Kim, and W. Wang, "Mutual awareness in collaborative design: An augmented reality integrated telepresence system," *Computers in Industry*, vol. 65, no. 2, pp. 314–324, Feb. 2014. doi: 10.1016/j.compind.2013.11.012

[170] X. Wang, L. Besançon, F. Guéniat, M. Sereno, M. Ammi, and T. Isenberg, "A vision of bringing immersive visualization to scientific workflows," in *Workshop CHI-IA*, Glasgow, UK, 2019.

[171] X. Wang, L. Besançon, D. Rousseau, M. Sereno, M. Ammi, and T. Isenberg, "Towards an understanding of augmented reality extensions for existing 3D data analysis tools," in *Proc. CHI*.   New York: ACM, 2020. doi: 10.1145/3313831.3376657

[172] C. Ware and P. Mitchell, "Visualizing graphs in three dimensions," *ACM Transactions on Applied Perception*, vol. 5, no. 1, pp. 2:1–2:15, Jan. 2008. doi: 10.1145/1279640.1279642

[173] A. D. Wilson and H. Benko, "Combining multiple depth cameras and projectors for interactions on, above and between surfaces," in *Proc. UIST*.   New York: ACM, 2010, pp. 273–282. doi: 10.1145/1866029.1866073

[174] B. G. Witmer and M. J. Singer, "Measuring presence in virtual environments: A presence questionnaire," *Presence*, vol. 7, no. 3, pp. 225–240, Jun. 1998. doi: 10.1162/105474698565686

[175] H.-K. Wu, S. W.-Y. Lee, H.-Y. Chang, and J.-C. Liang, "Current status, opportunities and challenges of augmented reality in education," *Computers & education*, vol. 62, pp. 41–49, Mar. 2013. doi: 10.1016/j.compedu.2012.10.024

[176] H. Xia, S. Herscher, K. Perlin, and D. Wigdor, "Spacetime: Enabling fluid individual and collaborative editing in virtual reality," in *Proc. UIST*.   New York: ACM, 2018, pp. 853–866. doi: 10.1145/3242587.3242597

[177] L. Xie, A. N. Antle, and N. Motamedi, "Are tangibles more fun? comparing children's enjoyment and engagement using physical, graphical and tangible user interfaces," in *Proc. TEI*.   New York: ACM, 2008, pp. 191–198. doi: 10.1145/1347390.1347433

[178] N. Yuill and Y. Rogers, "Mechanisms for collaboration: A design and evaluation framework for multi-user interfaces," *ACM Transactions on Computer-Human Interaction*, vol. 19, no. 1, pp. 1:1–1:25, May 2012. doi: 10.1145/2147783.2147784

[179] 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 *Proc. ISMAR*.   Los Alamitos: IEEE Computer Society, 2008, pp. 193–202. doi: 10.1109/ISMAR.2008.4637362

Mickael Sereno is a PhD student at University Paris-Saclay and Inria, France. He obtained his Master in HCI and engineering in computer science degrees at University Paris-Sud and Polytech Paris-Sud in 2018. His work focuses at understanding the benefits and limitations of AR-HMDs for collaborative scientific visualization scenarios.

Xiyao Wang is a PhD student at University Paris-Saclay and Inria, France. Previously, he received his master degree at University Paris-Sud, engineering degree at CentraleSupelec, bachelor degrees at University Pierre Marie Curie and Ocean University Of China. Currently, he works on studying novel 3D interaction and visualization techniques using AR environments.

Lonni Besançon is a postdoctoral fellow at Linköping University, Sweden. He received his Ph.D in computer science at University Paris Saclay, France. His thesis "An interaction continuum for 3D dataset visualization" received the second price of the prix de these GDR-IGRV. He is particularly interested in interactive visualization techniques for 3D spatial data relying on new input paradigms and his recent work focuses on the visualization and understanding of uncertainty in empirical results in computer science.

Michael J. McGuffin is an associate professor at ETS, a French-language engineering school in Montreal, Canada, where his students do research in AR, VR, HCI, and visualization. In 2009, his paper at the IEEE Information Visualization Conference (InfoVis 2009) received an Honorable Mention.

Tobias Isenberg is a senior research scientist at Inria, France. He received his doctoral degree from the University of Magdeburg, Germany, in 2004. Previously he held positions as post-doctoral fellow at the University of Calgary, Canada, and as assistant professor at the University of Groningen, the Netherlands. His research interests comprise topics in scientific visualization, illustrative and non-photorealistic rendering, and interactive visualization techniques. He is particularly interested in interactive visualization environments for 3D spatial data that rely on novel interaction paradigms such as augmented reality, tactile screens, and tangible devices.