Contents lists available at ScienceDirect

J. Vis. Commun. Image R.

journal homepage: www.elsevier.com/locate/jvci

Full length article

Supporting workspace awareness in remote assistance through a flexible multi-camera system and Augmented Reality awareness cues[☆]

Troels Rasmussen[a,*], Tiare Feuchtner[a], Weidong Huang[b], Kaj Grønbæk[a]

[a] Department of Computer Science, Aarhus University, Aarhus, Denmark
[b] University of Technology Sydney, Sydney, Australia

ARTICLE INFO

MSC:
68U35

Keywords:
Remote assistance
Augmented reality
Workspace awareness
Awareness cues
Calibration

ABSTRACT

Workspace awareness is critical for remote assistance with physical tasks, yet it remains difficult to facilitate. For example, if the remote helper is limited to the single viewpoint provided by the worker's hand-held or head-mounted camera, she lacks the ability to gain an overview of the workspace. This may be addressed by granting the helper view-independence, e.g., through a multi-camera system. However, it can be cumbersome to set up and calibrate multiple cameras, and it can be challenging for the local worker to identify the current viewpoint of the remote helper. We present *CueCam*, a multi-camera remote assistance system that supports mutual workspace awareness through a flexible ad-hoc camera calibration and various Augmented Reality cues that communicate the helper's viewpoint and focus. In particular, we propose visual cues presented through a head-mounted Augmented Reality display (Virtual Hand, Color Cue), and sound cues emitted from the cameras' physical locations (Spatial Sound). Findings from a lab study indicate that all proposed cues effectively support the worker's awareness of helper's location and focus, while the Color Cue demonstrated superiority in task performance and preference ratings during a search task.

1. Introduction

Remote assistance is the interactive process by which a remote user, typically called the "helper", assists a local user, often referred to as the "worker", in performing a physical task. In this paper, we will refer to the helper as female ("she") and the worker as male ("he") solely for the benefit of readability and easier understanding. In an industrial scenario, the worker may use a standard video communication application, like Skype, to convey information about a machine problem to the helper. In this case, information about the workspace is provided from the worker's point of view, and the helper is limited to this viewpoint. In other words, the helper may not be aware of the overall spatial layout of the workspace and cannot, for example, independently move to inspect another area of a large industrial machine. In contrast to the conventional phone-based remote assistance, *multi-camera remote assistance* involves the use of two or more cameras that are mounted in the environment to capture different areas and perspectives of a workspace. This offers a degree of view independence, since the helper can quickly "teleport" from one area of a workspace to another by simply looking through a different camera, without needing to negotiate the view with the worker. In previous research, view independence has been shown to be beneficial for remote assistance [1–4], in particular for large workspaces, which may require a worker and helper to troubleshoot multiple areas in order to identify and solve a problem. For example, a service technician's interactions with a human–machine interface (e.g., a button push) may result in the movement of mechanical parts at a different location on a large machine (e.g., a robot arm with a tool that moves up or down). In such a scenario, the helper may act like a second on-site worker by using multiple cameras to monitor machine movements that are beyond the technician's field of view (FoV).

However, a critical challenge of multi-camera remote assistance systems is introduced by the helper's freedom to navigate between cameras: due to the disjointed camera views, spatial and gestural information that is naturally available between co-located people is lost [5]. Importantly, the worker lacks awareness of the helper, including information about the helper's location (i.e., which camera the helper is currently viewing), what parts of the workspace the helper can see (i.e., the cameras' FoV), and what part of the scene the helper is focusing on (e.g., foreground or background objects). We address this issue by proposing the use of Augmented Reality (AR) cues for improving the worker's awareness of the helper's location and focus in a large workspace that is accessed through multiple cameras. To this

[☆] This paper has been recommended for acceptance by Dr. Zicheng Liu. The work reported here is funded by MADE Digital (Project 6151-00006B).
[*] Corresponding author.
E-mail address: troels.rasmussen@cs.au.dk (T. Rasmussen).

https://doi.org/10.1016/j.jvcir.2022.103655
Received 30 September 2021; Received in revised form 10 May 2022; Accepted 27 September 2022
Available online 18 October 2022

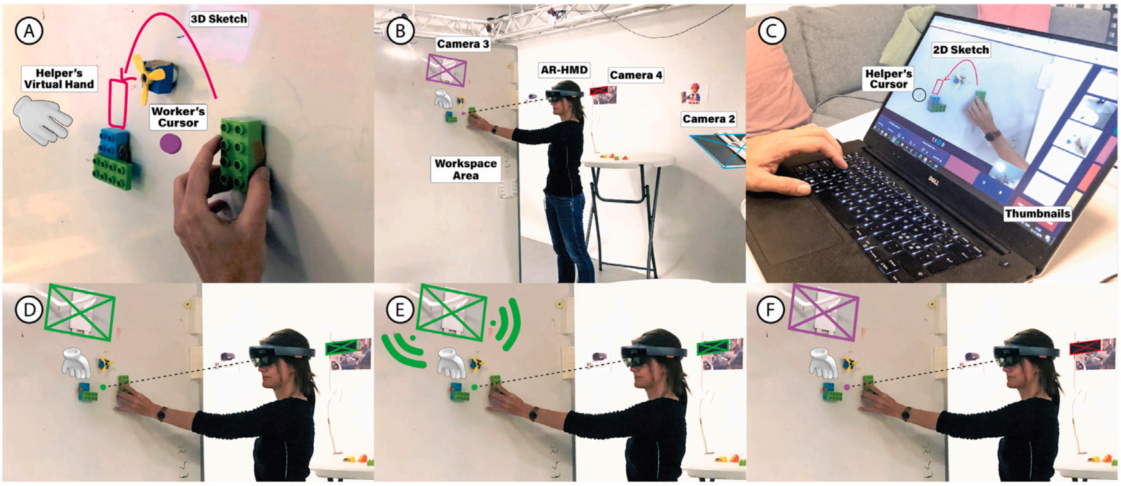


Fig. 1. We present a multi-camera remote assistance system, *CueCam*, through which a remote helper can guide a worker with annotations and pointing in a large workspace. The worker sees these annotations through an AR HMD (A). Critical areas of the workspace are captured by multiple cameras (B), and the helper can view and annotate these individual video feeds on a PC (C). To aid the worker in locating the helper in the workspace (i.e., find the camera viewed by the helper), our system supports three awareness cues. We evaluate these in a user study with the following conditions: (D) *Virtual Hand Only*, (E) *Virtual Hand + Spatial Sound*, (F) *Virtual Hand + Color Cue*.

end, we present our multi-camera remote assistance system, *CueCam*, that supports three AR awareness cues, as illustrated in Fig. 1: (D) Virtual Hand, (E) Spatial Sound, and (F) Color Cue. The visual cues, Virtual Hand and Color Cue, are perceived by the worker through an AR head-mounted display (HMD), while Spatial Sound is emitted from the cameras' integrated speakers. As described further in Section 3.4, these cues make use of the visual and auditory modalities and thereby support different strategies for identifying the helper's location and focus. We evaluated the effectiveness of our awareness cues in a user study, comparing the following awareness cue combinations (1) *Virtual Hand Only*, (2) *Virtual Hand + Spatial Sound*, and (3) *Virtual Hand + Color Cue*. The Virtual Hand is a basic tool for remote assistance, allowing the helper to point at objects. Thus, we assume this as the baseline and compares it to combinations, where the Virtual Hand is supplemented by either Spatial Sound or a Color Cue. Our findings reveal that all awareness cues are effective, but the addition of a Color Cue or Spatial Sound is beneficial. Further, the Color Cue led to overall best performance in locating the helper and was the preferred cue in most cases.

An additional challenge with multi-camera systems is the setup and calibration, which can be cumbersome and time consuming [6]. Similar to earlier work by Rasmussen et al. [7], we address this by proposing the use of off-the-shelf camera devices, such as regular webcams, smartphones, and tablets, and describe a novel AR-based ad-hoc camera calibration procedure in Section 3.2, which facilitates quick and flexible (re)configuration of multiple scene cameras (i.e., adding, moving, or removing cameras) during use of our *CueCam* system.

In addition to the awareness cues and camera calibration procedure, *CueCam* also supports elementary functionalities for remote assistance: Using a WIMP (windows, icons, menus, pointer) interface on a PC, the helper can point and draw on the video feed of each camera. The pointer position and drawings are then projected into the 3D reconstruction of the workspace and visualized to the worker through AR.

In summary, aiming to support mutual workspace awareness, our contribution is two-fold: (1) For the local worker, we reveal the remote helper's location and focus in a large shared workspace through AR awareness cues. We propose the use of three distinct visual and auditory cues and share insights about their effectiveness and applicability based on a lab study. (2) We enable quick and flexible configuration of a multi-camera setup in a large workspace through an AR-based ad-hoc camera calibration procedure. This is intended to facilitate the helper's independent workspace exploration and ability to negotiate an appropriate view of the workspace.

In the following sections, we first review related work on multi-camera remote assistance and awareness cues. Then, we describe and discuss the design of our multi-camera remote assistance system, *CueCam*, including core functionality, ad-hoc camera calibration, and awareness cues. This is followed by a report of our evaluation of the awareness cues. We conclude with a discussion of design implications, system and study limitations, and future work.

2. Related work

As we present a remote assistance system that establishes a shared visual space for the remote helper and local worker through multiple cameras, we begin by discussing existing research on multi-camera systems. We then proceed to review work on mixed reality (MR) awareness cues for remote assistance, as a basis for discussing our proposed AR awareness cues. Note that various visual communication cues, such as sketch cues [8,9], hand gestures [10,11], and pointers [8], have previously been explored for giving *explicit* remote instructions. In contrast to such cues for explicit communication, we focus on *implicit* awareness cues that convey information about a collaborator's location and activities in a shared visual space.

2.1. Sharing a visual space through multiple cameras

Multi-camera approaches have been explored to address a number of challenges, such as to support view independence of remote collaborators [7,12] or provide visual information at various levels of detail [5,7,12–14]. For example, Gaver et al. [5] aimed to support collaboration on physical tasks across remote office spaces. Multiple scene cameras captured both a contextual overview of the workspace and detailed views of task-related artifacts in each office. Among others, the researchers uncovered the following challenges for multi-camera remote assistance:

1. The worker lacks *awareness* of the helper's focus of attention, i.e., which camera she is viewing and what she is focusing on in the camera feed. This makes it difficult for the worker to know whether the helper can see his gestures and actions without constant verbal confirmation. To address this, we propose three AR awareness cues (see Section 3.4).

2. Referencing objects using a multi-camera setup lacks the ease of interpreting pointing and gaze direction during co-located collaboration. Our remote assistance system supports remotely pointing at an object by projecting the helper's mouse cursor into the workspace as a Virtual Hand in AR (see Section 3.3).
3. The camera views are presented to the helper in a disjointed manner, without conveying the spatial relationship between cameras, work areas, and the worker. This issue is addressed by research on focus-in-context systems [7,13,15–18] that provide a contextual overview of how the different work areas are spatially connected, while additional cameras offer focus views of the worker's manipulations in the areas.

Besides the above challenges, Aschenbrenner et al. [18] showed that permanently installing scene cameras in an industrial workspace can be problematic for obtaining the right view of a problem. We aim to address this challenge in *CueCam* with an AR-based calibration method that supports flexible ad-hoc configuration and movement of scene cameras (see Section 3.2).

Beyond regular RGB cameras, related research has also proposed the use of multiple RGB-D cameras to stitch together live 3D reconstructions of the worker's space. The remote helper can independently navigate this reconstructed space on a desktop [2,6], or in MR for increased sense of immersion and co-presence [19–21]. Lately, the use of 360-degree cameras for remote assistance in MR has also received increased attention [21–27]. When worn by the worker, a 360-degree camera can provide the helper with a panoramic view of the workspace, which technically corresponds to multiple camera perspectives and can be explored independently of the worker's orientation. Despite the appeal of using multiple RGB-D cameras to reconstruct the workspace or using a 360 camera to support full view independence, we intentionally focus on multi-camera remote assistance with plain RGB cameras, (i.e., no depth data is assumed available) to reduce hardware and processing demands. RGB multi-camera systems are less dependent on high network speeds, bandwidth, and computational power since depth data and live 3D reconstructions are omitted. They can therefore more easily be run on mobile devices (tablets/smartphones). Further, detailed high-resolution views of the workspace can be achieved by using commodity RGB cameras, whereas commodity RGB-D cameras only generate comparably low-resolution 3D reconstructions, which may not fulfill the users' needs: previously conducted interviews with real-world service technicians and remote experts from the manufacturing industry reveal that mobility and detailed high-resolution views of a workspace are important to the helper, due to urgent, nomadic problem-solving needs and the desire to make detailed comparisons between an industrial machine and its schematics [28].

More recent technological advances have enabled the use of a single camera to create a static 3D reconstruction [29–31] or light field [32] by panning the camera over the workspace to capture multiple images, before initiating the remote assistance session. Such approaches are ideal for scenarios that involve few structural changes in the workspace. However, they become cumbersome in dynamic scenarios, such as assembly or maintenance, where the state of task objects changes frequently, as updating the representation (i.e., 3D reconstruction or light field) after each change requires time-consuming manual camera work. In comparison, a multi-camera system provides a view independent live representation of the workspace, thus capturing dynamic changes to task objects.

2.2. Awareness cues for remote assistance

Gutwin et al. [33] introduced the concept of workspace awareness, which they regard as a specialization of situation awareness [34] and define as an *"up-to-the-moment understanding of another person's interaction with the workspace"* during collaboration. They categorized sources of workspace awareness information into consequential communication (body movement), feedthrough (object manipulations), and intentional communication (e.g., pointing). Since these sources of information are lost during remote assistance, awareness cues must be consciously designed and mediated by the system. Awareness cues that convey body movement as a consequence of a collaborator's interactions in a visually shared workspace are often used as an implicit source of information on location and focus of attention.

With the advent of AR/MR (mixed reality) for remote assistance, awareness cues for visualizing location and attention in 3D space have received increased interest. For instance, visual awareness cues, including the users' head pose, view frustum and eye-gaze ray, have been found to be helpful for a remote helper and a local worker to maintain awareness of each other's attention within a shared 3D reconstruction of a workspace in MR [29–31,35]. Such cues also have relevance when using 360-degree cameras for remote assistance, to address the challenge of communicating the collaborator's viewing direction in the shared panoramic view [22,24]. For example, Lee et al. [22] added an MR view frame in the form of a colored rectangle to indicate the collaborator's current perspective within the panoramic view. Their results indicate that this improved the collaborators' awareness of each other, helped reconcile perspectives and served as a pointer. Further, an additional arrow showing the direction to the collaborator's view frame was perceived as useful by the workers.

The Virtual Hand awareness cue supported by our system shares an important characteristic with the cues for head pose, view frustum, head gaze [29–31], and view frame [22,24] from related work: it must be in the worker's FoV to allow identification of the helper's location and focus of attention, and will otherwise require a visual search. We compare this to an awareness cue that relies on the auditory sense (Spatial Sound), similar to the recent work by Yang et al. [36], and an awareness cue that is visually persistent (i.e., always in the user's FoV) and depends on spatial memory and color mapping (Color Cue) to locate the helper. To enable the latter, we assume that remote assistance takes place in a discrete number of areas, which are captured by the scene cameras in our multi-camera setup (see details in Section 3.4).

The comparison of different guidance modalities for remote assistance is not novel in itself. For example, Gunther et al. [37] compared visual, auditory, and tactile communication cues. Similar to how we study the effect of combining awareness cues, Kim et al. [38] explored the combination of pointer, hand, and sketch. The most important distinction to our work is that they both explore guidance in a small desk-sized workspace covered by a single camera, while our system serves to study awareness cues in a room-sized workspace covered by multiple cameras. We found such remote assistance scenarios involving large workspaces, which require the worker and helper to navigate to multiple areas while maintaining awareness of each other, to be under-explored. We thus aim to contribute to this field by proposing a multi-camera remote assistance system that supports AR-based ad-hoc calibration of cameras, and evaluating three types of awareness cues that allow the local worker to adopt different search strategies for locating the remote helper in a large workspace.

3. *Cuecam*: AR multi-camera remote assistance

Our goal is to support mutual workspace awareness during remote assistance through a (re)configurable multi-camera system that allows the remote helper to "move" through a large workspace, while supporting the worker's awareness of the helper's location and focus through a variety of AR cues. In this section we describe the design and implementation of this AR multi-camera remote assistance system, *CueCam*, including its core functionality, support for ad-hoc calibration of cameras, and AR awareness cues.

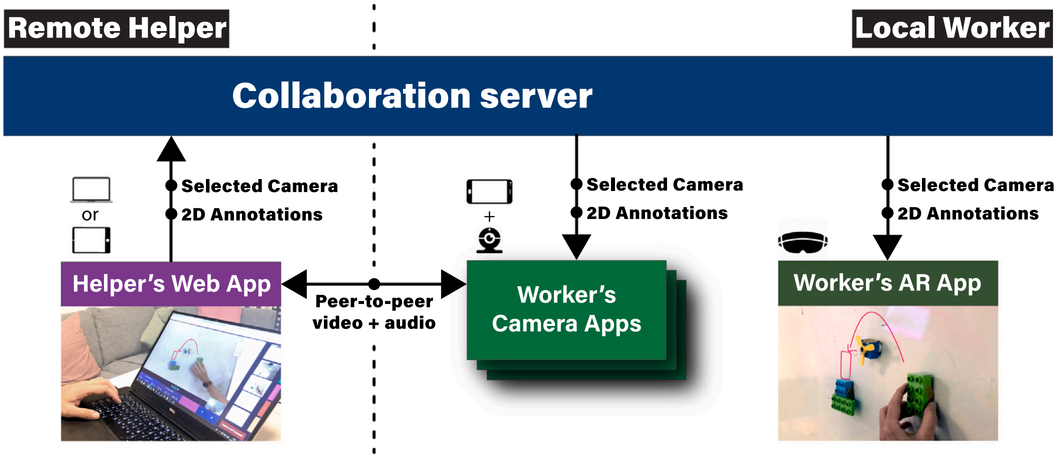


Fig. 2. System architecture of *CueCam*: A Collaboration Server coordinates the transmission of communication cues, by passing 2D annotations and the current camera selection from the Helper's Web App on to the Worker's Camera Apps and AR App. Live video and audio are shared through a peer-to-peer connection between the Helper's Web App and the Worker's Camera Apps.

3.1. Core functionality

The presented system supports remote assistance through a Web App for the helper and an AR App for the worker. Fig. 2 shows an overview of the different components of both applications, and the communication flow between them. In the following, each is described in more detail.

3.1.1. The worker's AR App and Camera Apps

In our particular implementation, the AR App is developed in Unity3D.[1] and runs on a head-mounted display, a Microsoft Hololens (version 1)[2] The Hololens, which relies on inside-out tracking, creates and continuously updates a virtual surface mesh reconstruction of the workspace. This allows the AR App to calibrate the cameras in the workspace (see Section 3.2), and supports projection of the helper's cursor and drawings into the worker's environment in 3D (see Section 3.3). The AR App further supports voice communication with the helper, by sending and receiving audio. Finally, it also enables presenting a range of awareness cues that indicate the location of the helper's viewpoint and focus (see Section 3.4).

Apart from the Hololens, the worker relies on several camera devices (so-called "scene cameras"), which he places throughout the workspace to capture relevant parts of the environment. Each of these camera devices runs a Camera App, which is a web application developed in JavaScript (ECMAScript 6+). This Camera App fulfills two purposes: (1) it transmits live video from the worker to the helper and audio both ways; (2) it is capable of displaying an AR marker on camera devices that include displays (e.g., smartphones, tablets), which are used for calibration. Importantly however, the Camera App is not an AR application: it is agnostic of its environment and leaves all surface reconstruction and tracking to the AR App. This allows the worker to use nearly any kind of computer device with an RGB camera as a scene camera, as long as it supports wireless connectivity and is browser compatible (e.g., smartphone, tablet, PC + webcam).

3.1.2. The helper's Web App

On the helper's side, the Web App runs in the internet browser, e.g., on a PC or tablet. This application is developed in JavaScript (ECMAScript 6+). Its main function is to receive live video and audio from the worker's cameras and present these to the helper, as well as to transmit the helper's audio stream and 2D annotations (i.e., cursor and drawings) to the worker. The Web App offers the helper simultaneous access to different views of the worker's space through each of the camera video feeds, which are presented as thumbnail previews on the right side of the interface (see Fig. 1, C). Upon clicking on a thumbnail, the selected video is shown in the large central video window, on which the helper can then make pointing gestures and drawings that are shown to the worker in AR. The helper's audio is transmitted to the selected camera device in the worker's space.

3.1.3. Communication protocols

The helper's Web App and the worker's multiple instances of the Camera App act as peers in a network, and transmit video and audio using WebRTC.[3] Further types of live information are exchanged between the worker and helper applications using a Node.js collaboration server and websockets. For example, the Web App transmits the helper's 2D annotations and information about the currently selected camera to the worker's AR App. The AR App then uses this information for projecting the annotations into the 3D workspace (see Section 3.3).

3.2. Ad-hoc calibration of cameras

Generally, cameras of AR remote assistance systems require calibration, i.e., calculation of the camera's intrinsic and extrinsic parameters, to enable correct augmentations in the workspaces they capture. When using *CueCam* the worker places multiple scene cameras in the workspace to provide the helper with views of different work areas that are relevant to the remote assistance task. To ensure correct presentation of the AR awareness cues and the helper's 2D annotations in the worker's 3D environment, the scene cameras must be calibrated: The intrinsic parameters and distortion coefficients of each camera are obtained through a standard calibration procedure before use,[4] while the extrinsic parameters are computed ad-hoc through the following calibration process, as also illustrated in Fig. 3.

1. The worker approaches each scene camera in the work space and scans a unique image target (marker) on the camera using the AR App on the HMD (see Fig. 3). This is achieved with the voice command "Scan Marker", which triggers capture of the marker with the front-facing camera integrated in the HMD. When using tablets or smartphones as scene cameras, the marker

[1] https://unity.com/

[2] https://docs.microsoft.com/en-us/hololens/hololens1-hardware

[3] https://webrtc.org/

[4] OpenCV Documentation: https://docs.opencv.org/master/dc/dbb/tutorial_py_calibration.html.

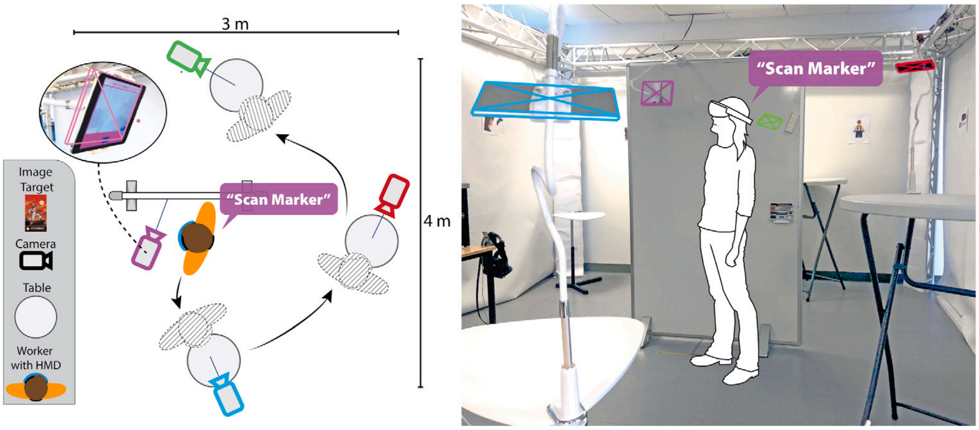


Fig. 3. Calibration process. (Left) Top-down view illustrating the worker's path through the workspace when populating the space with virtual cameras by scanning an image target on each real camera. (Right) View of the workspace after calibration, with the aligned virtual cameras superimposed on the real cameras as colored wireframes.

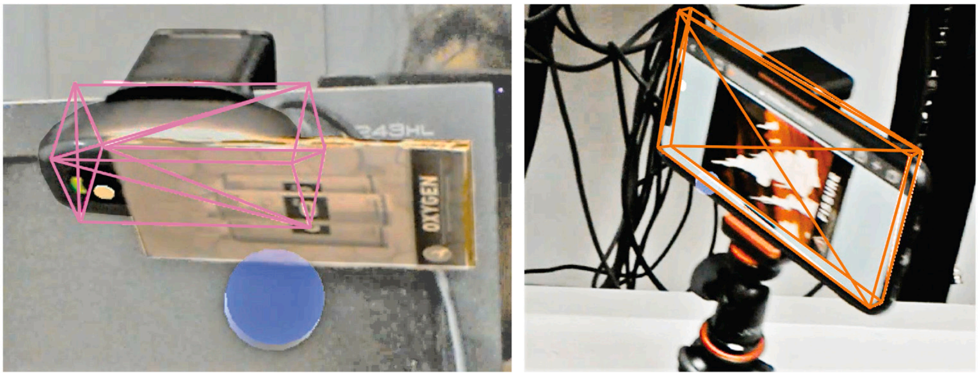


Fig. 4. Markers on the camera devices are used for calculating the extrinsic camera parameters and aligning virtual camera to physical camera. (Left) Physical marker on webcam. (Right) Virtual marker on screen of smartphone.

can simply be displayed on the device's screen, while regular cameras require attachment of a physical marker (e.g., printout on paper attached to webcam; Fig. 4). The AR library Vuforia is used for marker tracking.

2. Upon scanning the marker, its pose and the associated camera ID are extracted.
3. The pose of the associated camera is computed based on the tracked marker pose. For this, a rigid transformation between the marker and camera lens must be determined a priori and entered into the system programmatically. This is usually a simple translation due to the offset between marker and camera lens.
4. A virtual camera is created with the same extrinsic and intrinsic parameters as the physical camera. The virtual camera appears to the worker as a wireframe model that is superimposed on the physical camera (see Fig. 5), providing visual feedback for the achieved calibration.
5. As the worker moves about the workspace, the HMD continues to keep track of the virtual cameras in the scene through its inside-out tracking capabilities.

This process must be done for any camera that is added to, moved, or removed from the workspace. Hence, if a single camera in the setup is moved, only that particular camera must be recalibrated by simply re-scanning its marker with the HMD, while existing calibrations are maintained for the remaining (unchanged) scene cameras. This facilitates easy reconfiguration of the camera setup, allowing the worker to add and move scene cameras ad-hoc *during* remote assistance. In contrast, in related work the placement of scene cameras is usually assumed to be fixed after initial calibration, which must be done *before* remote assistance commences and often involves tracking of one central marker with known placement in the workspace [4,6]. Importantly, our approach enables the helper and worker to negotiate appropriate views of the workspace throughout their remote assistance session, which may contribute to the helper's workspace awareness. Furthermore, in comparison to using a central marker in the workspace for calibration that favors an outside-in multi-camera configuration with overlapping FoVs, our approach is agnostic to the particular spatial configuration of cameras: The calibration procedure works the same regardless of whether the cameras cover separate areas, or whether their views overlap. See Fig. 8 for examples of spatial camera configurations supported by our procedure.

3.3. 3D interpretation of 2D annotations

From knowing the pose (extrinsic parameters) and the intrinsic parameters of the scene cameras (see Section 3.2), the AR App can interpret the helper's 2D cursor position and drawings on a camera's video feed as 3D annotations. For this 3D interpretation, we currently make use of the spraypaint technique [39], where a 2D pixel position in screen space is projected into the world along a ray from the camera's focal point. The resulting 3D position of the pixel is at the intersection point of the ray with the reconstructed surface mesh of the workspace. Consequently, as is shown in Fig. 1 (A), the helper's cursor appears as a Virtual Hand that moves along the reconstructed surface of the 3D workspace. Similarly, when the helper makes drawings, the projection from screen space to world space is done for every drawn pixel. Hence, to the worker the drawings appear mapped onto the reconstructed 3D surfaces in the workspace. Precise 3D interpretation of 2D annotations

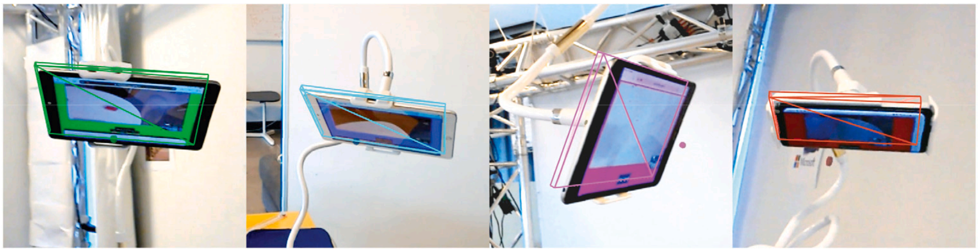


Fig. 5. From the point of view of the worker's Hololens in AR after calibration: The four scene cameras are augmented with the wireframes of their corresponding virtual cameras. In the *Virtual Hand + Color Cue* condition, the wireframe of each camera has a distinct color, while in the other awareness cue combinations all wireframes are green. (The misalignment of wireframes is a result of the screen capture method. Alignment is accurate, when seen through the HMD).

requires that we model the scene cameras with accurate values for camera extrinsics, intrinsics, and distortion coefficients. We achieve this through the calibration procedure described previously.

3.4. Augmented reality awareness cues

From the worker's point of view, a helper, who switches between multiple camera views in the Web App to access different areas in the workspace, appears to "teleport" from one location to another. Through his AR App, the worker might see the helper pointing at a component in one area of the workspace, and in the next instant sees her drawing in a completely different area. Thus, it can be difficult for the worker to remain aware of the helper's location and activities. We aim to address this issue with three different AR awareness cues: (1) Virtual Hand, (2) Spatial Sound, and (3) Color Cue, which are described in more detail below.

The design of these AR awareness cues was motivated by the aim to systematically explore multiple output modalities and different search strategies (see Section 4), thereby extending on the related work. Initially, a guiding AR arrow was designed as a fourth cue, based on the Hololens design guidelines from Microsoft. This arrow was persistently visualized next to the Hololens AR cursor and indicated the direction to the helper's viewpoint. However, in a pilot study the arrow was perceived as misleading and hampered participants' performance. Hence, this variant was not explored further.

Virtual Hand: The Virtual Hand is the helper's User Representation [40] in the physical workspace and serves as an explicit communication cue, as it represents the helper's mouse cursor and allows pointing and drawing within the worker's 3D space: When the helper moves her mouse cursor around on a live video from the currently viewed camera, the 2D cursor position is projected into the worker's space and shown as a 3D Virtual Hand (see Fig. 1, A and C). The pose of the Virtual Hand is determined by the direction of the vector from the focal point of the currently viewed camera to the intersection point with the reconstructed surface mesh. Hence, the Virtual Hand implicitly acts as an awareness cue, informing the worker about the helper's *location* (i.e., the location of the camera currently viewed by the helper) and *focus* (i.e., the object of interest according to the position of the Virtual Hand in the workspace). Further, when the helper draws on a video by dragging the mouse during left click, the helper's 2D drawing is projected into the worker's space as a 3D drawing. Hence, to identify the helper's location and focus, the worker can visually search the work areas for the Virtual Hand or an annotation that is being created.

Spatial Sound: Loudspeakers integrated in, or co-located with, the camera devices can emit Spatial Sound to provide feedback about the helper's location. For this awareness cue we propose that during conversation the worker hears the helper's voice from the camera that is currently being viewed (illustrated in Fig. 1, E). Further, the viewed camera device emits continuous white noise when the helper moves her cursor over the video window, and the sound of a pencil on paper when drawing. The latter is inspired by the concept of consequential communication: co-located collaborators see and hear each other as a consequence of their activities in the workspace without the need for constant verbal communication [33]. In similar manner, Spatial Sound can provide continuous information about the helper's location and activities and may also be used in a call-and-response fashion when needed (e.g., worker: "Where did you jump to now?"; helper: "I'm over on this side".). Thus, Spatial Sound can inform the worker about the helper's location and activities through an auditory search.

Color Cue: With this cue type the worker perceives a persistent Color Cue in his FoV-in our particular implementation we use the Hololens AR cursor, a colored disc located at the intersection between a ray cast from the head of the worker in gaze direction and the mesh-reconstruction of the environment. For the purpose of Color Cue, each of the cameras is assigned a unique color (green, blue, red, or pink), and the color of the cursor then reveals which camera the helper is viewing (see Fig. 1, F). The assigned camera colors are made evident by coloring the corresponding camera wireframes in the worker's AR view (Fig. 5). Thus, when a change in the Color Cue indicates that the helper has switched camera, the worker can locate the helper's new viewpoint by mapping the color of Color Cue to that of the cameras. This may require a visual search strategy, but is likely also supported by the worker's spatial memory of the camera-color mapping. Color Cue is inspired by the realization that the number of locations (or viewpoints) the helper can "navigate to" in a multi-camera remote assistance system are limited; in contrast to remote assistance systems that make use of shared 3D reconstructions [30,31] or 360-degree views [22,24], where the helper can navigate with 6DOF or 3DOF respectively and hence must be located in a continuum. The fixed number of locations in a multi-camera system makes it possible to represent these with colors or to use other symbolic representations.

The three proposed cues differ in the type of information they convey, the modality this information is conveyed by, and the search strategies they require from the worker. While Virtual Hand provides information both about the helper's location (i.e., currently viewed camera) and focus in the camera view (position of cursor), Spatial Sound and Color Cue only indicate the helper's location. The latter may not suffice, when a camera covers a large work area, or an area with a multitude of components that cannot easily be identified by verbal descriptions. Therefore to ensure that both the location and focus of the helper is available, and because the ability to point and draw annotations in the worker's space has become a common feature in remote assistance solutions, we propose to apply Virtual Hand in combination with Color Cue and Spatial Sound.

4. Evaluation of AR awareness cues

To evaluate the effectiveness and experience of the awareness cue combinations in our multi-camera remote assistance system, we conducted a controlled lab study with participant pairs, who took turns acting as worker and helper. Aiming to explore how the proposed the AR awareness cues contribute to the worker's awareness of the helper in a large workspace, the worker was tasked with repeatedly identifying the helper's location and focus in the workspace.

4.1. Study design

In a within-subjects design we compared the following three awareness cue combinations (conditions) in counterbalanced order:

1. ***Virtual Hand Only* (VH; baseline condition):** When the helper points to something in the live video feed with her mouse cursor, the worker sees this cursor projected into his workspace as a 3D virtual hand.
2. ***Virtual Hand* + *Spatial Sound* (VH+SS):** In addition to the Virtual Hand, a continuous sound is emitted from the camera device that the helper is currently viewing, while she moves her cursor on the video feed.
3. ***Virtual Hand* + *Color Cue*(VH+CC):** In addition to the Virtual Hand, the color of the worker's AR cursor changes to reflect the unique color assigned to the camera currently viewed by the helper.

For the worker, the default color of his circular AR cursor and the virtual outlines of all cameras is green; the cursor and outline colors only differ in the *Virtual Hand* + *Color Cue* condition. To aid understanding, Fig. 1 (D-F) provides an illustration of these three conditions.

To locate the helper in both the *Virtual Hand Only* and *Virtual Hand* + *Spatial Sound* conditions, the worker must search the workspace to find the hand or the sound source. In the first case a visual search for the virtual hand is needed. In the latter the helper's new location is advertised by a shift in sound source, which may be beneficial when the virtual hand is beyond of the worker's FoV or occluded by physical obstacles. The *Virtual Hand* + *Color Cue* condition further provides a persistent cue of the helper's location in form of the Color Cue, which is always in the worker's FoV. The worker can identify the helper's view, by matching the color of the cue cursor to the correspondingly outlined camera. Spatial memory may aid the recall of colors and camera locations, so that no visual search is needed.

Note that we do not compare to any no-cue condition, in which no aid is given for finding the helper's location. This is due to the Virtual Hand being elementary for enabling explicit remote assistance (i.e., guiding a worker through a physical task by pointing out components and demonstrating actions), which implicitly provides information about the helper's location and focus. Hence, this study evaluates the benefit of adding further awareness cues (Spatial Sound, Color Cue) to the baseline condition (*Virtual Hand Only*). The study design consists of one independent variable with three levels (awareness cue combinations) about which we have made the following hypotheses:

H1 Workers perform better at locating their helper with *Virtual Hand* + *Color Cue*, compared to the other awareness cue combinations. We reason that spatial memory facilitates the mapping of Color Cue to scene cameras with unique colors, thereby reducing the need for the worker to perform a visual or auditory search for the helper.

H2 Workers perform better with *Virtual Hand* + *Spatial Sound* than with *Virtual Hand Only*, since sound can be perceived and located without facing it directly. In contrast, the Virtual Hand can only be located when it is in the worker's FoV and a visual search for it may require more time.

H3 Workers prefer *Virtual Hand* + *Spatial Sound* over the other conditions, since it does not require a visual search for the Virtual Hand, nor memorizing the location and color of cameras for interpreting the Color Cue.

For H1 and H2 we evaluated the workers' Task Performance in the search task by counting the number of helper locations that were correctly identified with each awareness cue combination. To test H3, we evaluated the workers' Preference as a ranking of awareness cue combinations (from best to worst) in a post-study questionnaire. Further, after each trial, workers were asked to indicate the Ease of Use for each awareness cue combination on a 5-point Likert scale.

4.2. Experimental task and apparatus

To compare the effectiveness of our cues, we designed a search task in which the worker (AR user) was asked to repeatedly locate the helper (PC user) in a large, room-sized workspace for the duration of two minutes. It should be noted that this search task does not reflect the complexity of an authentic remote assistance scenario, but was chosen with the intention to limit confounding factors. This task was repeated for each awareness cue combination.

The worker's setup was constructed in the following way: Three tables and a whiteboard were distributed in a 3 × 4-meter workspace, as illustrated in Fig. 3. These represent different work areas. Four scene cameras (2 *x* iPad 6 with Safari 12, 1 *x* iPad Mini 4 with Safari 12, 1 *x* Galaxy s9 smartphone with Chrome) were mounted on the tables and whiteboard, such that their front-facing cameras captured a specific flat surface in the respective work area (i.e., tabletop/whiteboard surface). Importantly, the workspace was designed to warrant the use of multiple cameras, as is illustrated in Fig. 8 (Config 1). Therefore, the work areas were spatially configured such that it was not practical to capture the workspace in its entirety and in detail from the perspective of a head-worn camera or single mounted camera alone. Further, simulating scenarios that require fine grained manipulations, the cameras were mounted such that they capture a high-detail view of their respective work areas. As a result, the distance between the helper's viewpoint (camera location) and focus (location of Virtual Hand on work area) is small, allowing the worker to identify the helper's focus based on each camera location.

The helper was seated at a desk adjacent to the worker's lab space. To better emulate remote assistance, a wall divider ensured that participants could not see each other. The helper interacted with the Web App on a laptop computer (Dell XPS with an Intel Core i7 2.80 GHz CPU) and was instructed to move her mouse cursor around on the active video view of a work area. To the worker, the helper's cursor appeared as a Virtual Hand moving on the reconstructed surface mesh of the respective work area. Upon identifying the helper's location and focus, the worker indicated this by placing his palm on the surface where the Virtual Hand was projected. The helper then immediately "moved" to another work area by selecting another camera thumbnail in her Web App. A colored icon in the Web App indicated, which of the four color-coded camera views to select next. Worker–participants were instructed to correctly locate the helper in as many work areas as possible, without running. The number of successfully identified helper locations was visualized to the worker after each trial, to motivate high performance.

To ensure fair comparisons between participants, the order of consecutive camera views/work areas was pre-computed for each condition as a random sequence, while ensuring the same level of difficulty and same overall minimum path length for all participants. Each of these sequences consisted of 200 entries, referring to one of the four installed scene cameras. The same camera view never appeared twice in a row (i.e., there was always a change of work area/camera view).

4.3. Participants

We recruited 12 voluntary participants (2 female, 10 male), ranging from 23 to 34 years of age (avg. age: 27). They were primarily recruited from our university campus, via email and online groups. English language proficiency and normal or corrected-to-normal vision were acceptance criteria for participation. Participants were paired up in 6 groups, whereby 5/6 pairs knew each other before the study. Participants were compensated for their time with treats. Based on 7-point Likert scale ratings ("never" to "very often"), 8/12 participants had never or very rarely used AR, 3 indicated often or very often, and one participant was undecided. One participant, P11, was red-green color blind, thus the colors of the cameras and Color Cue used in the *Virtual Hand* + *Color Cue* condition had to be adjusted for his session. He was asked to confirm before the experiment, that four new camera colors (white, blue, red, pink) were easily distinguishable.

4.4. Procedure

The workspace with the scene cameras was calibrated by the experimenter before the study. Participants arrived in pairs, received information about the experiment together, and then each gave their informed consent and completed a pre-study demographics questionnaire. They then received a detailed introduction to the multi-camera remote assistance system and took turns trying out the worker's AR App and the helper's Web App. To help them attain the right mental model of how the system worked, we pointed out that 2D annotations made on the video window by the helper were converted into 3D AR annotations in the worker's space. Furthermore, in a small training task, they each practiced locating the respective helper with the three proposed awareness cues. Participants were then randomly assigned the role of worker or helper for the first part of the study, in which they performed the 2-minute search task described above for each awareness cue combination in counterbalanced order. Before each trial, the current awareness cue combination was again explained to the participant in the worker role. During the trials, participants were asked not to communicate verbally during trials, but to rely solely on the awareness cues. After each trial, the participant in the worker role was asked to fill out a brief questionnaire, rating the Ease of Use and their qualitative experience with the respective awareness cue combination. After conclusion of all trials, the worker filled out a post-study questionnaire, ranking the awareness cues from best to worst and providing additional qualitative feedback. Then followed the second part of the study, in which participants repeated the above procedure in swapped roles.

5. Results

We report on the results of our controlled study with three dependent variables, i.e. the different awareness cue combinations. Task Performance, Ease of Use, and Preference data as well as qualitative feedback was collected. For statistical analysis in R,[5] we define an α-level of 0.05.

5.1. Task performance

Fig. 6 shows the collected Task Performance data for the search task for each condition. Mauchly's test shows no violation of sphericity ($W(2) = 0.72, p = 0.19$). With one-way repeated-measure ANOVA, we found a significant effect of awareness cue combination on Task Performance ($F(2,22) = 15.34$, p<0.01, $\eta^2 = 0.58$ with $CI = [0.18, 0.63]$). Post-hoc comparisons using paired t-tests revealed the significant differences between all conditions: On average, participants correctly identified 34.33 ($SD = 7.46$) helper locations with *Virtual Hand + Color Cue*, thereby performing significantly better than with *Virtual Hand Only* with 25.17 locations ($SD = 5.86$; $p < 0.01$), as well as *Virtual Hand + Spatial Sound* with 30 locations ($SD = 4.97$; $p < 0.05$). This lends support to *H1*. Also *Virtual Hand + Spatial Sound* led to significantly better Task Performance than *Virtual Hand Only* ($p < 0.05$), supporting *H2*.

According to Kendall's rank correlation test, participants' Task Performance did not correlate with their indication of prior AR experience ($\tau(31) = 0.13, p = 0.34$).

5.2. Ease of use and user preferences

After each trial we asked the worker to rate their agreement with the statement *"It was easy to use [awareness cue condition] to locate the helper"* on a scale from 1 ("strongly disagree") to 5 ("strongly agree"). The results are visualized in Fig. 6. A Friedman test revealed a significant effect of awareness cue combination on ratings of Ease of Use ($X^2(2) = 8, p < 0.05$). Posthoc pairwise comparisons using Wilcoxon rank sum test with Bonferroni correction showed that participants found the *Virtual Hand + Color Cue* significantly easier to use than the *Virtual Hand Only* ($p < 0.05, r = 0.54$). No significant difference were found between *Virtual Hand Only* and *Virtual Hand + Spatial Sound* ($p = 0.22$), or *Virtual Hand + Spatial Sound* and *Virtual Hand + Color Cue* ($p = 0.22$).

After the *Virtual Hand + Color Cue*-trial we asked participants to rate their agreement with the statement, "Instead of locating the helper by the color of the cursor, I looked for the helper's virtual hand": 8 out of 12 participants rated the statement with a 1 ("very rarely") and 1 participant rated it with a 2 ("rarely"). So, the majority of participants did not rely on the Virtual Hand, when the Color Cue was available. However, some participants had issues with the Color Cue as is evident from the qualitative results below.

Participants were asked to rank the awareness cue combinations by indicating their first (rating: 1), second (rating: 2) and third choice (rating: 3). Pairwise Fisher's exact tests revealed that preference significantly differed by awareness cue combination (VH vs. VH+SS: $p < 0.05$; VH vs. VH+CC: $p < 0.01$; VH+SS vs. VH+CC: $p < 0.01$). As can be seen from the visualization of these results in Fig. 7, *Virtual Hand + Color Cue* was predominantly the first choice (ranked 1st by 9/12 and 2nd by 3/12 participants). *Virtual Hand + Spatial Sound* achieved second place ranked 1st by 2/12, 2nd by 7/12 and 3rd by 3/12 participants, while *Virtual Hand Only* was consistently least preferred ranked 1st by 1/12, 2nd by 2/12 and 3rd by 9/12 participants. Thus, H3 is not supported.

5.3. Qualitative results

In the post-study questionnaire, the worker was asked to explain their indicated preference of awareness cue combination. Most participants responded that they preferred the Color Cue for locating the helper, because it was easy to remember the colors and positions of the different cameras. Participants explained that, *"Knowing that the color of the cursor could be matched to the color of a camera, I just had to remember the position of the colors"* (P2); *"I think the strength of the color cursor is that you learn, which color is where in the room, so you do not have to rely on looking through the narrow field of view"* (P4); and *"After memorizing the colors' location, it was the easiest cue"* (P9). This is also reflected by additional data collected after the *Virtual Hand + Color Cue*-trial: all participants rated the statement "After a while I had remembered the colored cameras" either with a 4 ("agree") or 5 ("strongly agree"). P3 mentioned that he liked the color cursor better, because *"moving your head is not required"*. P4 and P8 justified their dislike of the Virtual Hand as follows: *"I liked the moving hand the least because every time the helper moved, you might have to look around at all the other locations"* (P4); *"You have to look for the hand all the time"* (P8). These statements align with observations of the experimenter that the worker would often continuously move their head in search for the cursor in the *Virtual Hand Only* condition. Only one participant, P11, who was color blind, preferred the Virtual Hand. To summarize, the Color Cue was preferred, since with it the worker did not have to search for the helper, in contrast to both Spatial Sound and the Virtual Hand.

While most people performed best with and preferred the Color Cue, this cue still posed some challenges. One issue mentioned was visibility: *"It is difficult to see the color (of the cursor), if the background is not white"* (P3) and "The red and pink color are too similar" (P8). Similarly, P11 had difficulties distinguishing between the white and pink color of the cursor and virtual cameras, so he would rely on the virtual hand

[5] R-Project: https://www.r-project.org/.

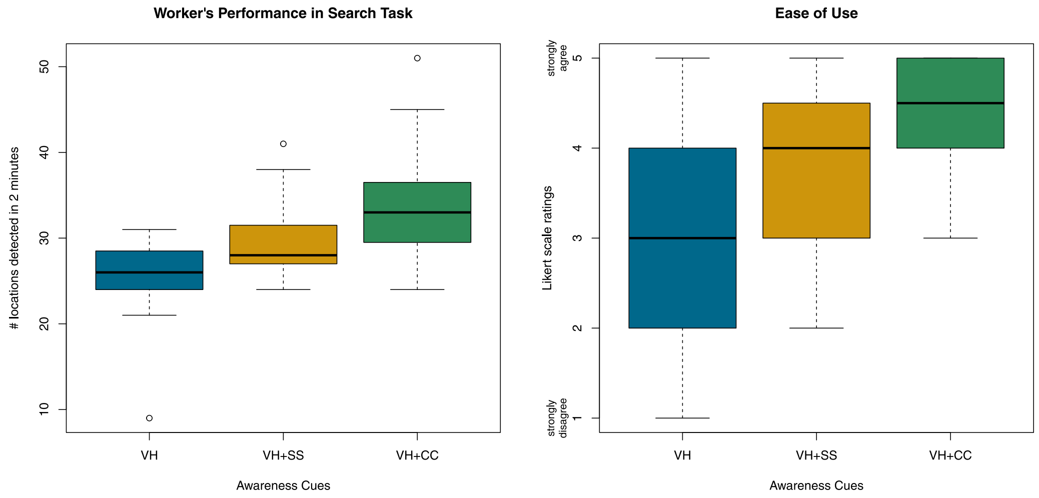
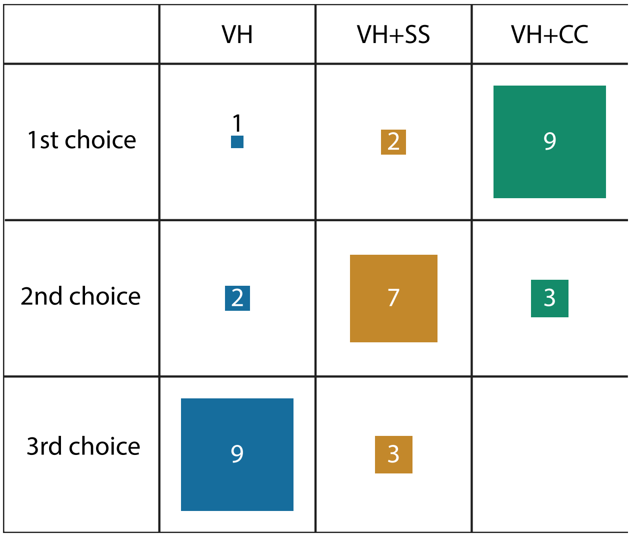


Fig. 6. (left) The worker's performance was significantly affected by awareness cue combination. Performance was lowest in *Virtual Hand Only* condition (VH) and highest for *Virtual Hand + Color Cue* (VH+CC). Y-axis shows number of areas visited within a 2-minute period. (right) Participants rated their agreement with the statement *"It was easy to use [awareness cue condition] to locate the helper"*. *Virtual Hand + Color Cue* (VH+CC) was the easiest to use, while *Virtual Hand Only* (VH) was the least easy to use.



	VH	VH+SS	VH+CC
1st choice	1	2	9
2nd choice	2	7	3
3rd choice	9	3	

Fig. 7. Participants ranked the awareness cue combinations, indicating their 1st, 2nd, and 3rd choice. *Virtual Hand + Color Cue* (VH+CC) was the first choice for 9/12 participants, while 3 rated it as second choice. Only 2 participants preferred *Virtual Hand + Spatial Sound* (VH+SS), while most named this as their second choice. Almost all participants named *Virtual Hand Only* (VH) as their least preferred (10/12).

to locate the helper in those situations. The issues with Color Cue is also evident from data collected after the *Virtual Hand + Color Cue*-trial: 2 out of 12 participants rated the statement, "Instead of locating the helper by the color of the cursor, I looked for the helper's virtual hand", with a 4 ("often"). A further challenge was the projection of the Hololens cursor onto the surface reconstruction of the environment: P9 volunteered that *"If the marker was stationary instead of disappearing into the room, it would have been easier"* and suggested to display the Color Cue as head-stabilized information, instead of using an AR cursor.

P1 and P7 preferred the Spatial Sound, even though they performed better with the Color Cue: *"Sound is the most intuitive. Looking for the color requires me to focus more on the task of searching"* (P1). Similarly, P7 preferred sound, because it provided him with a cue for what direction to move in, unlike the other awareness cues. *"I could locate next position while looking for ways to move there"* (p7). This meant that he could start moving in the general direction of the helper right away, without knowing the exact location. P3 suggested to improve Spatial Sound, by making the sound of each camera unique, thereby giving it similar properties as Color Cue, which supports memorization.

6. Discussion

In the following, we discuss design implications of our awareness cues and reflect on the sensory modalities involved. We further discuss the camera configuration chosen for our study, and present pros and cons of our camera calibration procedure.

6.1. Design implications of our AR awareness cues

While our results show that all awareness cues were effective, we found that workers were significantly faster at locating the helper in the *Virtual Hand + Color Cue* condition compared to the other conditions. The majority of workers explained that they could easily memorize the color mappings, which enabled them to perform faster in this condition. This lends support to our first hypothesis (*H1*). Thus, our study suggests that a Color Cue, which is always in the worker's FoV, is efficient and easy to use for identifying a small and memorizable set of helper locations. While other persistent indicators for the helper's location, such as arrows or a compass [41], share some qualities with the Color Cue, the latter may have an advantage in leveraging remembered information in addition to interpreting visual information in the world. We recommend the use of a Color Cue in industrial scenarios, in particular when the noise from machines makes Spatial Sound impractical to use.

However, some study participants experienced difficulties in perceiving the color of the cursor or distinguishing the colors of the virtual cameras with *Virtual Hand + Color Cue*. Particularly the colors pink and red sometimes looked quite similar. Due to the optical see-through HMD, these perception issues were further exacerbated by lighting conditions in the environment. Color blindness may be an additional factor to consider. We therefore wish to highlight the importance of carefully choosing contrasting colors and designing a Color Cue that can be clearly distinguished, for example by also varying the cursor's shape and using symbolic representations for work areas.

Another design implication refers to the violation of depth cues when persistently displaying awareness cues that are *behind* an object. This choice to NOT occlude the Virtual Hand and virtual cameras was made consciously to ensure visibility of these cues, even when the worker's line of sight to the helper's location may be obstructed. For example, in our experimental setup, a scene camera was attached to the whiteboard in the middle of the room, providing a view of the whiteboard's front surface. A worker pointed out that it was confusing

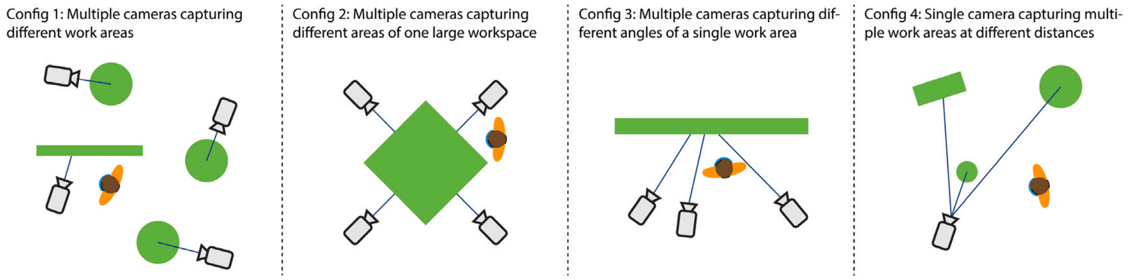

Config 1: Multiple cameras capturing different work areas
Config 2: Multiple cameras capturing different areas of one large workspace
Config 3: Multiple cameras capturing different angles of a single work area
Config 4: Single camera capturing multiple work areas at different distances

Fig. 8. Config 1 depicts the camera configuration evaluated in the presented study. Configs 2–4 show alternative examples of camera configurations. Notably, Config 3 involves capturing only one work area, and Config 4 relies on the use of only a single camera.

to still see the Virtual Hand through the whiteboard when moving behind it. The lack of occlusion briefly gave the misleading impression that the helper was viewing the back side of the whiteboard. However, the worker was quickly able to correct this interpretation, based on knowledge of which work areas were covered by the scene cameras. While we argue that visualizing information about a thing or person that is occluded by a physical obstacle can be useful for maintaining workspace awareness, we recommend to represent occluded cues distinctly to support correct depth interpretation, e.g., using an "X-ray" mode as in [42].

6.2. Sensory modalities and attention

In our experiment, participants were given an artificial search task that allowed direct comparison of three different awareness cue combinations. In this controlled study, no other tasks competed for the participants' attention. This was desired to avoid confounding factors, but is not faithful to a real remote assistance scenario, where a local technician may be simultaneously engaged in the tasks of repairing a machine, monitoring its status, and communicating with the remote helper. Thus, their visual and/or auditory attention is divided between the physical domain task and maintaining awareness of the helper. According to the multiple resource model of Wickens [43], tasks can be performed better when distributed across sensory modalities, because they compete to a lesser degree for the same cognitive resources. This implies that a redundant use of awareness cues - e.g. simultaneous use of Spatial Sound and Color Cue - may be worth considering in a real remote assistance scenario. For example, if the worker's auditory attention is occupied during conversation with a remote helper, the Color Cue may be most useful for locating the helper. Vice versa, if the worker's visual attention is occupied during an assembly/repair task, it may be best to use Spatial Sound for conveying the helper's location. The work of Yang et al. [36] supports this in demonstrating that combination of a visual awareness cue (the helper's view frustum) with spatial sound improved the worker's feeling of social and spatial presence compared to using spatial sound alone. It even seems reasonable to assume that a combination of all three awareness cues (Virtual Hand, Spatial Sound and Color Cue) might be most advantageous, since it allows workers to resort to various strategies for locating the helper depending on their situation. However, we refrain from making assumptions about whether redundant cues could impact the worker's cognitive load. As future work it would be interesting to conduct a controlled study with the dual tasks of completing physical manipulations of task objects while maintaining awareness of the helper. This would allow us to better understand the benefits and challenges of simultaneous, redundant awareness cues and the user's reliance on visual and auditory modalities when awareness cues and physical domain tasks compete for the worker's attention.

It should be noted that our awareness cues are limited to the visual and auditory modalities, and including the haptic modality may provide obvious benefits over sound, especially if the objects in a search task are close to each other [37]. Nevertheless, this supposition remains for future work to verify.

6.3. Camera configurations

For the experimental search task we distributed the scene cameras such that each covered its own separate work area within a large space. However, our system also allows the exploration of other scene camera configurations. Some examples are illustrated in Fig. 8. We chose the current configuration (*Config. 1* in Fig. 8), since it facilitated that the worker had to search for and navigate to the helper's location, enabling us to better study the effectiveness of our awareness cues in a large space. Furthermore, this configuration reflects a potential workspace layout in an actual remote assistance scenario in industry, e.g., when troubleshooting a large manufacturing machine that consists of multiple areas of interest with both vertical and horizontal surfaces. *Config. 2* in Fig. 8 similarly describes a large workspace, with each camera covering a distinct area on a large object, e.g., a machine. However, all cameras are arranged in an outside-in manner, pointing at different sides of the object. While such a scenario would have supported a comparison of our awareness cues, the regular and homogeneous layout was deemed an unreasonable simplification. Conversely, *Config. 3* would not have been useful for our evaluation of awareness cues. In such a configuration, a change of the helper's view point leads to a change in perspective, but the area of attention remains the same, thereby strongly reducing the need for awareness cues. Finally, *Config. 4* shows a single camera on one side of a room, providing the helper with a wide view of multiple work areas at various distances. This scenario serves nicely to discuss some limitations of Spatial Sound and the Color Cue, since these cues only provide information about the helper's location in terms of the viewed camera. Thus, there can be no clear mapping between the helper's location (camera position) provided by the cues and the helper's focus of attention. Nevertheless, Color Cue and Spatial Sound could be modified to be useful in such a scenario. For instance, Spatial Sound could be implemented as a spatial 3D sound emanating from the position of the Virtual Hand, by playing it through the HMD instead of the scene cameras. Furthermore, the Color Cue could map to particular areas in the task space instead of mapping to cameras, for instance by presenting cues as colored virtual arrows hovering over each area (see Fig. 9). This would require, however, that the areas can be identified from one camera view — either predefined by a user or identified by a computer vision algorithm.

6.4. Calibration of multi-camera remote assistance systems

Calibration processes for multi-camera AR remote assistance commonly require that camera devices track the pose of one central marker on the physical surface that is to be augmented [4,6,44]. This marker must therefore be visible to all scene cameras, enforcing an outside-in camera configuration with overlapping FoVs. This approach may cause difficulties, for example during maintenance of a large injection molding machine, where there may be no obvious way to attach a central marker to the machine, while ensuring that multiple disconnected areas can be observed. Further, using multiple fixed cameras without

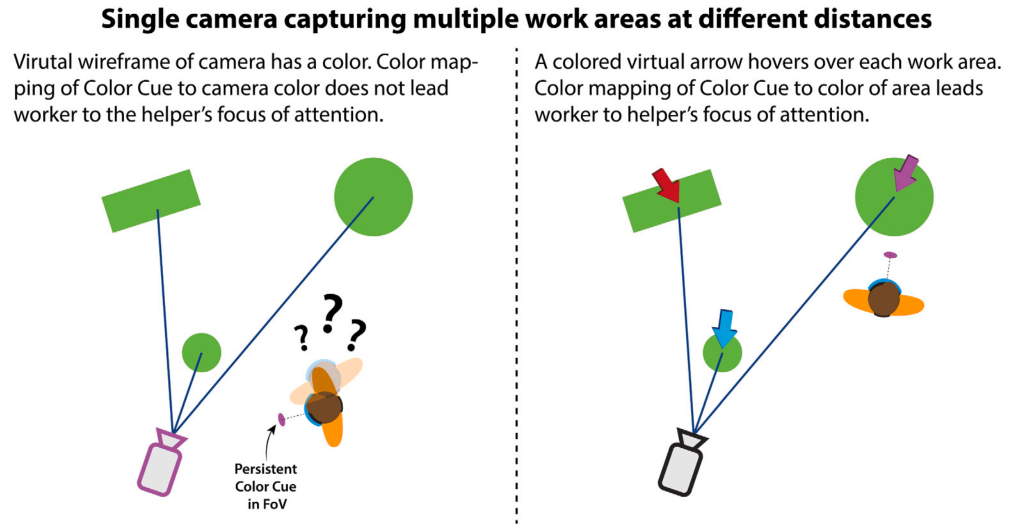

Single camera capturing multiple work areas at different distances
Virutal wireframe of camera has a color. Color mapping of Color Cue to camera color does not lead worker to the helper's focus of attention.
A colored virtual arrow hovers over each work area. Color mapping of Color Cue to color of area leads worker to helper's focus of attention.
? ? ?
Persistent Color Cue in FoV

Fig. 9. Improving Color Cue by mapping color of cursor to areas instead of to cameras.

the option to move them arguably reduces the usefulness of the multi-camera setup [18]. Finally, altering such multi-camera setups *during* remote assistance can be challenging, as maintaining the overlap of FoVs limits the possibilities for camera placement or requires moving markers around, and recalibration of the entire workspace can be time consuming.

To overcome these challenges, our approach involves a marker attached to each camera, either virtually on tablets/smartphones or physically on webcams (see Fig. 4). This enables the worker to add, move, and recalibrate cameras in the workspace in an ad-hoc manner, by simply re-scanning individual markers with the HMD. Our camera calibration process thereby differs from and extends on related work in two important respects: Firstly, it supports configurations where cameras cover distinct, non-overlapping areas, as the need for placing markers in the workspace and ensuring that they are within the camera's FoV is eliminated entirely. Secondly, our approach supports ad-hoc reconfiguration of the setup, when adding, moving, or removing individual cameras during remote assistance. Our work also presents an alternative to the approach demonstrated by Piumsomboon et al. [26], who track the pose of a movable 360-degree camera with an outside-in tracking system (HTC Vive). In comparison, our calibration process only relies on the inside-out tracking of the AR-HMD to keep track of camera poses and is thus not restricted to a space with an installed tracking system. Using *CueCam*, multiple cameras can flexibly be set up in any space that has suitable features for inside-out tracking.

However, off-loading the calibration process onto the HMD also has some disadvantages. First, a rigid transformation must be calculated for each marker and camera lens pair. A current limitation of our AR App is that these transformations (translations in our case) must be measured and input manually, *once* per camera, prior to calibration. Future work involves may address the intuitive definition of these transformations, e.g., by aligning a virtual camera lens with its physical counterpart through direct manipulation in an AR application. An additional limitation of our calibration process is the possible drift over time due to the inside-out tracking of the HMD. Over extended use, this causes the virtual camera poses to deviate from their physical counterparts and therefore the 3D interpretation of 2D annotations and visual awareness cues to become slightly offset from their true positions. However, the performance of inside-out tracking is ever improving and we plan to migrate the AR App to Hololens v2, to alleviate this issue in future work.

7. Limitations and future work

We wish to mention a number of additional limitations of the presented work, which have not yet been addressed. Most importantly, our findings relate to our specific experimental setup and our particular implementation of each awareness cue. Future observations may therefore differ for alternative designs or hardware.

First of all, visual search for the Virtual Hand was likely affected by the narrow FoV of the supported AR-HMD (Hololens version 1) [45]. This may have especially impacted search performance in the *Virtual Hand Only* condition. It remains for future work to explore whether a larger FoV improves the worker's awareness of visual cues. In addition, a better implementation of a guiding arrow, such as the one in [3], may have led to an improved visual search.

In respect to the design of our Spatial Sound, a participant suggested improvement by adding a uniquely identifiable sound to each camera device, thus enabling the worker to use both recognition and recall, as in *Virtual Hand + Color Cue*. Further, one participant commented that the continuous sound of Spatial Sound might annoy and distract the worker in a real-word setting. In a real industrial application requiring prolonged use, more attention needs to be paid to the audio design. The continuous sound might also be avoided, by allowing the worker to query the system about the location of the helper, for instance by issuing a voice command “Locate viewpoint” upon which a sound is played once at the location of the helper. This is similar to a co-located scenario, where one collaborator yells out “Where are you?” upon which the other collaborator responds, “Over here!”. While *CueCam* implicitly supports workspace awareness through voice communication with localized sound (i.e., the worker can orient himself based on the direction from which the remote helper's voice is heard), the study presented in this paper involved only artificial sound cues and no voice communication. This was done to avoid confounding sound cues, since study participants were physically co-located in the same room and merely separated by a screen. This aspect warrants further exploration in a study with non-co-located users.

Furthermore, in our system the worker perceives the helper as teleporting from one work area to another when switching between scene cameras. Thus, the direction and position of AR awareness cues changes abruptly, which makes the Virtual Hand particularly difficult to use for quick location of the helper. In contrast, in a co-located collaborative scenario, a worker walking from one work area to another continuously provides information about their location. Even when they have left a co-workers FoV, the direction in which they walked and the layout of the facility will support locating them. Future work could aim at recreating such continuous information about the helper's change in location, e.g., by presenting an animation of a virtual character navigating from one area to another or by displaying movement trajectories [46], and by playing spatial sounds through the worker's AR HMD instead of the camera devices. This would allow systematic comparison of teleportation and animation tweening of awareness cues. Directional arrows, as proposed by [22,47], might further support the visual search for the Virtual Hand. It remains for future work to explore such arrows as awareness cues in comparison to the ones proposed in this paper.

Lastly, we made the deliberate choice to inform study participants about the number of visited areas after each condition, because we wished to create a gamified, competitive experience that would encourage them to perform their best. This may introduce a confounding variable when evaluating preferences, since participants might be inclined to favor the condition in which they performed best. Yet, some participants performed best with *Virtual Hand + Color Cue* but preferred *Virtual Hand + Spatial Sound*, indicating that objective performance measurements and subjective preferences do not always align.

We aim to explore the usefulness of add-hoc reconfiguration of scene cameras during AR remote assistance in a future field study in the manufacturing industry. We thereby hope to gain a better understanding about the number of cameras and configurations needed for particular workspaces, and how camera work unfolds in real use cases. This research area is yet under-explored and important for improving remote assistance in large workspaces.

8. Conclusion

This paper presents a multi-camera remote assistance system that aims to support shared workspace awareness between the local worker and remote helper. The worker places multiple cameras in the workspace, enabling the helper to independently explore multiple perspectives. To allow for ad-hoc reconfiguration of the cameras by the worker (e.g., upon request of the helper during the task), we developed a novel AR-based camera calibration procedure. Further, to help increase the worker's awareness of the disembodied helper's current viewpoint and focus throughout the collaboration, we propose three AR awareness cues. In particular, we implemented a Virtual Hand, indicating the location of the camera viewed by the helper and her assumed focus, as well as a Spatial Sound cue and a Color Cue, which allow the worker to locate the helper's point of view using different strategies. We propose to apply the latter cues in combination with Virtual Hand, and evaluated these in a user study comparing three awareness cue combinations. While we find that all implemented cues were effective, the combination of *Virtual Hand + Color Cue* was superior to *Virtual Hand + Spatial Sound* and *Virtual Hand Only* regarding performance and preference, and both *Virtual Hand + Color Cue* and *Virtual Hand + Spatial Sound* were more performant than *Virtual Hand Only*. With these new insights on AR awareness cues and our proposed ad-hoc calibration method, we aim to contribute to the design of future remote assistance solutions.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Appendix A. Supplementary data

Supplementary material related to this article can be found online at https://doi.org/10.1016/j.jvcir.2022.103655.

References

[1] J. Lanir, R. Stone, B. Cohen, P. Gurevich, Ownership and control of point of view in remote assistance, in: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, CHI '13, ACM, New York, NY, USA, 2013, pp. 2243–2252, http://dx.doi.org/10.1145/2470654.2481309.

[2] M. Tait, M. Billinghurst, The effect of view independence in a collaborative AR system, Comput. Support. Cooper. Work 24 (6) (2015) 563–589, http://dx.doi.org/10.1007/s10606-015-9231-8, URL:https://link.springer.com/article/10.1007/s10606-015-9231-8.

[3] G.A. Lee, T. Teo, S. Kim, M. Billinghurst, A user study on mr remote collaboration using live 360 video, in: 2018 IEEE International Symposium on Mixed and Augmented Reality (ISMAR), 2018, pp. 153–164, http://dx.doi.org/10.1109/ISMAR.2018.00051.

[4] S. Kim, M. Billinghurst, G. Lee, The effect of collaboration styles and view independence on video-mediated remote collaboration, Comput. Supported Coop. Work 27 (3–6) (2018) 569–607, http://dx.doi.org/10.1007/s10606-018-9324-2.

[5] W.W. Gaver, A. Sellen, C. Heath, P. Luff, One is not enough: Multiple views in a media space, in: Proceedings of the INTERACT '93 and CHI '93 Conference on Human Factors in Computing Systems, CHI '93, ACM, New York, NY, USA, 1993, pp. 335–341, http://dx.doi.org/10.1145/169059.169268.

[6] M. Adcock, S. Anderson, B. Thomas, RemoteFusion: Real time depth camera fusion for remote collaboration on physical tasks, in: Proceedings of the 12th ACM SIGGRAPH International Conference on Virtual-Reality Continuum and Its Applications in Industry, VRCAI '13, ACM, New York, NY, USA, 2013, pp. 235–242, http://dx.doi.org/10.1145/2534329.2534331.

[7] T.A. Rasmussen, W. Huang, SceneCam: Improving multi-camera remote collaboration using augmented reality, in: 2019 IEEE International Symposium on Mixed and Augmented Reality Adjunct (ISMAR-Adjunct), 2019, pp. 28–33, http://dx.doi.org/10.1109/ISMAR-Adjunct.2019.00023.

[8] S.R. Fussell, L.D. Setlock, J. Yang, J. Ou, E. Mauer, A.D.I. Kramer, Gestures over video streams to support remote collaboration on physical tasks, Human Comput. Interact. 19 (3) (2004) 273–309, http://dx.doi.org/10.1207/s15327051hci1903_3.

[9] W. Huang, S. Kim, M. Billinghurst, L. Alem, Sharing hand gesture and sketch cues in remote collaboration, J. Vis. Commun. Image Represent. 58 (2019) 428–438, http://dx.doi.org/10.1016/j.jvcir.2018.12.010, URL:https://www.sciencedirect.com/science/article/pii/S1047320318303365.

[10] W. Huang, L. Alem, HandsinAir: A wearable system for remote collaboration on physical tasks, in: Proceedings of the 2013 Conference on Computer Supported Cooperative Work Companion, CSCW '13, ACM, New York, NY, USA, 2013, pp. 153–156, http://dx.doi.org/10.1145/2441955.2441994.

[11] W. Huang, L. Alem, F. Tecchia, H.B.-L. Duh, Augmented 3D hands: a gesture-based mixed reality system for distributed collaboration, J. Multi. User Interf. 12 (2) (2018) 77–89, http://dx.doi.org/10.1007/s12193-017-0250-2.

[12] S.R. Fussell, L.D. Setlock, R.E. Kraut, Effects of head-mounted and scene-oriented video systems on remote collaboration on physical tasks, in: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, CHI '03, ACM, New York, NY, USA, 2003, pp. 513–520, http://dx.doi.org/10.1145/642611.642701.

[13] K. Yamaashi, J.R. Cooperstock, T. Narine, W. Buxton, Beating the limitations of camera-monitor mediated telepresence with extra eyes, in: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, CHI '96, ACM, New York, NY, USA, 1996, pp. 50–57, http://dx.doi.org/10.1145/238386.238402.

[14] D. Palmer, M. Adcock, J. Smith, M. Hutchins, C. Gunn, D. Stevenson, K. Taylor, Annotating with light for remote guidance, in: Proceedings of the 19th Australasian Conference on Computer-Human Interaction: Entertaining User Interfaces, ACM, pp. 103–110.

[15] A. Ranjan, J.P. Birnholtz, R. Balakrishnan, Dynamic shared visual spaces: Experimenting with automatic camera control in a remote repair task, in: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, CHI '07, ACM, New York, NY, USA, 2007, pp. 1177–1186, http://dx.doi.org/10.1145/1240624.1240802.

[16] J. Norris, H. Schnädelbach, G. Qiu, CamBlend: An object focused collaboration tool, in: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, CHI '12, ACM, New York, NY, USA, 2012, pp. 627–636, http://dx.doi.org/10.1145/2207676.2207765.

[17] J. Norris, H.M. Schnädelbach, P.K. Luff, Putting things in focus: Establishing co-orientation through video in context, in: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, CHI '13, ACM, New York, NY, USA, 2013, pp. 1329–1338, http://dx.doi.org/10.1145/2470654.2466174.

[18] D. Aschenbrenner, F. Sittner, M. Fritscher, M. Krauß, K. Schilling, Cooperative remote repair task in an active production line for industrial internet telemaintenance**funded by the bavarian ministry of economic affairs, infrastructure, transport and technology in its R&D program 'information and communication technology'., IFAC-PapersOnLine 49 (30) (2016) 18–23, http://dx.doi.org/10.1016/j.ifacol.2016.11.116, URL:https://www.sciencedirect.com/science/article/pii/S240589631632554X.

[19] S. Orts-Escolano, C. Rhemann, S. Fanello, W. Chang, A. Kowdle, Y. Degtyarev, D. Kim, P.L. Davidson, S. Khamis, M. Dou, V. Tankovich, C. Loop, Q. Cai, P.A. Chou, S. Mennicken, J. Valentin, V. Pradeep, S. Wang, S.B. Kang, P. Kohli, Y. Lutchyn, C. Keskin, S. Izadi, Holoportation: Virtual 3D teleportation in real-time, in: Proceedings of the 29th Annual Symposium on User Interface Software and Technology, UIST '16, ACM, New York, NY, USA, 2016, pp. 741–754, http://dx.doi.org/10.1145/2984511.2984517, event-place: Tokyo, Japan.

[20] M. Joachimczak, J. Liu, H. Ando, Real-time mixed-reality telepresence via 3D reconstruction with HoloLens and commodity depth sensors, in: Proceedings of the 19th ACM International Conference on Multimodal Interaction - ICMI 2017, ACM Press, Glasgow, UK, 2017, pp. 514–515, http://dx.doi.org/10.1145/3136755.3143031, URL:http://dl.acm.org/citation.cfm?doid=3136755.3143031.

[21] T. Teo, L. Lawrence, G.A. Lee, M. Billinghurst, M. Adcock, Mixed reality remote collaboration combining 360 video and 3D reconstruction, in: Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems, CHI '19, Association for Computing Machinery, Glasgow, Scotland Uk, 2019, pp. 1–14, http://dx.doi.org/10.1145/3290605.3300431.

[22] G.A. Lee, T. Teo, S. Kim, M. Billinghurst, Sharedsphere: MR collaboration through shared live panorama, in: SIGGRAPH Asia 2017 Emerging Technologies on - SA '17, ACM Press, Bangkok, Thailand, 2017, pp. 1–2, http://dx.doi.org/10.1145/3132818.3132827, URL:http://dl.acm.org/citation.cfm?doid=3132818.3132827.

[23] M. Speicher, J. Cao, A. Yu, H. Zhang, M. Nebeling, 360Anywhere: Mobile Ad-hoc collaboration in any environment using 360 video and augmented reality, Proceedings of the ACM on Human-Computer Interaction 2 (EICS) (2018) 1–20, http://dx.doi.org/10.1145/3229091, URL:http://dl.acm.org/citation.cfm?doid=3233739.3229091.

[24] T. Teo, G.A. Lee, M. Billinghurst, M. Adcock, Hand gestures and visual annotation in live 360 panorama-based mixed reality remote collaboration, in: Proceedings of the 30th Australian Conference on Computer-Human Interaction, OzCHI '18, Association for Computing Machinery, Melbourne, Australia, 2018, pp. 406–410, http://dx.doi.org/10.1145/3292147.3292200.

[25] J. Kangas, A. Sand, T. Jokela, P. Piippo, P. Eskolin, M. Salmimaa, R. Raisamo, Remote expert for assistance in a physical operational task, in: Extended Abstracts of the 2018 CHI Conference on Human Factors in Computing Systems, in: CHI EA '18, ACM, New York, NY, USA, 2018, pp. LBW127:1–LBW127:6, http://dx.doi.org/10.1145/3170427.3188598, event-place: Montreal QC, Canada.

[26] T. Piumsomboon, G.A. Lee, A. Irlitti, B. Ens, B.H. Thomas, M. Billinghurst, On the shoulder of the giant: A multi-scale mixed reality collaboration with 360 video sharing and tangible interaction, in: Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems, CHI '19, Association for Computing Machinery, Glasgow, Scotland Uk, 2019, pp. 1–17, http://dx.doi.org/10.1145/3290605.3300458.

[27] T. Teo, A. F. Hayati, G. A. Lee, M. Billinghurst, M. Adcock, A technique for mixed reality remote collaboration using 360 panoramas in 3D reconstructed scenes, in: 25th ACM Symposium on Virtual Reality Software and Technology, VRST '19, Association for Computing Machinery, Parramatta, NSW, Australia, 2019, pp. 1–11, http://dx.doi.org/10.1145/3359996.3364238.

[28] T.A. Rasmussen, K. Grø nbæk, Tailorable remote assistance with RemoteAssistKit: A study of and design response to remote assistance in the manufacturing industry, in: International Conference on Collaboration and Technology, Springer, 2019, pp. 80–95, http://dx.doi.org/10.1007/978-3-030-28011-6_6.

[29] L. Gao, H. Bai, R. Lindeman, M. Billinghurst, Static local environment capturing and sharing for MR remote collaboration, in: SIGGRAPH Asia 2017 Mobile Graphics & Interactive Applications, SA '17, ACM, New York, NY, USA, 2017, pp. 17:1–17:6, http://dx.doi.org/10.1145/3132787.3139204.

[30] T. Piumsomboon, A. Day, B. Ens, Y. Lee, G.A. Lee, M. Billinghurst, Exploring enhancements for remote mixed reality collaboration, in: SA '17, 2017, http://dx.doi.org/10.1145/3132787.3139200.

[31] T. Piumsomboon, A. Dey, B. Ens, G. Lee, M. Billinghurst, The effects of sharing awareness cues in collaborative mixed reality, Frontiers in Robotics and AI 6 (2019) http://dx.doi.org/10.3389/frobt.2019.00005, URL:https://www.frontiersin.org/articles/10.3389/frobt.2019.00005/full, Publisher: Frontiers.

[32] P. Mohr, S. Mori, T. Langlotz, B.H. Thomas, D. Schmalstieg, D. Kalkofen, Mixed reality light fields for interactive remote assistance, in: Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems, CHI '20, Association for Computing Machinery, New York, NY, USA, 2020, pp. 1–12, http://dx.doi.org/10.1145/3313831.3376289.

[33] C. Gutwin, S. Greenberg, A descriptive framework of workspace awareness for real-time groupware, Comput. Supported Coop. Work 11 (3) (2002) 411–446, http://dx.doi.org/10.1023/A:1021271517844.

[34] M.R. Endsley, Toward a theory of situation awareness in dynamic systems, Human Factors J. Hum. Factors Ergon. Soc. 37 (1) (1995) 32–64, http://dx.doi.org/10.1518/001872095779049543, URL:http://journals.sagepub.com/doi/10.1518/001872095779049543.

[35] R.S. Sodhi, B.R. Jones, D. Forsyth, B.P. Bailey, G. Maciocci, Bethere: 3D mobile collaboration with spatial input, in: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, CHI '13, ACM, New York, NY, USA, 2013, pp. 179–188, http://dx.doi.org/10.1145/2470654.2470679.

[36] J. Yang, P. Sasikumar, H. Bai, A. Barde, G. Sörös, M. Billinghurst, The effects of spatial auditory and visual cues on mixed reality remote collaboration, J. Multi. User Interfac. 14 (4) (2020) 337–352, http://dx.doi.org/10.1007/s12193-020-00331-1.

[37] S. Günther, S. Kratz, D. Avrahami, M. Mühlhäuser, Exploring audio, visual, and tactile cues for synchronous remote assistance, in: Proceedings of the 11th PErvasive Technologies Related To Assistive Environments Conference on - PETRA '18, ACM Press, Corfu, Greece, 2018, pp. 339–344, http://dx.doi.org/10.1145/3197768.3201568, URL:http://dl.acm.org/citation.cfm?doid=3197768.3201568.

[38] S. Kim, G. Lee, W. Huang, H. Kim, W. Woo, M. Billinghurst, Evaluating the combination of visual communication cues for HMD-based mixed reality remote collaboration, in: Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems, CHI '19, Association for Computing Machinery, New York, NY, USA, 2019, pp. 1–13, http://dx.doi.org/10.1145/3290605.3300403.

[39] B. Nuernberger, K. Lien, T. Höllerer, M. Turk, Interpreting 2D gesture annotations in 3D augmented reality, in: 2016 IEEE Symposium on 3D User Interfaces (3DUI), 2016, pp. 149–158, http://dx.doi.org/10.1109/3DUI.2016.7460046.

[40] S. Seinfeld, T. Feuchtner, A. Maselli, J. Müller, User representations in human-computer interaction, Hum. Comput. Interact. (2020) 1–39, http://dx.doi.org/10.1080/07370024.2020.1724790.

[41] P.A. Olin, A.M. Issa, T. Feuchtner, K. Grø nbæk, Designing for heterogeneous cross-device collaboration and social interaction in virtual reality, in: 32nd Australian Conference on Human-Computer Interaction, 2020, pp. 112–127, http://dx.doi.org/10.1145/3441000.3441070.

[42] B. Avery, C. Sandor, B.H. Thomas, Improving spatial perception for augmented reality X-ray vision, in: 2009 IEEE Virtual Reality Conference, 2009, pp. 79–82, http://dx.doi.org/10.1109/VR.2009.4811002, ISSN: 2375-5334.

[43] C.D. Wickens, Multiple resources and performance prediction, Theoret. Iss. Ergon. Sci. 3 (2) (2002) 159–177, http://dx.doi.org/10.1080/14639220210123806.

[44] O. Fakourfar, K. Ta, R. Tang, S. Bateman, A. Tang, Stabilized annotations for mobile remote assistance, in: Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems, CHI '16, ACM, New York, NY, USA, 2016, pp. 1548–1560, http://dx.doi.org/10.1145/2858036.2858171.

[45] C. Trepkowski, D. Eibich, J. Maiero, A. Marquardt, E. Kruijff, S. Feiner, The effect of narrow field of view and information density on visual search performance in augmented reality, in: 2019 IEEE Conference on Virtual Reality and 3D User Interfaces, VR, IEEE, 2019, pp. 575–584.

[46] H.P. Klemen Lilija, K. Hornbk, Who put that there? Temporal navigation of spatial recordings by direct manipulation, in: CHI ACM Conference on Human Factors in Computing Systems, Association for Computing Machinery, 2020.

[47] S. Gauglitz, B. Nuernberger, M. Turk, T. Höllerer, World-stabilized annotations and virtual scene navigation for remote collaboration, in: Proceedings of the 27th Annual ACM Symposium on User Interface Software and Technology, UIST '14, ACM, New York, NY, USA, 2014, pp. 449–459, http://dx.doi.org/10.1145/2642918.2647372.