A Conceptual Model and Taxonomy for Collaborative Augmented Reality

Bernardo Marques, Samuel Silva, João Alves, Tiago Araújo, Paulo Dias, and Beatriz Sousa Santos, *Senior Member, IEEE*

Abstract—To support the nuances of collaborative work, many researchers have been exploring the field of Augmented Reality (AR), aiming to assist in co-located or remote scenarios. Solutions using AR allow taking advantage from seamless integration of virtual objects and real-world objects, thus providing collaborators with a shared understanding or common ground environment. However, most of the research efforts, so far, have been devoted to experiment with technology and mature methods to support its design and development. Therefore, it is now time to understand where the field stands and how well can it address collaborative work with AR, to better characterize and evaluate the collaboration process. In this article, we perform an analysis of the different dimensions that should be taken into account when analysing the contributions of AR to the collaborative work effort. Then, we bring these dimensions forward into a conceptual framework and propose an extended human-centered taxonomy for the categorization of the main features of Collaborative AR. Our goal is to foster harmonization of perspectives for the field, which may help create a common ground for systematization and discussion. We hope to influence and improve how research in this field is reported by providing a structured list of the defining characteristics. Finally, some examples of the use of the taxonomy are presented to show how it can serve to gather information for characterizing AR-supported collaborative work, and illustrate its potential as the grounds to elicit further studies.

Index Terms—Collaboration, augmented reality, conceptual model, taxonomy, human-centered, systematization

1 Introduction

Collaboration can be described as the process of joint and interdependent activities between co-located or remote collaborators performed to achieve a common goal [1], [2], [3], [4], [5], [6], [7]. Collaboration integrates high levels of characteristics by focusing on different types of collaborators, common tasks and in encompassing dynamic environments with contextual data. The resources required to address collaboration have been growing in terms of scale, complexity, and interdisciplinarity entailing not only the mastery of multiple domains of knowledge, but also a strong level of proficiency in each [8]. Therefore, ensuring the conditions to support collaboration is paramount [4], [5], [9].

The field of Computer-Supported Cooperative Work (CSCW) has been concerned with understanding and designing solutions to support collaboration, aiming to enable communication, cooperation, assistance, training, learning as well as knowledge sharing between collaborators [9], [10], [11]. As the field of Augmented Reality (AR) matured, researchers started to explore how it can provide a common ground, similar to their understanding of the physical space, i.e., serve as a basis for situation mapping, allowing identification of issues, and making assumptions and beliefs visible [4], [9], [12], [13], [14], [15], [16], [17], [18]. Augmented Reality is well suited for human-to-human interactions, since it allows overlaying responsive computer-generated information on top of the real-world environment, resulting in the creation of solutions that combine the advantages of virtual environments and the possibility for seamless interaction with the real-world objects and other collaborators [19], [20], [21], [22].

In the past decade researchers have devoted their efforts to experiment with technology and mature methods for Collaborative AR [9], [23], [24], [25] and it is now time to understand where do we stand and how well can we address the domain of collaborative work with AR. In this context, there are two aspects that need to be tackled to obtain a proper perspective: (1) what does it take to address the question at hand, e.g., which dimensions need to be considered (i.e., what needs to be done); and (2) how is existing research tackling each of these dimensions (i.e., how it is done). In this regard, literature reviews can help us understand which research can provide an answer to a specific research question or problem [26], [27] and, particularly for areas with an already high level of maturity, existing work helps to identify enough of these two aspects. However, there are cases for which the literature addressing a particular research question does not provide enough information to understand if all relevant dimensions of the problem are being covered. Therefore, performing a review without a first effort at identifying these dimensions can provide limited

- *Bernardo Marques, Samuel Silva, João Alves, Paulo Dias, and Beatriz Sousa Santos are with IEETA, DETI, University of Aveiro, 3810-193 Aveiro, Portugal. E-mail: {bernardo.marques, sss, jbga, paulo.dias, bss}@ua.pt.*
- *Tiago Araújo is with the PPGCC, Federal University of Pará, Belém 66075-110, Brasil. E-mail: tiagoaraujo@ufpa.br.*

Manuscript received 19 Dec. 2020; revised 9 July 2021; accepted 28 July 2021. Date of publication 4 Aug. 2021; date of current version 27 Oct. 2022. (Corresponding author: Bernardo Marques.) Recommended for acceptance by S. Zollmann. Digital Object Identifier no. 10.1109/TVCG.2021.3101545

insight, since it might precisely miss those aspects that, albeit important, are still not addressed by existing research. We are not only seeking what the community has achieved, but also if it fully addresses all aspects to solve domain-specific problems.

In this line of thought, before performing a literature review on collaborative work supported by AR, perhaps wrongly assuming that every relevant aspect has already been addressed, we need to take a step back to enable a wide perspective that goes outside the AR research boundaries to also encompass the context it strives to support collaboration. This entails going beyond Collaborative AR literature, considering other domains that may be relevant to characterize the collaborative effort, to identify which aspects (dimensions) should be taken into account when we move from asking what existing systems can do to understanding what they were able to do in particular contexts, i.e., the value of the solutions they provide. The importance of such effort is, in our perspective, twofold: first, it allows gathering a structured insight on the defining dimensions of Collaborative AR, fostering a more detailed understanding of the field; and, second, by doing so, contributes to support research that places the solutions in close relation with the collaborative context they address and reflects on the extent of its contributions. And, as the Collaborative AR community, having matured technologies and methods [9], approaches domain experts to address their collaborative needs, this latter aspect is paramount to ensure that the research adds to the body of knowledge and provides enough context and evidence to enable a transparent account [28] and transferability [29].

To this end, it is important to materialize the devised dimensions of Collaborative AR into a conceptual framework and taxonomy that might foster a harmonization of perspectives for the field creating a common ground for systematization and discussion of past, present, and future works [27]. The proposal of a taxonomy should influence and improve how research is reported [30] by providing a structure and, in a way, a check-list to the defining characteristics that need to be clarified. A more systematic reporting can, in turn, lead to a community setting that enables easier building on existing research. By gathering dimensions that both cover the collaborative context and the AR solution, the taxonomy may also foster going beyond the description of the methods and into the methodology [31], i.e., how the research moved from the problem to the choice of the methods. Additionally, a taxonomy should also improve the awareness of researchers about different dimensions of the contexts they target. In this regard, the work presented here contributes to research on Collaborative AR by:

- providing an explicit consideration of the work in Collaborative AR in tight relation with several characteristics identified as defining collaborative work;
- performing an analysis of different dimensions to be considered when developing collaborative AR-based systems;
- proposing an extended human-centered taxonomy for the categorization of the main features of Collaborative AR stemming from the identified dimensions.

The remainder of this document is organized as follows. Section 2 introduces essential concepts of Collaborative AR and overviews existing categorization efforts. Section 3 explains the methodology adopted, defines a conceptual model and proposes a taxonomy, as well as illustrations of the use of the taxonomy through a visualization diagram. Finally, concluding remarks and future research opportunities are drawn in Section 5.

2 Background and Related Work

In this section we describe how AR can support collaborative work according to existing works, present essential concepts and characteristics and report on existing categorization efforts.

2.1 Collaborative Technologies and AR

The field of CSCW has long been concerned with understanding and designing solutions to support collaboration, resulting in theories that have influenced the design of collaborative technologies [10]. From this research, Johansen [32] proposed a categorization for CSCW systems based on time and spatial location. According to the time-space matrix, CSCW systems can be organized depending on when collaborators work together at the same or different times (synchronous versus asynchronous collaboration) and the physical arrangement of the work place (at the same location or in different places). The time-space matrix organization is still, up until today, a cornerstone of the categorization of software tools for collaborative activity [9]. Moreover, two essential elements that CSCW solutions need to address to support collaborative work have emerged: 1) enable mutual awareness about the collaborators in the workspace, as well as about the tasks they are performing; and 2) understand and articulate how information is used to support collaboration [9], [33]. To support interaction between multiple collaborators, CSCW requires the creation of Groupware [34], which can be defined as: ”a computer systems that support groups of people engaged in a common task (or goal) and that provide an interface to a shared environment” [35].

Researchers from the CSCW domain have been investigating the use of AR to provide collaborators with a shared understanding or common ground environment [36]. Augmented Reality (AR) allows overlaying responsive computer-generated information on top of the real-world combining real and virtual content, real time interaction, and 3D registration with real content [37]. AR has the potential to create new forms of visualization of information and interaction, by adding new content (including 3D models, animations, images and video) aligned with the physical world. Moreover, the information displayed is context-sensitive, which means that it depends on the observed objects [13], [38]. Nowadays, AR is considered an interdisciplinary area, transcending boundaries between concepts (traditionally not thought to be related to AR) ranging from Optics, Computer Graphics, Computer Vision, Human-Computer Interaction, Ergonomics, to other concepts like ethics, art, philosophy, and social sciences [19], [39]. Through the years, numerous studies have proven that AR has the potential to reduce cognitive load, task duration and number of errors, improving learning activities, and facilitate more effective training over most current practices [40], [41], [42], [43], [44], [45]. AR solutions are often isolated and applicable in static work environment [44], being commonly used to enhance a single user perception of reality. However, an area with, perhaps, the greatest potential for AR is collaborative work [12], [46], [47].

The concept of Collaborative AR can be described as a system where: "multiple collaborators share the same augmented environment locally or remotely [48] and which enables knowledge transfer between different them [49]". Moreover, the augmentation of the real environment of one collaborator occurs through the actions of other collaborators and does not merely rely on information previously stored in the computer [34]. By creating a common ground environment, it can enhance alertness, awareness, and understanding of the situation in the form of visual communication cues (e.g., pointers, annotations, hand gestures, among others) to enhance a scene as it is captured by a collaborator and provide real-time spatial information about objects, events and areas of interest [9], [50], [51]. Although Collaborative AR research is still in its infancy, it has the potential to support effective knowledge transfer between multiple collaborators allowing them to interact with each other in a context-sensitive manner. These benefits may result in significant cost-savings and better service for customers [16], [20], [49]. Co-located AR solutions can be used to elicit and promote the performance of specific tasks between a co-located group of users. Such solutions allow to interact with shared AR content as naturally as with physical objects while maintaining important natural face-to-face communication cues [46]. While early work focused mainly on co-located scenarios, there is a great interest for remote scenarios as technological limitations are being overcome [9], [15]. Remote AR solutions can be used to empower remote collaboration that may be needed in very specific situations which require know how and additional information from professionals unavailable on-site [50], [52]. Remote users can use AR solutions regardless of their localization to guide local users, providing real-time spatial information, highlighting specific areas, or sharing annotations [15], [51], [53].

2.2 Categorization Efforts

Throughout the years, several categorizations have been proposed for AR and collaborative technologies. The ability to draw inferences is a critical condition for a useful categorization. Taxonomies are a good example, allowing to structure the knowledge of a field, understand the relationships among concepts, analyze complex domains, and provide relevant input to the development of theories [54]. Nickerson *et al.* [55] present a literature survey in several disciplines, and discuss thoroughly the problem of taxonomy development. The authors proposed the following qualitative attributes for the creation of a useful taxonomy: it must be `concise, robust, comprehensive` and `extensible`. Likewise, a good taxonomy must also be `explanatory, and not descriptive`: it must contain dimensions and characteristics that do not describe in detail specific objects of interest, but rather provide useful explanations of their nature, allowing the taxonomy to be useful for several purposes.

2.2.1 Augmented Reality

Several taxonomies have been proposed in the field of AR, starting with the one by Milgram and Kishino (1994) [56], which performs a categorization based on the types of visual displays used. The taxonomies by Mackay (1998) [57], Suomela and Lehikoinen (2004) [58], Lindeman and Noma (2007) [59], Bráz and Pereira (2008) [60], Toennis and Plecher (2011) [61], and Hugues and Fuchs (2011) [62] are fairly general, not addressing any particular type of AR technique or area of application. These are summarized and organized by Normand *et al.* [63] into four different types:

- *Technique oriented* refers to taxonomies that group concepts related with the system environment knowledge, realistic representations, centricity of the type of display (egocentric or exocentric), congruency of control-display mapping and sense of presence.
- *User-centered* encompasses taxonomies that categorize stimuli based on the insertion point: the real world (in Spatial AR) or the virtual world (when the content is only visible through a device). Other taxonomies categorize AR application based on other users properties: mobility (stationary/ mobile), number of users and space (co-located/ remote).
- *Interaction-centered* taxonomies focus on interaction aspects, such as the target of augmentation (user or physical object), input and output devices, system and persons, and connections between the system and the real world.
- *Information-centered* taxonomies focus on concepts related to the data available: model dimensionality (ranging from 0D to 3D), viewpoints (first or third person), temporality (continuous or discrete presentation of information), registration and referencing (objects that present information about other objects in the environment).

2.2.2 Collaborative Augmented Reality

Regarding Collaborative AR, some categorization efforts can be found in literature. Early work by Benford *et al.* [64] proposed an interaction-centered taxonomy for classification of MR approaches according to the shared spaces based on three dimensions: `transportation` - the extent to which a group and objects leave behind their on-site space and enter into a new remote space in order to meet with others; `artificiality` - the extent to which a space is either synthetic or is based on the physical world; and `spatiality` - the level of support for fundamental physical spatial properties such as containment, topology, distance, orientation, and movement.

Billinghurst *et al.* [65] defined the following characteristics as relevant for a Collaborative AR environment: `virtuality` - virtual objects with no direct relation with the real environment that can be seen and examined in AR; `augmentation` - virtual objects (e.g., annotations, visual guides, etc.), directly related to real objects existing in the scene; `cooperation` - possibility for multiple users to see each other and cooperate; `independence` - each user controls their own independent view; `individuality` - the displayed data/representation might be different for each user.

Then, Wang *et al.* [34] focused into the design of effective AR systems to mediate human-human collaboration for shared production tasks in the Construction sector. Three dimensions were identified for categorizing AR systems in such contexts: `mobility` - user's location and orientation, divided in small local area environments and large distances; `number of users` - single-user AR and multi-user AR systems; `space` - distance between users in a multi-user AR system, which can be classified as either co-located or distributed systems.

In addition, Brockmann *et al.* [66] provided a categorization for collaborative AR-applications, based on a literature review focused on six dimensions: `space`, `time`, `mobility`, `virtual content`, `role concept`, and `visualization hardware`. According to the authors, their user-centered taxonomy "shall support the user in identifying the most appropriate collaborative AR-application fitting the respective communication and collaboration scenario".

In the same way, the research by Jalo *et al.* [49] reported the following characteristics for the development of Collaborative AR systems for the Industry sector: `dimensions` - depending on whether the collaboration happens synchronously or asynchronously and whether the users are located in the same place or not; `stakeholders` - collaboration inside a company, between companies or between a company and its customers; `type` - depends on the number of participants and can be divided into: one-on-one, one-on-many and many-on-many categories; `functionalities` - visual digital information, such as text, pictures, videos and models; `device` - collaboration using AR can happen though the use of a multitude of hardware; `senses` - all human senses can be used in AR.

Speicher *et al.* [67] also present the notion of mixed reality as a meeting of AR and VR users that are potentially physically separated. Although the authors emphasize the MR landscape is highly fragmented, a conceptual framework with seven dimensions was created to categorize MR applications in terms of `number of environments` - total of physical and virtual environments; `number of users` - users required for a certain type of activity; `level of immersion` - how immersed the user feels based on the digital content they perceive; `level of virtuality` - how much digital content the user perceives (whether or not restricted to a specific sense); `degree of interaction` - which can be divided into implicit (e.g., walking around a virtual object registered in space) and explicit (e.g., intentionally providing input to); `input` - refers to input besides explicit interaction, used to inform the experience, which can be anything sensors can track; `output` - considers output to one or more of the user's senses in order to change their perception.

Another example, is the work by Belen *et al.* [68], who performed a systematic review of the current state of collaborative MR technologies, published from 2013 to 2018. This review presented a high-level overview of collaborative MR influence across several research disciplines. A total of 259 papers have been categorized based on their `application areas`, `types of displays` used, `collaboration setups`, `user interaction and user experience` aspects.

Ens *et al.* [9] revisited collaboration through MR, taking into account the evolution of groupware. The authors reviewed investigated how common taxonomies and frameworks in CSCW and MR research could be applied to such systems. A set of six dimensions were defined, namely: `time` and `space` - including the values synchronous/asynchronous and co-located/remote respectively; `symmetry` - whether collaborators have the same basic roles and capabilities (symmetric) or whether they have different roles or capabilities (asymmetric); `artificiality` - extent to which a space is based on the physical world or either synthetic, spanning between physical, mostly digital, or hybrid; `focus` - primary target of collaborative activity, which can be defined as environment, workspace, person and object; `scenario` - overall concept of a system according to the users and use case. The authors emphasize that MR systems have been facing significant engineering hurdles, being limited by the contemporary capabilities of technology, and have only recently started to mature to the point where researchers can focus on the nuances of supporting collaboration (focus squarely on the human concerns that underlie communication and collaboration), instead of the need to focus on creating the enabling technology, that makes AR collaboration possible.

Finally, recent work by Sereno *et al.* [11] presented a systematic survey, reviewing 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`. The authors derived design considerations for collaborative AR environments, and identified research topics to further investigate, such as the use of heterogeneous hardware and 3D data exploration. The survey also contemplated collaborative immersive analytics using AR technologies to provide an overview of the field for newcomers, researchers and domain experts.

In fact, the effort of searching for a categorization in this recent work is evidence that the research community is trying to bring forward a systematic view over the literature. An aspect this manuscript tries to address.

2.3 Summary

While existing categorization efforts focus on specific use cases or aspects of Collaborative AR, they do not intend to cover the complete landscape as described in this paper. Considering the reviewed literature, there are several aspects that are deemed to deserve further attention. Most works report the characteristics of the technologies developed to address collaborative efforts, which is a valuable contribution to the community but, to the best of our knowledge, there are no efforts to explicitly convey the concrete contexts that each solution is designed to serve, defining the characteristics of, e.g., the team, environment, and collaborative context. In practical terms, given a concrete application context and a problem, the research community is still not able to provide an overall definition of the collaborative AR system that addresses it. Furthermore, this also makes it more difficult to establish an explicit relation between collaborative dimensions and how they are served by AR, which could help identify potential gaps.

Additionally, existing efforts are mostly oriented towards technology. A human-centered approach, i.e., focusing on the feature that needs to be provided to serve the collaborators, instead of the technology, might bring forward a perspective that is not rapidly deprecated with the advancements of technology [54].

3 Conceptual Model and Taxonomy

This section explains the methodology adopted, describes a conceptual model and a human-centered taxonomy for Collaborative AR. An illustration of the use of the taxonomy is also presented.

3.1 Method

To reach a first proposal of the dimensions defining the work carried out in Collaborative AR, we adopted a conceptual-to-empirical methodology, adapted from the work of Nickerson *et al.* [55] and partially inspired by the research method used by Collazos *et al.* [27]. The methodology followed for the creation of the taxonomy was based on a participatory design process [69], i.e., actively involving stakeholders in focus groups and brainstorm sessions [70], [71]. To this effect, we gathered a set of multidisciplinary experts with several years of expertise (minimal of 6 years, and a maximum of 30 years of experience) in the areas of Human-Computer Interaction (HCI), Virtual and Augmented Reality (VR/AR), Visualization, Multimodal Interaction, as well as remote assistance and maintenance in industrial contexts, involved in various types of collaborative work. In total, 15 experts were involved, although, in many cases, not all at once. These individuals had various professions, e.g., PhD students, researchers, faculty members, project managers, maintenance technicians, remote technical instructors, thus benefiting from on-going collaborations with partners from the Industry sector.

In this vein, we conducted several face-to-face and remote meetings, focus group and brainstorm sessions (sometimes with different combinations of experts according to their availability) over several months. In this process, we used illustrative materials like storyboards, diagrams and videos of our own work in the field of Collaborative AR. In addition, multiple collaborative tools were used for discussion, analysis and brainstorm, e.g., Evernote, SimpleMind, NodeD, VoiceRecorder, Zoom, OneDrive, Microsoft Word and PowerPoint, Overleaf. A moderator facilitated the discussion using scripts to elicit richer discussion, and as the work progressed, different iterations of the conceptual model and taxonomy were used to generate debate and deliberation. During these sections, we focused on the following phases:

1) *Explore collaborative realities* of each individual and progressively introduce and discuss the subject of Collaborative AR through the use of different materials, e.g., images and videos of a collaborative AR-based solution being used to address a maintenance problem between two distributed collaborators;
2) *Identification of relevant terms*, i.e., determine a set of objects of interest, based on the area of interest and the expected use of the taxonomy by the research community. In this process, several terms may be added, removed and renamed. Based on the experience of the individuals, an effort was made to identify the different defining blocks of a collaborative effort supported by AR;
3) *Conceptual Model definition* using the terms identified, which then elicited an analysis of the literature, not only about Collaborative AR, but covering the key pieces of the conceptual model to support their definition;
4) *Categorization of objects of interest*, i.e., identification of reoccurring objects and conceptualization of components that may be appropriate to differentiate between those using a graphical process as suggested in [27];
5) *Creation of the Taxonomy*, i.e., form the initial dimensions of the taxonomy, following a similar approach to the one used by Zollmann *et al.* [72]. Their definitions must be clarified and agreed upon. Moreover, each dimension contains categories and characteristics that are mutually exclusive and collectively exhaustive [55], [73];
6) *Detailed explanation* of all objects of interest included in the taxonomy, and description of the main decisions and design alternatives related to them, as well as inclusion of the main related bibliographical references;
7) *Application, discussion and refinement*of the taxonomy to verify if the established dimensions, categories and characteristics were well defined, need to be merged, or if new ones could be identified [55], which resulted in several iterations to the initial taxonomy.

3.2 Conceptual Model for Collaborative AR

During the creation of the conceptual model for Collaborative AR, there was no particular concern with specific supporting technologies, but mostly with the steps required to accomplish it. In this regard, the first step consisted of a conceptualization for the single user AR scenario to establish a baseline. Our goal was to represent the collaboration nature and the most common tasks (which can be present in more than one scenario), the context (co-located or remote collaboration), the collaborative setups, i.e., necessary apparatus to capture and share AR, the collaborator role, the predominant interaction modalities, the level of engagement, and the multisensory context. Moreover, in the context of co-located

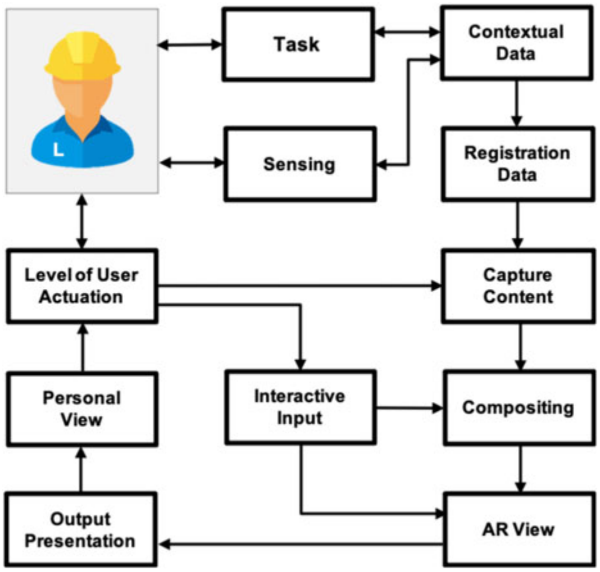

Fig. 1. Conceptual model illustrating the elements associated with how a single user interacts with AR to accomplish a given task.

scenarios, we had in consideration the collaborative setup and cooperation modalities for different levels of engagement with virtual and real objects. Likewise, for remote collaboration scenarios, other modalities can be foreseen, allowing to differentiate the role of each collaborator, according to abilities and prerogatives.

The boxes represent different key elements with arrows loosely indicating a flow between elements. Before looking into how collaboration is performed using AR, let's look into AR when used by a single user to identify the main elements for such a goal. In a typical scenario supported by an AR system (see Fig. 1), users try to accomplish a *task* by interacting with the environment, while their senses may be provided with some *contextual data*. The contextual data is considered from the very beginning of the pipeline that is building the AR view, and it should be assumed that it goes through the different entities and can be used accordingly. *Registration data* is considered to enable the identification of points of interest in the *capture content* and providing the grounds for augmentation spatially (and temporally) aligned with the reality. An *AR view* is generated and presented to the user through some *output presentation*. Different *levels of user actuation* may be possible entailing, e.g., the ability to modify how the scene is augmented through *compositing*, i.e., which elements are visible, exploring particular aspects of the scene, or interacting with content through the *interactive input*.

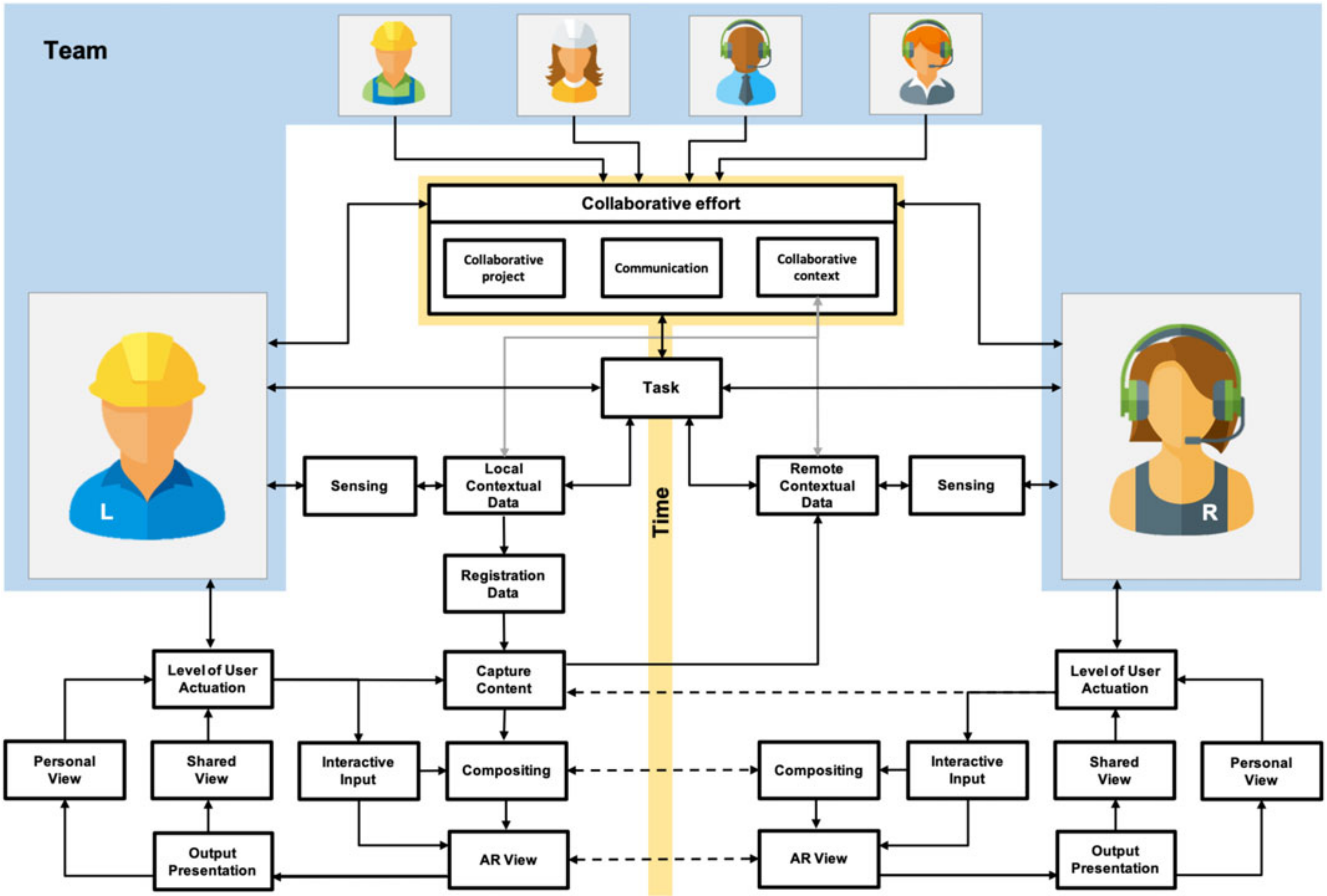

Fig. 2. Conceptual model for Collaborative AR. For the sake of simplicity, the diagram only shows the different conceptual blocks for one local and one remote user in detail. Dash lines imply the existence of connections between elements, which are not mandatory, but may occur if needed during the collaborative effort.

From this single-user model, we evolve to a collaborative conceptual model (see Fig. 2). When moving into the collaborative setting, several aspects are replicated for the different users involved, as further discussed below, but there are a few additional elements that take shape. First, a *team* is now involved with its characteristics, e.g., the number of elements, their profiles, and their location. For the sake of simplicity, the diagram only shows the different conceptual blocks for one local and one remote user in detail, despite having present additional team members. With the collaboration effort, emerges the need for *communication* among the team members considering the channels suited for the *context* and *task*, which may be affected by different *time* aspects. Within the field of CSCW, the term awareness can be defined as "an understanding of the activities of others, which provides a context for your own activity" [74]. Awareness relates to the knowledge one has of other team members' actions. Such knowledge is used to inform one's own action in a way that makes the whole team move forward in the collaborative effort. While awareness is associated to the knowledge of what is going on at a particular moment, the notion of common ground refers to the common understanding of joint goals, shared resources and the state of the *task* solving process [75]. More formally, it can be defined as "a state of mutual understanding among participants about the topic at hand" [76]. The existence of a mutual understanding between team members is based on working vocabulary, practices and norms, that contribute to a sense of shared knowledge and awareness [77] and is of utmost importance. These allow team members to work together effectively, adjusting their activities as necessary through different shared context sources.

The *contextual data* of the remote user is updated according to the *capture content* to present him an updated view of the *local context data*. On the remote side, the user can be provided with different levels of information regarding the *local context*, including, e.g., a video feed, a virtual scene or, even, tangible landmarks or a reproduction of the physical context. In the simplest situation, a view of the *task* setting is provided and can, if possible, ask the local user to provide different views. The view(s) available can also include augmented content whether sharing what other team members are seeing or adding information relevant for the remote member's function. Additionally, the remote user might have some control over the *capture content* and be able to select particular views, e.g., controlling a camera. The coexistence of more than one team member might also motivate the consideration of more than one view on each side: one that is shared, e.g., the remote user seeing the same view as the local user; and an additional view, e.g., for an overall analysis of the environment or selecting a different augmentation to explore additional information. These realities can be shared with more or less fidelity at the counterpart of each user: local user can see aspects of the physical reality and vice-versa. Also, both local and remote user can interact and modify some attributes of the augmented elements in their counterpart view. Finally, each user might have a *shared view* (common between both) and a *personal view* not shared.

The conceptual diagram was continuously analyzed to identify potential grouping of blocks according to their overall purpose and interrelation. Finally, the terminology considered for each of the elements was harmonized, as best as possible and without affecting the overall concept, to adopt nomenclature already used in recent literature [72]. The outcomes were then considered as the basis for the proposed taxonomy, as presented in what follows. To elaborate, the team, time, task, communication and the level of user actuation dimensions were kept directly from the conceptual model as is. The scene capture and tracking dimension are the result of merging the *registration data* and the *capture content* from the conceptual model. Moreover, the shared context sources dimension is the result of adapting the collaborative context (including the local and remote contexts particularities like physical context, sensory context and sensing) in order to encompass more aspects like the human, environmental and collaborative factors. In turn, the input modalities dimension, as well as the output and augmentation were created based on the *output presentation* and *interactive input*. Finally, the research dimension resulted from the analysis and evaluation of the taxonomy itself, not being directly related to the conceptual model.

3.3 Taxonomy for Collaborative AR

Our taxonomy aims to not only propose a first systematic approach to the more intrinsic (technological) characteristics of Collaborative AR, but also to put them in relation to key aspects of collaborative work (Fig. 3). In what follows the different dimensions included in the taxonomy are presented and the categories and characteristics detailed. Additionally, for easier reference, a companion table is provided for each dimension.

TEAM	
physical distribution	
	co-located, distributed, mixed-presence
role structure	
	functional, divisional
size	
	two, three or more
life-span	
	short-term, long-term
turnover	
	low, intermediate, high
technology usage	
	sporadically, systematically
multidisciplinarity	
	yes, no

The characteristics of a team involved in a collaborative effort define much of how the tasks progress occurs [77].

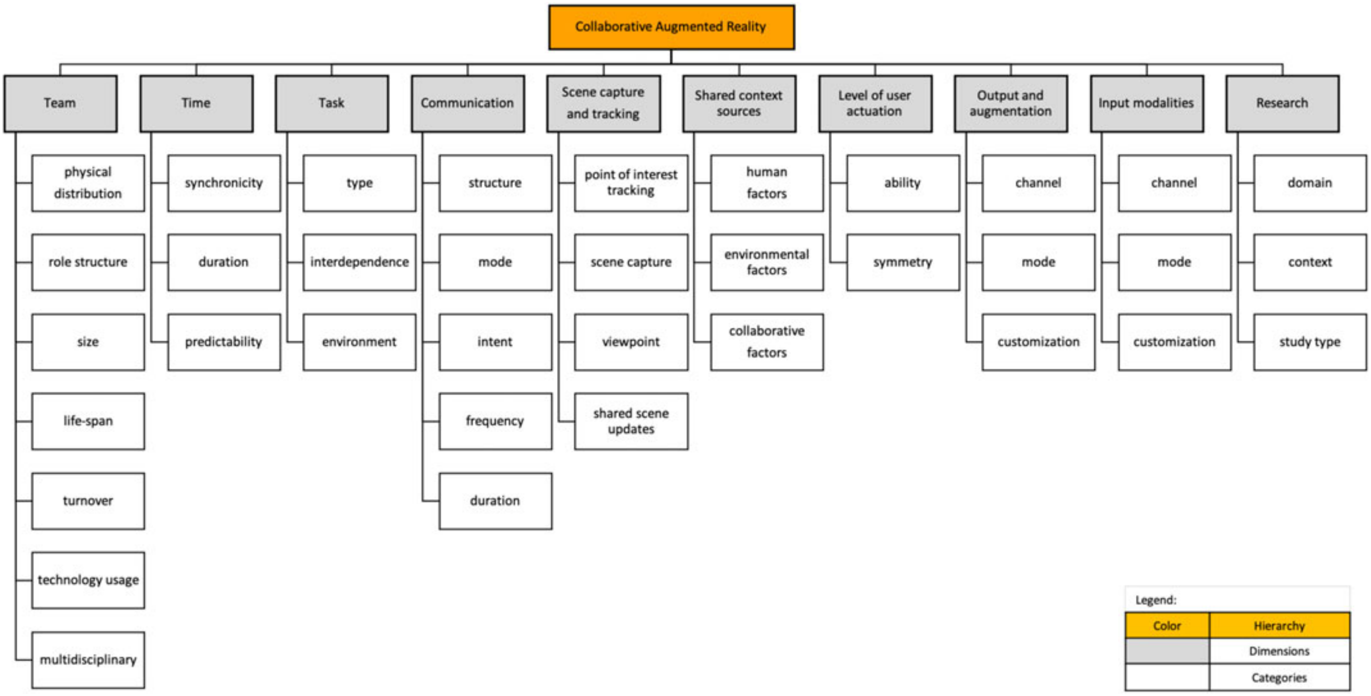

Fig. 3. Taxonomy including the different dimensions and categories identified for Collaborative AR categorization.

First of all, the *physical distribution* of its team-members, if they are in the same location (co-located), if they are all in remote locations (distributed) or a mix of these two cases (mixed-presence [78]). This corresponds to one of the dimensions identified in the seminal work of Johansen [32].

Another important aspect dictating how the teams need to work is their *role structure*. If the team is functional, it means that each member has a specific function or expertise, but if the team is divisional, all elements have the same level of expertise, but collaborate to divide the work [79], [80].

The team's *size* is also an important aspect to consider, since it can have impact on several aspects during design and run-time, for instance, in how to make it clear, for all, who is intervening or performing a certain task, at a certain time. At this stage, we distinguish between two and three or more elements [81].

Team *life-span* (or permanence), refers to the amount of time a team exists as so. If a team is assembled to tackle a particular task and, then disbanded, it is said to be short-term; however, if the team persists, over time, across multiple tasks it is classified as long-term.

Team *turnover* [81] refers to the amount of expected change in the elements intervening in the collaborative effort, i.e., how often team members leave and/or new team members are added, ranging between low, intermediate or high.

Technology usage includes the amount of effort devoted to the use of technology, i.e., how frequently technological solutions are used during the collaboration effort. Team members may use it sporadically, i.e., once in a while, not a common practice or systematically, i.e., technology has been established as one of the practices for collaboration and used often [77], [82].

Finally, a *multidisciplinary* team, i.e., the presence of members with different backgrounds and perspectives over the task (analogous to the number of communities of practice discussed by Lee and Paine [81]), might pose particular challenges regarding how, e.g., a more elaborate context needs to be provided, communication is supported, or adaptation needs to be available to allow custom discipline specific augmentation.

Time	
synchronicity	synchronous, asynchronous, mixed
duration	short, intermediate, long
predictability	scheduled, unscheduled

This dimension groups characteristics that have to do with how the different elements of collaboration relate over time, considering: *synchronicity*, i.e., if all team-members are present and can act in real time (synchronous), if collaborative actions, performed by different elements, take place at different times (asynchronous), or mixed synchronicity, which might be relevant, for some tasks;

Duration refers to the time required to accomplish short (less than 30min), intermediate (between 30min and 90min), or long (more than 90min) tasks. This is important, since a certain setup might be adequate for short usage, but uncomfortable for longer periods; and *predictability* of the collaboration [83] describes if it happens at well-defined `scheduled` (predictable) or `unscheduled` (unpredictable) times.

TASK	
type	
	management, advisory, negotiation, psycho-motor, defined problem, ill-defined problem
interdependence	
	pooled, sequential, reciprocal, intensive
environment	
	indoor, outdoor, mixed

The task is central in a collaborative effort and dictates much of the communication, information, and augmentation requirements [6], [77]. Regarding the task *type*, it can be divided in [84]: 1) `management`, where someone assumes the supervision and coordination of others; 2) `advisory`, entailing professional support, e.g., providing expert advice; 3) `negotiation`, when two or more parties need to resolve conflicts and reach agreement; 4) `psycho-motor` action, referring to those tasks consisting of the manipulation of a machine or product involving elaborate movements and/or psychological processing, whether in physical or virtual reality. 5) `defined problem`, i.e., problems with well defined answers, e.g., when a remote expert has the solution and provides instructions; 6) `ill-defined problem`, when no partner has an immediate solution and they, e.g., generate/share ideas or plans through brainstorming.

Interdependence describes how actions of team-members may be influenced or limited by others and may vary among [84]: 1) `pooled`, when each member can make their contribution independently from others, possibly asynchronously; 2) `sequential`, when the actions are performed in a well established sequence of team member contributions; 3) `reciprocal`, when the different actions to accomplish the task are performed in sequence, but there is a continuous adjustment to how the task is progressing, entailing a certain level of unpredictability, e.g., choosing the kind of expertise required for particular situations. To clarify, sequential interdependence presumes a fixed and well defined sequence of steps with, typically, a precise definition of the team member involved in each, while in reciprocal interdependence, team members need to work in sequence, but there is a back and forth adjustment depending on how the task progresses and the expertise required, at each time; 4) `intensive`, when all team-members need to work simultaneously, i.e., synchronously to accomplish goals of a given task.

Finally, it may be relevant to know if the task is performed in an *environment* located `indoor, outdoor`, or `mixed`, since this might impact on how the AR-based system is designed and how the collaborative process occurs.

COMMUNICATION	
structure	
	hub-and-wheel, chain, star
mode	
	verbal, textual, graphical, gestural
intent	
	inform, commit, guide, request, express, decide, propose, respond, record
frequency	
	never, sometimes, continuous
duration	
	short, intermediate, long

The communication *structure* describes how the message flows inside the team [77]. Inheriting from the work of Wildman *et al.* [84], three structures can be considered: 1) `hub-and-wheel`, where all communication passes through one team element (e.g., a leader), and flows to others through him/her; 2) `chain`, where the message reaches each team element through a hierarchy; and 3) `star`, where every team member freely passes and receives information from others.

The communication *mode* [85] characterizes what elements are possible, such as `verbal, textual` (e.g., messaging), `graphical` (e.g., sketch), or `gestural` (e.g., hand gestures).

The *intent* [85], [86] emphasizes the existence of an explicit goal: `inform, commit, guide, request, express, decide, propose, respond,` and `record`. It marks a difference to other aspects of the collaboration dynamics which, although related to communication, are not explicit. For instance, using hand gestures to, convey or complement a message (e.g., pointing to a specific area) is communication related [87], while sharing hands' position to contextualize how the task is being performed would not.

Frequency characterizes how often communication can (or needs to) occur to accomplish the task: `never, sometimes, and continuous`, and *duration*: e.g., `short` (less than 5 seconds), `intermediate` (between 5s and 5min) or `long` (more than 5min). Both these aspects are dependent (or might face challenges) on a number of other factors such as the type of task and team distribution. For instance, for particular tasks, frequent communication might be mandatory.

Following, some clarifications are presented regarding the difference between some categories of the previous dimensions, namely: "Time:duration", "Team:life-span", and "Communication:duration". While the Time:duration and Communication:duration may vary in similar characteristics, i.e., short, intermediate or long task/long term, Time duration refers to the total amount of time required to accomplish a given task, while Communication duration refers to the amount of time used for sharing information between team-members. Finally, the Team life-span refers to a different aspect, in particular, the amount of time a team exists with its members, either long-term, if the team persists, over time or short-term, if a team is assembled to tackle a particular task and disbanded at the end. So, for instance, a short-term team assembled to solve an emergency may perform a long task supported on small duration communications for, e.g., action synchronization.

Scene Capture and Tracking	
point-of-interest tracking	
	computer vision
	marker, markerless
	sensor
	electromagnetic (RFID), GPS
scene capture	
	camera, stereo camera, depth camera, 360 camera
viewpoint	
	fixed, mobile dependent, mobile independent
shared scene updates	
	static, dynamic, live

An aspect to consider is *point-of-interest tracking*, i.e., how the system knows where are the relevant key features (e.g., objects, location) enabling proper registration of the augmented content. It may vary among: 1) `computer vision` methods, resorting to marker or markerless approaches; 2) `sensor` (e.g., electromagnetic, Global Positioning System (GPS), Inertial Measurement Unit (IMU); or 3) `non-existent`. The latter option encompasses situations for which augmentation is done without a direct connection with scene elements, e.g., instructions provided by remote teammates presented to an on-site technician field-of-view, while he performs a maintenance procedure.

The *apparatus* is the technological device used to support scene capture (and tracking) and can range from a simple `camera`, to more complex devices, such as `stereo cameras`, `depth cameras` or `360 cameras` [5].

Finally, *viewpoint* refers to the nature of the views of the scene that are available. In this regard, we consider three alternatives: 1) `fixed`, e.g., from a fixed overhead camera; 2) `mobile, dependent` on the user, e.g., from a handheld device or a user mounted camera providing POV; and 3) `mobile, independent`, e.g., a camera mounted in a robotic arm that can be oriented to provide any particular view [5].

The *shared scene updates* are associated with how up-to-date the environment is updated. It may be 1) `static`, e.g., 360 image; 2) `dynamic`, e.g., 360 video; or 3) `live`, e.g., 360 video stream.

Shared Context Sources	
human	
	personal
	age, abilities and knowledge, perceptual, cognitive, affective state
	task-specific
	gaze orientation and focus, availability, location, motor activity
	social factors
	system sociability, social space, social presence
environmental	
	physical, digital
collaborative	
	action timeline, progression

Context-awareness is a field of research deserving strong attention in multiple areas. In this regard, it is important to distinguish two main purposes: 1) to provide the computational system with a context that might enable adaptive behaviors; and 2) provide the users with information that contextualizes the task they are involved in [88], [89]. Regarding collaboration and AR, two notable examples of a systematic (taxonomic) approach to the subject are the works by Grubert *et al.* [90], performed in the context of their proposal of the concept of Pervasive AR, and by Collazos *et al.* [27], proposing a taxonomy for context information sources in the scope of their descriptive theory of awareness for groupware development. Here, we leave out the subject of discussing which context sources can be considered for providing adaptive system features (see [90] for an in-depth discussion on system context-awareness). Instead, we focus on these context elements considering their potential importance to increase the awareness of team members regarding the collaborative context. Since system-side context-awareness has a parallel with collaborative context awareness (by sharing a range of context sources), we consider the nomenclature proposed by Grubert *et al.* [90] selecting those dimensions and characteristics with a more immediate relevance for collaborative scenarios and inherit, where deemed relevant, from [27]. To clarify, we do not consider as context those elements that arise from explicit communication by any team member [27]. Furthermore, it is important to note that context sharing is not only useful for remote collaboration, but can also be an important resource in co-located efforts [91]. In fact, the increasingly common multi-device ecologies often generate team and task awareness fragmentation [92], [93] that might be tackled by an explicit presentation of context elements.

Human factors, i.e., those pertaining team members and their performance encompass `personal`, `task-specific`, and `social` aspects. Personal human factors specifically relate to individual user characteristics or states including, e.g., age, abilities and knowledge, perceptual, cognitive (e.g., cognitive load) and affective states. Additionally, it can also encompass aspects directly specifying the user's context within the task, such as gaze orientation and focus (point-of-interest), availability, location, and motor activity. Social human factors account for the broader scope of interaction among people. The dimension of social interaction in collaborative efforts is highly relevant as it fosters improved learning, group formation and group dynamics [94]. Considering the framework proposed by Kreijns *et al.* [94], social interaction depends on the systems sociability, on the creation of social space, and on social presence. `Presence` goes beyond the simple information regarding the location or availability of a team member, as it entails a sense of someone being present and following what the person is doing [27]. The inclusion of such feature (e.g., through avatars [95], [96]) is relevant as it enables a remote collaboration experience and performance that is closer to what is possible in co-located work. Additionally, research hints that, as happens with co-located work, the sense of someone being present might have an influence on team member performance. For instance, Miller *et al.* [97] have shown how a sense of presence might have a social facilitation or inhibition effect depending on the difficulty of the task: having the sense of someone present, when performing a difficult task, lowered performance. The extent to which these elements are present depends on several aspects, such as the supported level of communication, but also greatly depends on how social cues are made available to team members.

Environmental factors concern everything about the `physical` and `digital` (i.e., augmented) environment being experienced. The physical environmental factors describe the characteristics of the place where the user is positioned including elements like temperature, ambient noise and light intensity, and the spatial and geometric configuration of relevant artefacts [27], [82], [98], [99]. Digital environmental factors refer to elements that provide context about the characteristics of the AR environment and the features it provides. This is important in certain contexts, in which teams rely on different technological resources which should be known by other members [27] to guarantee success of collaboration [82]. For example, the amount of information each team member views, if the tracking/alignment mechanism are working properly, if all virtual elements are being rendered in a satisfactory manner, the availability of adequate infrastructure factors, like bandwidth for distance technology tools, state-of-the-art workstations or the availability of technical support.

Collaborative factors pertain information that provides a wide contextualization of the collaboration effort, further supporting coordination, complementing communication, an aspect of the utmost relevance for collaborative work, by contributing to a shared team cognition and potentially driving anticipatory behavior and implicit coordination [100]. The collaborative `action timeline` [27] refers to information regarding sequences of past actions of different team members, and annotations or outcomes that provide a reference procedure to solve a problem, a context of what has been attempted or performed, so far, or support auditing procedures. In turn, `progression` refers to a less granular level of information. While the timeline can work similarly to a logbook, progression entails additional detail providing a runtime performance monitoring of each member on their current task. This is important to enable a team-level perspective of the ongoing work, serve to support coordination, and help create conditions for the team to adjust to different phases of the tasks [101], e.g., informing when expert support may be required and, facilitating articulation of individual actions with the collaborative efforts [102].

Level of User Actuation	
ability	
	passive-view
	on-site, remote
	interact / explore
	none, on-site, remote
	share / create
	none, on-site, remote
symmetry	
	symmetric, asymmetric, fully asymmetric

The user's actuation *ability* can range from `passive-view`, which can be on-site or remote, to `interacting/exploring`, e.g., manipulating content present in the scene, and to `sharing/creating`, e.g., adding annotations to the scene or new views or content that others can see [103]. According to Isenberg *et al.* [103], these are associated to the level of engagement and may apply to none, on-site and/or remote users. We chose 'actuation' to avoid a clash with 'engagement' also being used, in the literature, to refer to the amount of motivation and commitment a user is devoting to a task [104], [105].

A user involved in AR-supported collaboration is also influenced by the level of *symmetry*, which represents if all parts have the same level of actuation: `symmetric`, i.e., whether collaborators have the same basic roles and capabilities; `asymmetric`, i.e., whether they have different roles or abilities [9]; or `fully asymmetric`, i.e., a remote user is equipped with the abilities that can help solve an onsite user's problem without any help being provided onsite [106]. The inclusion of full asymmetry is motivated by the passive role of the onsite user, which creates a context with specific challenges to address, beyond those of asymmetry, since the onus of action is on the remote user.

Output and Augmentation	
channel	
	vision
	self, modifier, wearable
	audition
	airborne, structure, wearable
	touch
	tactility, haptics, vibration, wearable
	kinesthetics
	proprioception, equilibrioception, kinematics
	olfaction
	ambient, wearable
	gustation
	neural oscillation
	galvanism
mode	
	unimodal, redundant, complementary
customization	
	adaptable, adaptive, non-customizable

We choose to have a level devoted to the sensory *channel* receiving the output rather than directly addressing the technological apparatus since this enables an easier grasp of which channels are specifically considered, to avoid uncertainty when the device might serve many channels (e.g., a tablet might provide visual, haptic, and auditory output). Additionally, centering the categories on the users, it should enable a more versatile categorization to encompass novel technologies and devices. In this line of thought, our proposal inherits from the detailed work of Augstein and Neumayr [54] proposing a human-centered taxonomy for interaction. The authors identify six modalities (sensory channels), related to human perception capabilities from which our work inherits to characterize output augmentation. Output and Augmentation can be performed through: `vision`, including standalone self appearance changing devices (e.g., monitor), their wearable alternatives (e.g., HMD), and external medium appearance changing devices, i.e., devices that can change the appearance of an external element (e.g., video projector); `audition`, considering airborne sound propagation (e.g., sound speakers), through a structure (e.g., bone), and possibly wearable (e.g., headphones); `touch` [107], including tactility (i.e., devices that simulate being touched), haptics (i.e., devices that shift their physical properties, e.g., shape, temperature) and vibration, also considering wearable alternatives; kinesthetics, considering our senses of proprioception (i.e., body orientation and position), equilibrioception (i.e., body balance), and kinematics (i.e., acceleration); olfaction through a device located in the ambient or wearable (e.g., olfactometer [108]); and gustation. Additionally, Augstein and Neumayr [54] distinguish between the set of channels above, which entail a perception/action from one of the senses that is further processed by (or originates at) the brain, i.e., indirect processing, and those that directly deal with brain or muscle activity, i.e., direct processing. For the latter, the authors identify: neural oscillation and galvanism.

To explicitly convey if a system allows multimodal augmentation, i.e., through multiple channels, *mode* can be: unimodal, if only a channel is used, at each time, regardless of how many are available; redundant, and/or complementary, if multiple augmentation channels are used to reinforce or add information.

Finally, *customization* refers to the possibility of the user (adaptable) and/or the system (adaptive) to automatically choose or customize the most suited channels for output. It can also be non-customizable.

Input Modalities	
channel	
	vision
	fixed, wearable
	audition
	airborne, structure, wearable
	touch
	tactility, haptics, vibration
	kinesthetics
	proprioception, equilibrioception, kinematics
	olfaction
	gustation
	neural oscillation
	galvanism
mode	
	unimodal, redundant, complementary
customization	
	adaptable, adaptive, non-customizable

For the input modalities, we adopt a similar rationale as the one adopted for the output modalities, considering a human-centered characterization aligned with the work of Augstein and Neumayr [54], encompassing the following six *channels* related to human perception: vision, covering fixed (e.g., kinect) or wearable (e.g., eyetraker glasses) devices that capture/process visual data; audition, including devices that capture airborne sound waves (e.g., microphone), through structural propagation in other materials (e.g., ear-bone microphone), and if they are wearable; touch, encompassing tactility (i.e., a device sits passively and is touched), haptics (manipulation of an explorable physical surface, e.g., braille keyboard) and vibration (a device sensing vibrations, e.g., tremors); kinesthetics, considering proprioception (i.e., position and orientation of the body), equilibrioception (i.e., body balance), and kinematics (i.e., acceleration); olfaction; and gustation. Additionally, two channels are considered to cover input though brain or dermal activity, i.e., neural oscillation and galvanism.

Mode refers to how the available modalities can be used to perform interaction. As for the *output and augmentation* dimension, we consider the options unimodal, when only one modality can be used, at once, and when these are explored together: redundant, when modalities can be used simultaneously to perform the same action, or complementary, when multiple modalities are used in sequence to provide different parts of an action (e.g., pointing to an annotation and saying "delete").

Finally, *customization* refers to the possibility of the user (adaptable) and/or the system to automatically choose (adaptive) the most suited channel. It can also be non-customizable.

Research	
domain	
	industrial, education / training, architecture / construction, tourism / heritage, medicine, entertaining / gaming, among others
context	
	basic research, applied research
study type	
	pilot, informal, formal, field

The last dimension we have considered is devoted to research, allowing to clarify the maturity and detail of the collaborative work being reported. In this context, the research *domain*, or topic is associated to the area of application, ranging between medicine, industrial, education/training, architecture/construction, tourism/heritage, entertaining/gaming, among others [47], [68].

According to the collaborative effort and the tasks being addressed, the research *context* may vary between basic research, i.e., the technologies and/or methods investigated are novel and have not matured, yet, to be usable in real scenarios, often considering dummy tasks as the case study (e.g., assembly of Lego blocks, tangram puzzles) and evaluation; and applied research, i.e., the technologies and methods are implemented in practice using problems inspired by real-world scenarios (e.g., industry related procedures), and an evaluation of the technique is conducted [109], [110] for those scenarios.

There are various types of scientific studies [47]. The choice of *study type* mainly depends on the research goal, and may vary between pilot, i.e., small-scale preliminary studies aimed to investigate crucial components of a main study; informal, i.e., studies aimed at getting more input, in a quicker manner, without following any structured method; formal, i.e., studies that follow structured methods to obtain measures; field, i.e., studies conducted outside a laboratory environment.

3.3.1 Critical Analysis and Refinement

To understand if the taxonomy can be applied to recent research for assessing how the reporting of the works, along with a possible ambiguity of some characteristics of the taxonomy might affect its use, we asked four experts to critically analyse our proposal. The selected group (one female), aged from 28 to 63 years old, included PhD students, researchers and faculty members, sharing several years of expertise in HCI, VR/AR and Visualization. They also had previous experience in remote collaboration, co-authored multiple publications, and participated in international projects on these subjects over the years.

We choose to use an approach focused on utility demonstration, in which the experts were required to classify subject matter examples, i.e., publications selected by the experts in the field of Collaborative AR [111], [112]. In this context, the experts were provided with instructions, in conjunction with the definitions above. In summary, they were asked to select at least two subject matter examples each and ensure that they could be clearly, concisely and thoroughly classified into the taxonomy, to verify if the established dimensions and categories were well defined, needed to be merged, or if new ones could be identified.

After a period of understanding, adjustment and use of the taxonomy, the experts reported that the taxonomy was globally straightforward to use and apply. Overall, most of the dimensions were easy to follow and comprehend, and in case of doubts, as reported by one of the experts, the description was generally enough to understand and systematically organize the subject matter examples with a significant level of confidence.

The main difficulty reported was related to the *Task, Team and Shared Context Source* dimensions. In the examples analyzed, the source of the information could only be vaguely found in the evaluation sections, which was done resorting to simple tasks (e.g., involving Lego bricks). According to all reviewers, this lack of information on the aforementioned dimensions meant they could not be easily mapped. The emphasis of the experts' feedback regarding this issue, particularly for such recent works, may hint that many of the works exploring Collaborative AR aim to address real life scenarios, but are not mature enough to be used in such cases. As such, many of the dimensions in the taxonomy incorporate categories and characteristics currently not properly reported in existing works, which are still focused on technology aspects of collaboration. Nevertheless, the experts believe that this makes the taxonomy even more interesting, since it opens questions in areas where most researchers are not yet focused, which may be interesting opportunities for the future. Moreover, one expert suggested feeling the need to identify which side of the collaborative process is being described by this dimension. The expert proposed that the dimension was revisited, in order to include characteristics that allow to ease this gap.

Afterwards, we focused on identifying patterns regarding the experts' concerns, e.g., dimensions that could require further attention, structure of some categories, or characteristics missing some examples to better be understood. Also, how easy it was to use the taxonomy, what was their approach in case of doubt, and how they decided to proceed when some information could not be quickly identified in the publications analysed.

Then, we carefully examined and addressed the main observations that were raised. This refinement process resulted in a new iteration of the taxonomy, which are reflected in the aforementioned proposal. More specifically, we performed updates to address doubts pertaining some dimensions, mostly tackled by adding examples of contextualization and improving descriptions. Examples of these include, for instance, clearer definitions of what was considered as short, intermediate or long duration in some dimensions, and a better definition of *Task* interdependence. The *Level of User Actuation* was also improved to reflect which side of the collaboration (i.e., on-site or remote) was being considered. Finally, the main difficulty that arises was addressed by creating the *Research* dimension, devoted to clarifying the context, maturity and detail of the collaborative work.

3.3.2 Applying the Taxonomy to Collaborative AR Works

To illustrate the use and utility of the taxonomy [111], [112], ten publications that explored different aspects of collaboration (e.g., collaborative systems, aspects being addressed and evaluated, among others) were selected and analyzed. The publications were thoroughly classified into the proposed taxonomy dimensions, categories and characteristics to reflect the full extent of the reported information.

To foster insight on the examples analyzed, as well as the way they are classified using the proposed taxonomy we created a quick illustration through a visual representation based on a sunburst (as shown in Fig. 4). We choose to use a sunburst diagram, thus allowing to visualize the hierarchical data of the taxonomy, depicted by three levels of concentric rings. Each ring corresponds to a level in the hierarchy, with the inner ring representing the root node associated to the proposed dimensions. The hierarchy moves outward from the center to represent categories in the center ring and the characteristics in the outer ring. The data of each publication is represented by slicing and dividing rings based on their hierarchical relationship to the parent slice. In addition, a graphical representation of data in the outer ring is depicted using color to highlight the number of publications that addressed each specific characteristics. Mapping the number of publications that address each characteristic to colour helps getting an overall understanding on how they are classified using our taxonomy. This approach allows to comprehend which categories and characteristics get the most attention and identify existing patterns or gaps in a visual way.

The overview provided by Fig. 4 shows that for the *Team*, only distributed collaboration cases were addressed by the selected publications. Likewise, they all focused on teams composed by two collaborators, with a short-term life-span. From these, 7 out of 10 teams were functional, 2 teams reported divisional aspects and the remaining publication did not report the type of role structure. Regarding technology usage, 7 publications acknowledge their teams had continuous use of collaborative tools. In addition, 2 reported low turnover, and 1 high turnover, with the others not reporting any information on this characteristics. Regarding multidisciplinarity, only 3 reported the presence of team members with such background.

Concerning *Time*, all publications focused on synchronous collaboration, with their efforts divided between 5

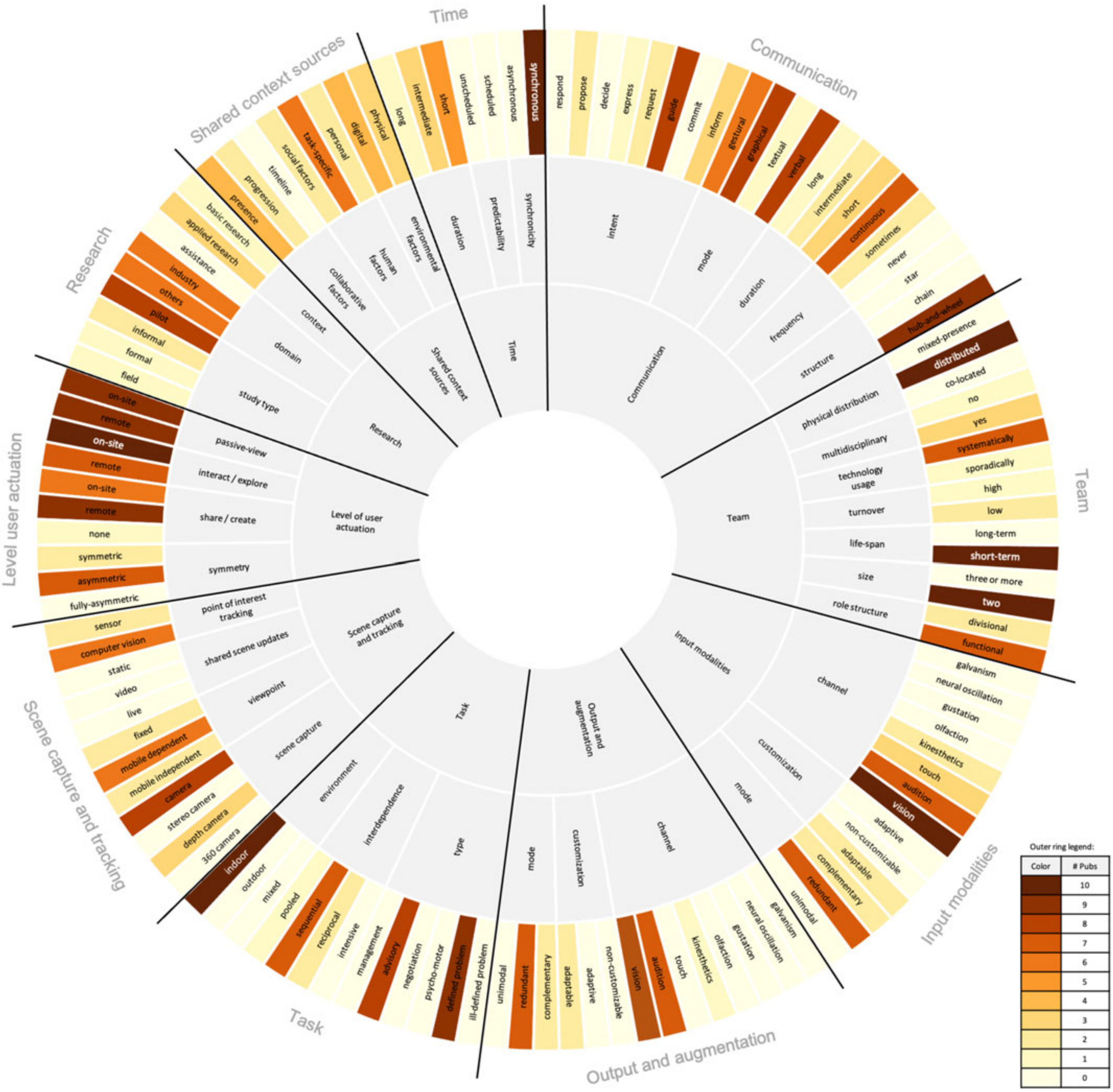

Fig. 4. Sunburst diagram displaying the hierarchical levels of the taxonomy: the inner ring represents the dimensions while categories and characteristics are showed as moving away from the center, respectively. The color scale shows the number of publications (out of a total number) addressing each characteristics. This example presents the results for the classification of the publications: [5], [52], [113], [114], [115], [116], [117], [118], [119], [120].

reporting short and 3 intermediate periods of collaboration. There was no report regarding predictability.

As for the *Task*, only indoor environments (10 publications) were considered while addressing advisory (8 publications) and defined problems (9 publications). Sequential interdependence was explored by 7, reciprocal by 2 and pooled by only 1.

The preferred structured type associated to *Communication* was hub-and-wheel, being used by 9 publications. Moreover, 3 reported the communication duration was short, 2 intermediate, and only 1 publication long. In addition, 8 used verbal and graphical mode for communication, 6 explored gestural and 1 textual communication. Likewise, on the subject of communication frequency 7 publications reported it as continuous. Finally, the intent of communication was split among guide (8 publications), propose and request (2 publications) and inform (3 publications).

In respect to *Scene capture and tracking*, 2 publications reported the use of sensors for tracking, while 6 used computer vision. Moreover, the majority used cameras for the scene capture (8 publications), and 3 resorted using a depth camera. Likewise, 6 used a mobile dependent viewpoint, 2 used mobile independent viewpoint and another 2 explored a fixed viewpoint.

Regarding the *Shared context sources*, the classification shows that 6 publications mention task-specific and 2 personal aspects, as well as social aspects in the human factors category. Besides, 3 publications report physical and 4 digital aspects in the environmental factors category. As for the collaborative factors category, 4 publications mention presence and 2 progression.

Additionally, the *Level of user actuation* emphasizes that for the distributed teams, 9 publications reported on-site and remote team members had access to a passive-view of the task context. In the same way, all on-site collaborators could interact/explore, and only 5 publications report that the remote collaborator could do the same. In terms of share/create, the opposite situation occurs, with 9 publications reporting this ability being available to collaborators. Asymmetric level of user actuation was reported by 6 publications, while only 2 included mentions to symmetric possibilities.

For the *Output and Augmentation*, vision (10 publications), audition (7 publications) and kinesthetics (1 publication) were used as output channel. In addition, 7 publications explored a redundant approach and 2 a complementary. Customization was only reported by 2 using an adaptable approach. In the same way, the *Input modalities* shows that vision (8 publications), audition (5 publications) touch (3 publications) and kinesthetics (1 publication) were used as input channel. The input mode was reported by 6 publications and focused on a redundant approach (5 publications) and complementary approach (1 publication). Regarding customization, only 1 publication explored an adaptable approach, suggesting that there is a lack of information in the selected papers or that this is not being addressed.

Regarding *Research*, 8 publications focused on basic research in the assistance domain, while 2 concentrated on applied research in the industry domain. Moreover, 8 publications described having conducted a formal study and reported on their results, while 1 publication reported an informal study and another a pilot study.

We further observe that some of our taxonomy categories and characteristics are not totally filled, due to lack of information being reported in the selected publications, as previously mentioned in the critical analysis and refinement sub sections by independent experts during their use of the taxonomy.

4 Discussion

While existing categorization efforts are suited for specific use cases or aspects of Collaborative AR, they do not cover the complete landscape of the field. Moreover, since different authors may use different notions when referring to the same aspects, depending on their context, it is important to make this context clear and provide a coherent common ground for systematization and discussion, thus fostering harmonization of perspectives, as well as reporting, and thus making comparative analyses easier.

4.1 Design and Validation

To explore this opportunity, we focused on a participatory design process and adopted a conceptual-to-empirical approach to understand the defining aspects of collaborative work supported by AR, i.e., what needs to be described about a work to provide a full account of its characteristics. As a result, our proposal is different from other efforts described in the literature, which use existing works as grounds and, then, propose a taxonomy that encompasses them. These previous works have their merits and usefulness. However, such an approach implicitly assumes that existing research already covers all the different aspects required to fully address the problem. Therefore, the outcomes speak about where we are, but not necessarily if we are ticking all the important requirements and where should we go next research-wise, particularly if the field has not matured, as a whole. In the case of Collaborative AR, research has evolved tremendously regarding the supporting AR technologies, but at the onset of our work we argued that the field now needs to devote more efforts to understand how collaboration is being served in this context.

To address this goal, we started by proposing and refining a conceptual model which then allowed the identification of ten dimensions that embody an extended human-centered taxonomy for the categorization of the main features of Collaborative AR. The work presented here proposes a set of dimensions that can be used to characterize collaborative AR not only addressing the technological features, but also encompassing the characteristics of the context they serve in the collaborative effort.

Considering the overall methodology adopted to reach the current stage of the taxonomy, it is important to note that its suitability to provide a structured view of Collaborative AR is not inferred solely from how well the four experts managed to classify ten recent articles. Although this is, naturally, a relevant outcome, and one could be tempted to increase the number of articles, the taxonomy is the result of an iterative participatory design method composed of several stages and it is the overall process that ensures its validation. In this context, the classification of ten articles works as one more refinement stage and the outcomes are presented here as further clarification on how to interpret the taxonomy given a set of recent research.

4.2 Utility and Impact

One of the purposes served by the adoption of the taxonomy is to increase the awareness of the research community regarding the characterization of the collaborative process, while also helping to identify those aspects that remain as research gaps. Additionally, the taxonomy provides a common ground to structure the different elements when conducting research. By increasing the awareness for the different dimensions, the taxonomy can foster additional transparency of the research, through better reporting, which, in turn, enables easier assessment by the community, fosters replicability, and the transferability of knowledge. By being provided with detailed information pertaining the different dimensions covered by a work, a researcher can assess the applicability and relevance of what is proposed to a new problem.

The range of the taxonomy results from the fact that Collaborative AR is an interdisciplinary area integrating different aspects from other research fields. Although in some particular situations one may advocate that simpler approaches as the ones described in Section 2.2 may be used, the proposed taxonomy may also be used as a checklist of relevant dimensions to take into consideration, avoiding oversimplifying the collaborative process characterization. In this regard, nothing precludes researchers from considering a subset of the taxonomy, in particular cases, as the scope of their work. However, this will also put evidence on what is left out, and on the need to provide a rationale for it, based on the targeted research goals, along with a discussion of the contributions that is properly adjusted to the selected scope.

One essential point to note is that the proposed taxonomy is not intended as a closed work, but should, instead, be taken as the grounds that might enable the community to elaborate, expand, and refine it. Although some of the proposed dimensions might still not reflect the full scope of some categories by encompassing all possibilities, we consider that they create a clear enough organization to make itself evident where to insert new characteristics. In addition, as boundaries can sometimes be considered blurred, our taxonomy proposal may also be extended to classify collaborative works that do not fit current definitions of AR, VR, or MR.

4.3 Identification of Novel Research Opportunities

One important aspect that should be addressed capitalizing on the work presented here is planning and conducting a systematic classification of literature to get a thorough characterization and understanding of the Collaborative AR landscape. The classification examples show how the taxonomy can be applied to existing publications, but a larger use of the taxonomy is a paramount step expected to hint on interesting trends, identify research gaps (i.e., concrete areas which are not yet fully addressed or reported) and layout future directions in light of the proposed taxonomy.

However, since most of the research efforts on Collaborative AR have been devoted to creating the enabling technology and proposing novel methods to support its design and development [9], it is expected that a majority of the data being reported corresponds to a subset of the dimensions covered by the taxonomy. As the field matures to focus on the nuances of supporting collaboration, the remaining dimensions will also flourish (e.g., the consideration of social aspects [94] as context sources) and accompany the growth already shown by the leading dimensions.

5 Concluding Remarks and Future Work

In the past decade, researchers have devoted their efforts to experiment and mature methods for Collaborative AR, focusing on providing a common ground to enable rich shared experiences. Since AR technology is starting to evolve to the point where research can focus on the nuances of supporting collaboration, it is now time to understand where do we stand and how well can we address the domain of collaborative work with AR.

In this paper, we performed an analysis of the different dimensions that should be taken into account when analysing the contributions of an AR system to the collaborative work effort. The primary contribution of this paper is to bring these dimensions forward into a conceptual framework and propose an extended human-centered taxonomy for the categorization of the main features of Collaborative AR, aiming to create a common ground for systematization and discussion.

The taxonomy was analyzed by four experts through a utility demonstration method, which reported that our proposal was easy to understand and to apply, allowing to thoroughly classify publications into the taxonomy with a significant level of confidence. In addition, the experts believe the taxonomy can identify interesting topics in areas where most researchers are not yet focused, but may be relevant opportunities for the future.

Finally, we selected ten publications for categorization, which explored different aspects of collaboration to demonstrate the use and utility of the taxonomy. Plus, we presented a hierarchical visualization overview based on a sunburst diagram. It is possible to relate the characteristics of each publication to the inner rings and understand how each individual publication is categorized. We also used color to highlight hierarchical groups or specific categories, which allows identifying existing patterns and gaps.

One notable question that the work presented here can trigger pertains how the different aspects composing this multidimensional vision of Collaborative AR can be assessed when we evaluate such systems. To this end, and in line with the methods presented here, our team is currently looking into the evaluation of Collaborative AR systems. Collaboration processes entail high levels of contextual data by focusing on different types of collaborators, on common tasks that explore different levels of difficulty and diversity and by encompassing dynamical environments. As such, and with the growing number of systems and prototypes using AR, it is important to work towards an understanding of evaluation able to bring forward rules, guidelines, and metrics. To that end, we need to fully grasp the characteristics of the evaluation methodologies that should serve Collaborative AR.

In addition, a structured and systematic approach to the field of Collaborative AR, as made possible by the proposed taxonomy, along with the resulting amount of data, creates interesting challenges on how best to take advantage of it to foster insights. In this regard, we also plan to explore interactive visualisation to propose tools providing the ability to process data faster and properly explore, analyse and compare the characteristics of the collaborative effort reported by existing publications, while providing an explicit representation of the current state of the field.

Appendix A

Next, illustrative sunbursts regarding some example publications are presented (Figs. 5, 6, 7, and 8 as well as a summary table for all publications considered (Table 1).

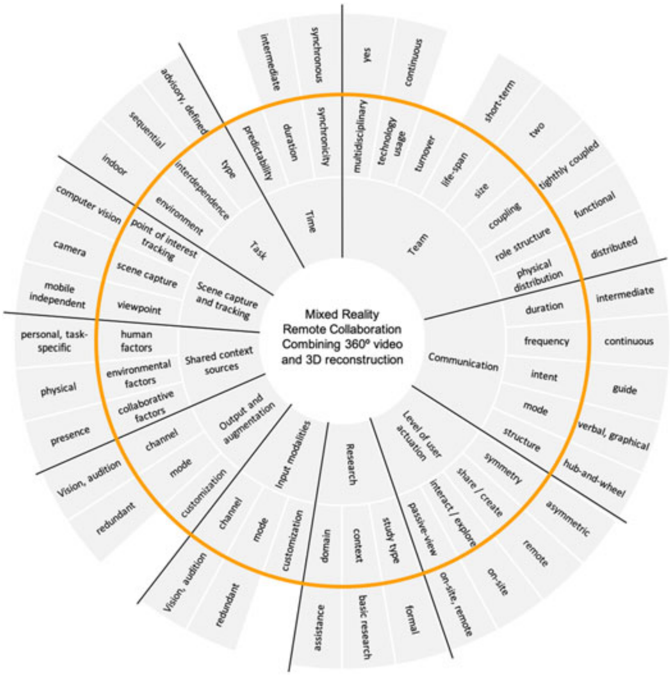

Fig. 5. Visualization overview of the publication: "Mixed reality remote collaboration combining 360° video and 3D reconstruction"[52].

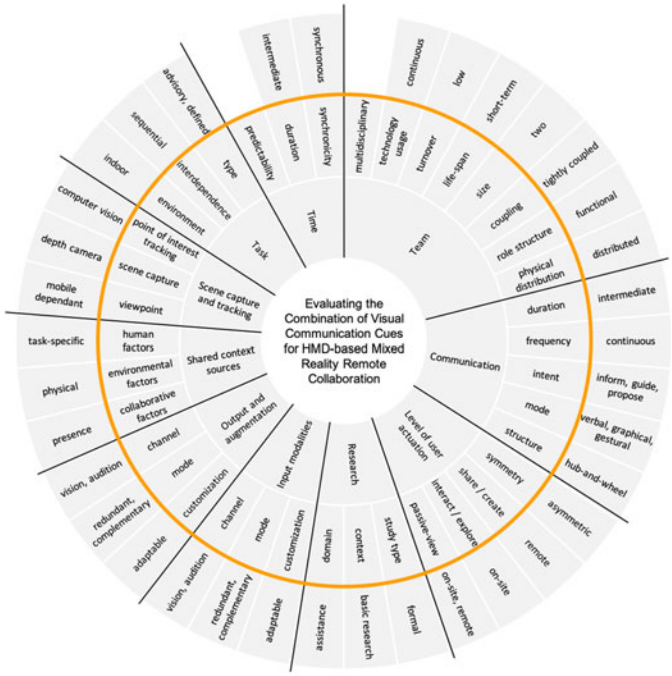

Fig. 7. Visualization overview of the publication: "Evaluating the combination of visual communication cues for HMD-based mixed reality remote collaboration" [115].

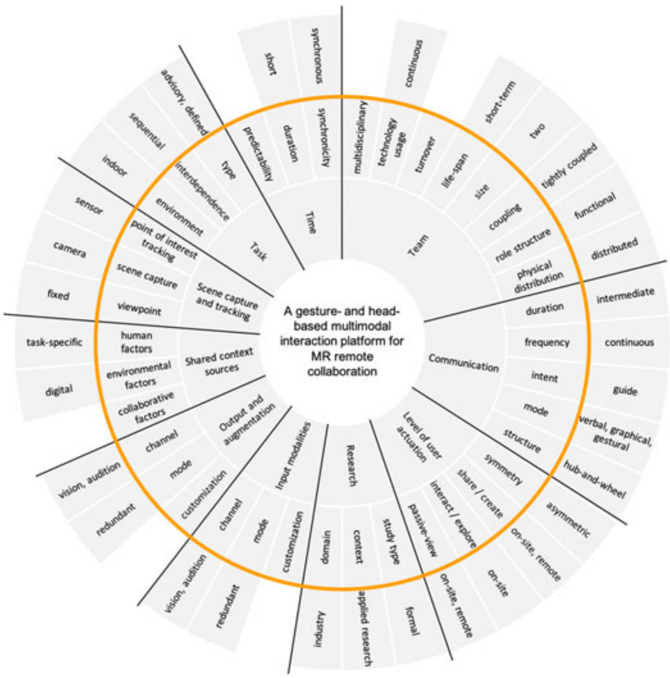

Fig. 6. Visualization overview of the publication: "A gesture- and head-based multimodal interaction platform for MR remote collaboration"[114].

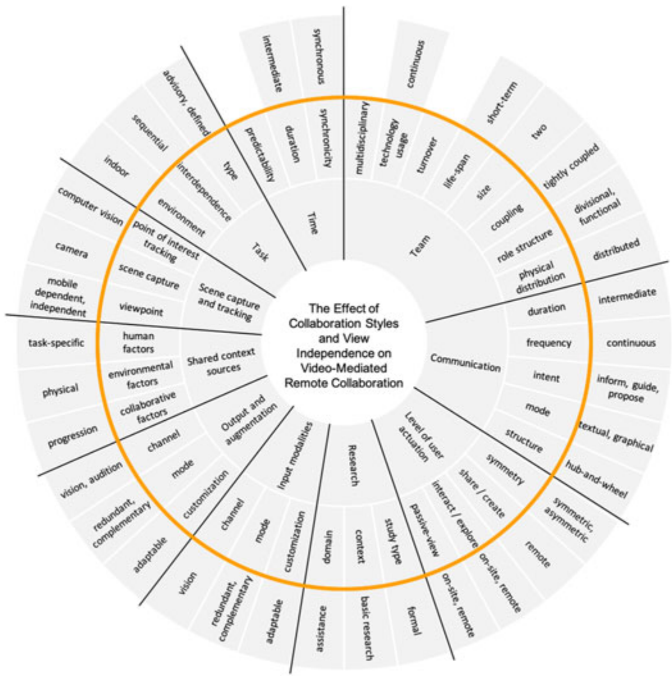

Fig. 8. Visualization overview of the publication: "The effect of collaboration styles and view independence on video-mediated remote collaboration" [5].

ACKNOWLEDGMENTS

The authors would like to thank the reviewers for their thoughtful comments and suggestions towards improving on an earlier version of this manuscript. The authors would also like to thank everyone involved in discussion groups and case studies for their time and expertise. This work was supported in the scope of the PhD under Grant SFRH/BD/143276/2019, funded by the FCT - Foundation for Science and Technology. This work was also supported in part by the IEETA - Institute of Electronics and Informatics Engineering of Aveiro, funded by National Funds through FCT under Grant UID/CEC/00127/2019, in part by the Portugal2020 under the Competitiveness and Internationalization Operational Program, and in part by the European Regional Development Fund through project SOCA – Smart Open Campus under Grant CENTRO-01-0145-FEDER-000010.

TABLE 1
Summary Table of of the Dimensions, Categories, and Characteristics of the Publications: [5], [52], [113], [114], [115], [116], [117], [118], [119], [120]

		Pub. 1	Pub. 2	Pub. 3	Pub. 4	Pub. 5	Pub. 6	Pub. 7	Pub. 8	Pub. 9	Pub. 10
		Gupta et al. 2016	Teo et al. 2019	Wang et al. 2019	Kim et al. 2019	Kim et al. 2018	Aschenbrenner et al. 2018	Teo et al. 2019	Piumsomboon et al. 2019	Piumsomboon et al. 2017	Obermair et al. 2020
Venue		IEEE Transactions on Visualization and Computer Graphics (TVCG)	ACM Symposium on Virtual Reality Software and Technology (VRST)	International Journal of Advanced Manufacturing Technology	ACM Conference on Human Factors in Computing Systems (CHI)	Computer Supported Cooperative Work (CSCW)	IEEE International Symposium on Mixed and Augmented Reality (ISMAR)	ACM Conference on Human Factors in Computing Systems (CHI)	Frontiers in Robotics and AI	Symposium on Mobile Graphics & Interactive Applications	International Conference on Industrial Engineering and Applications
Team	physical distribution	distributed	distributed	distributed	distributed	distributed	distributed	distributed	distributed	distributed	distributed
	role structure	functional	functional	functional	functional	divisional, functional	functional		functional	divisional	functional
	size	two	two	two	two	two	two	two	two	two	two
	life-span	short-term	short-term	short-term	short-term	short-term	short-term	short-term	short-term	short-term	short term
	turnover				low						
	technology usage	continuous	continuous	continuous	continuous	continuous	continuous	continuous	continuous	continuous	Continuous
	multidisciplinary		yes								no
Time	synchronicity	synchronous	synchronous	synchronous	synchronous	synchronous	synchronous	synchronous	synchronous	synchronous	synchronous
	duration	short	intermediate	short	intermediate	intermediate	intermediate	short	long		short
	predictability										
Task	type	advisory, defined	advisory, defined	advisory, defined	advisory, defined	advisory, defined	advisory, defined	advisory, defined	advisory, defined	advisory, defined, psychomotor	advisory, defined
	interdependence	sequential	sequential	sequential	sequential	sequential	sequential		sequential	reciprocal	Sequential
	environment	indoor	indoor	indoor	indoor	indoor	indoor	indoor	indoor	indoor	Indoor
Communication	structure	hub-and-wheel	hub-and-wheel	hub-and-wheel	hub-and-wheel	hub-and-wheel	hub-and-wheel		hub-and-wheel	hub-and-wheel	hub-and-wheel
	mode	verbal, graphical	verbal, graphical	verbal, graphical, gestural	verbal, graphical, gestural	textual, graphical	verbal, graphical, gestural	verbal, graphical, gestural	verbal, graphical, gestural	graphical, gestural	verbal, graphical
	intent	guide	guide	guide	inform, guide, propose	inform, guide, propose	inform, guide, propose	guide	guide, request	inform, guide, propose	guide
	frequency	continuous	continuous	continuous	continuous	continuous	continuous		continuous	continuous	continuous
	duration	intermediate	intermediate	intermediate	intermediate	intermediate	intermediate		long		Intermediate
Scene capture and tracking	point of interest tracking	computer vision	computer vision	sensor	computer vision	computer vision	computer vision		computer vision, sensor	computer vision	computer vision
	scene capture	camera	camera	camera	depth camera	camera	camera	camera, depth camera	camera	depth camera	camera
	viewpoint	mobile dependent	mobile independent	fixed	mobile dependant	mobile dependent, independent	fixed, mobile dependant	mobile dependent, independent	mobile dependent	mobile dependant	mobile dependent
Shared context sources	human factors	task-specific	personal, task-specific	task-specific	task-specific	task-specific	personal	task-specific	task-specific	task-specific, social factors	task-specific
	environmental factors		physical	digital	physical	physical	digital	digital	digital	physical	
	collaborative factors		presence		presence	progression		presence	presence	presence	
Level of user actuation	passive-view	on-site, remote	on-site, remote	on-site, remote	on-site, remote	on-site, remote	on-site, remote	on-site, remote	on-site, remote		on-site, remote
	interact / explore	on-site	on-site	on-site	on-site	on-site, remote	on-site	on-site, remote	on-site, remote	on-site, remote	on-site
	share / create	remote	remote	remote	remote	remote	remote	on-site, remote	on-site, remote	none	remote
	symmetry	asymmetric	asymmetric	asymmetric	asymmetric	symmetric, asymmetric	asymmetric	asymmetric	asymmetric	symmetric	asymmetric
Output and augmentation	channel	vision, audition	vision, audition	vision, audition	vision, audition	vision, audition	vision, audition	vision, audition	vision, audition, kinesthetics	vision	vision, audition
	mode	redundant	redundant	redundant	redundant, complementary	redundant, complementary	redundant, complementary	redundant	redundant	complementary	Redundant
	customization				adaptable	adaptable	adaptable				
Input modalities	channel	vision, audition	vision, audition	vision, audition	vision, audition	vision, audition	vision, audition	vision, audition	vision, audition, kinesthetics, touch	vision	vision, audition
	mode	redundant	redundant	redundant	redundant, complementary	redundant, complementary	redundant, complementary	redundant	redundant	complementary	Redundant
	customization				adaptable	adaptable	adaptable			adaptable	
Research	domain	assistance	assistance	industry	assistance	assistance	industry	assistance	assistance	others	Industrial
	context	basic research	basic research	applied research	basic research	basic research	applied research	basic research	basic research	basic research	basic research
	study type	formal	formal	formal	formal	formal	formal	formal	formal	informal	pilot

REFERENCES

[1] D. J. Wood and B. Gray, "Toward a comprehensive theory of collaboration," *J. Appl. Behav. Sci.*, vol. 27, no. 2, pp. 139–162, 1991.

[2] L. G. Terveen, "Overview of human-computer collaboration," *Knowl. Based Syst.*, vol. 8, pp. 67–81, 1995.

[3] P. J. Thomas, *CSCW Requirements and Evaluation*. London, U.K.: Springer, 1996.

[4] S. Kim, M. Billinghurst, C. Lee, and G. A. Lee, "Using freeze frame and visual notifications in an annotation drawing interface for remote collaboration," *Trans. Internet Inf. Syst.*, vol. 12, pp. 6034–6056, 2018.

[5] S. Kim, M. Billinghurst, and G. Lee, "The effect of collaboration styles and view independence on video-mediated remote collaboration," *Comput. Supported Cooperative Work*, vol. 27, no. 3, pp. 569–607, 2018.

[6] S. Kim, G. Lee, M. Billinghurst, and W. Huang, "The combination of visual communication cues in mixed reality remote collaboration," *J. Multimodal User Interfaces*, vol. 14, no. 4, pp. 321–335, 2020.

[7] S. Kim, M. Billinghurst, and K. Kim, "Multimodal interfaces and communication cues for remote collaboration," *J. Multimodal User Interfaces*, vol. 14, no. 4, pp. 313–319, 2020.

[8] E. Arias, H. Eden, G. Fischer, A. Gorman, and E. Scharff, “Transcending the individual human mind-creating shared understanding through collaborative design,” *ACM Trans. Comput.-Hum. Interact.*, vol. 7, no. 1, pp. 84–113, 2000.
[9] B. Ens *et al.*, “Revisiting collaboration through mixed reality: The evolution of groupware,” *Int. J. Hum.-Comput. Stud.*, vol. 131, pp. 81–98, 2019.
[10] J. Grudin and S. Poltrock, *The Encyclopedia of Human-Computer Interaction*, 2nd Ed. Aarhus, Denmark: Interaction-Design.org Foundation, 2013.
[11] M. Sereno, X. Wang, L. Besancon, M. J. Mcguffin, and T. Isenberg, “Collaborative work in augmented reality: A survey,” *IEEE Trans. Vis. Comput. Graph.*, to be published, doi: 10.1109/TVCG.2020.3032761.
[12] M. Billinghurst, A. Clark, and G. Lee, “A survey of augmented reality,” *Found. Trends Hum.–Comput. Interact.*, vol. 8, no. 2–3, pp. 73–272, 2015.
[13] S. K. Ong, M. L. Yuan, and A. Y. C. Nee, “Augmented reality applications in manufacturing: A survey,” *Int. J. Prod. Res.*, vol. 46, pp. 2707–2742, 2008.
[14] X. Wang, S. K. Ong, and A. Y. C. Nee, “A comprehensive survey of augmented reality assembly research,” *Adv. Manuf.*, vol. 4, no. 1, pp. 1–22, 2016.
[15] M. Hall, C. A. McMahon, P. Bermell-Garcia , A. Johansson, and R. Ravindranath, “Capturing synchronous collaborative design activities: A state-of-the-art technology review,” in *Proc. Int. Des. Conf.*, 2018, pp. 347–358.
[16] D. Aschenbrenner *et al.*, “Comparing human factors for augmented reality supported single and cooperative repair operations of industrial robots,” *Front. Robot. AI*, vol. 6, 2019, Art. no. p. 37.
[17] T. Madeira, B. Marques, J. Alves, P. Dias, and B. S. Santos, “Exploring annotations and hand tracking in augmented reality for remote collaboration,” in *Proc. Int. Conf. Hum. Syst. Eng. Des.*, 2021, pp. 83–89.
[18] B. Marques, S. Silva, A. Rocha, P. Dias, and B. S. Santos, “Remote asynchronous collaboration in maintenance scenarios using augmented reality and annotations,” in *Proc. IEEE Conf. Virtual Reality 3D User Interfaces Abstr. Workshops*, 2021, pp. 567–568.
[19] Y. Altug and A. M. Mahdy, “A perspective on distributed and collaborative augmented reality,” *Int. J. Recent Trends Hum. Comput. Interact.*, vol. 7, no. 2, pp. 23–41, 2016.
[20] 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 Trans. Vis. Comput. Graph.*, vol. 24, no. 11, pp. 2947–2962, Nov. 2018.
[21] E. Bottani and G. Vignali, “Augmented reality technology in the manufacturing industry: A review of the last decade,” *IISE Trans.*, vol. 51, no. 3, pp. 284–310, 2019.
[22] F. Bruno, L. Barbieri, E. Marino, M. Muzzupappa, L. D’Oriano, and B. Colacino, “An augmented reality tool to detect and annotate design variations in an industry 4.0 approach,” *Int. J. Adv. Manuf. Technol.*, vol. 105, no. 1, pp. 875–887, 2019.
[23] B. Marques, A. Teixeira, S. Silva, J. Alves, P. Dias, and B. S. Santos, “A conceptual model for data collection and analysis for AR-based remote collaboration evaluation,” in *Proc. IEEE Int. Symp. Mixed Augmented Reality*, 2020, pp. 1–2.
[24] B. Marques, S. Silva, P. Dias, and B. S. Santos, “An ontology for evaluation of remote collaboration using augmented reality,” in *Proc. Eur. Conf. Comput.-Supported Cooperative Work*, 2021, pp. 1–8.
[25] L. Merino, M. Schwarzl, M. Kraus, M. Sedlmair, D. Schmalstieg, and D. Weiskopf, “Evaluating mixed and augmented reality: A systematic literature review (2009–2019),” in *Proc. IEEE Int. Symp. Mixed Augmented Reality*, 2020, pp. 438–451.
[26] K. Petersen, S. Vakkalanka, and L. Kuzniarz, “Guidelines for conducting systematic mapping studies in software engineering: An update,” *Inf. Softw. Technol.*, vol. 64, pp. 1–18, 2015.
[27] C. A. Collazos, F. L. Gutiérrez, J. Gallardo, M. Ortega, H. M. Fardoun, and A. I. Molina, “Descriptive theory of awareness for groupware development,” *J. Ambient Intell. Humanized Comput.*, vol. 10, no. 12, pp. 4789–4818, 2019.
[28] P. T. Sukumar *et al.*, “Transparency in qualitative research: Increasing fairness in the CHI review process,” in *Proc. Extended Abstr. CHI Conf. Hum. Factors Comput. Syst.*, 2020, pp. 1–6.
[29] M. Meyer and J. Dykes, “Criteria for rigor in visualization design study,” *IEEE Trans. Vis. Comput. Graph.*, vol. 26, no. 1, pp. 87–97, Jan. 2020.
[30] A. F. Hadwin, C. L. Gress, and J. Page, “Toward standards for reporting research: A review of the literature on computer-supported collaborative learning,” in *Proc. 6th IEEE Int. Conf. Adv. Learn. Technol.*, 2006, pp. 1007–1011.
[31] M. Sedlmair, M. Meyer, and T. Munzner, “Design study methodology: Reflections from the trenches and the stacks,” *IEEE Trans. Vis. Comput. Graph.*, vol. 18, no. 12, pp. 2431–2440, Dec. 2012.
[32] R. Johansen, *GroupWare: Computer Support for Business Teams*. New York, NY, USA: Free Press, 1988.
[33] T. T. H. Nguyen and T. Duval, “A survey of communication and awareness in collaborative virtual environments,” in *Proc. Int. Workshop Collaborative Virtual Environ.*, 2014, pp. 1–8.
[34] X. Wang and P. S. Dunston, “Groupware concepts for augmented reality mediated human-to-human collaboration,” in *Proc. Joint Int. Conf. Comput. Decis. Making Civil Building Eng.*, 2006, pp. 1836–1842.
[35] C. A. Ellis, S. J. Gibbs, and G. Rein, “Groupware: Some issues and experiences,” *Commun. ACM*, vol. 34, no. 1, pp. 39–58, 1991.
[36] D. Gergle, R. E. Kraut, and S. R. Fussell, “Using visual information for grounding and awareness in collaborative tasks,” *Hum.–Comput. Interact.*, vol. 28, no. 1, pp. 1–39, 2013.
[37] R. T. Azuma, “A survey of augmented reality,” *Presence, Teleoperators Virtual Environ.*, vol. 6, no. 4, pp. 355–385, 1997.
[38] R. Azuma, Y. Baillot, R. Behringer, S. Feiner, S. Julier, and B. Macintyre, “Recent advances in augmented reality,” *IEEE Comput. Graph. Appl.*, vol. 21, no. 6, pp. 34–47, Nov./Dec. 2001.
[39] C. Sandor *et al.*, “Breaking the barriers to true augmented reality,” 2015, *arXiv:1512.05471*,.
[40] U. Neumann and A. Majoros, “Cognitive, performance, and systems issues for augmented reality applications in manufacturing and maintenance,” in *Proc. IEEE Virtual Reality Annu. Int. Symp.*, 1998, pp. 4–11.
[41] S. Henderson, S. Member, and S. Feiner, “Exploring the benefits of augmented reality documentation for maintenance and repair,” *IEEE Trans. Vis. Comput. Graph.*, vol. 17, no. 10, pp. 1355–1368, Oct. 2011.
[42] W. Li, A. Y. C. Nee, and S. K. Ong, “A state-of-the-art review of augmented reality in engineering analysis and simulation,” *Multimodal Technol. Interact.*, vol. 1, no. 3, 2017, Art. no. 17.
[43] J. Jetter, J. Eimecke, and A. Rese, “Augmented reality tools for industrial applications: What are potential key performance indicators and who benefits?,” *Comput. Hum. Behav.*, vol. 87, pp. 18–33, 2018.
[44] M. Quandt, B. Knoke, C. Gorldt, M. Freitag, and K. D. Thoben, “General requirements for industrial augmented reality applications,” *Procedia CIRP*, vol. 72, pp. 1130–1135, 2018.
[45] P. Wang *et al.*, “A comprehensive survey of AR/MR-based co-design in manufacturing,” *Eng. Comput.*, vol. 36, pp. 1715–1738, 2020.
[46] S. Lukosch, M. Billinghurst, L. Alem, and K. Kiyokawa, “Collaboration in augmented reality,” in *Proc. Comput. Supported Cooperative Work*, vol. 24, no. 6, 2015, pp. 515–525.
[47] A. Dey, M. Billinghurst, R. W. Lindeman, and J. E. S. Ii, “A systematic review of 10 years of augmented reality usability studies: 2005 to 2014 ”, *Front. Robot. AI*, vol. 5, 2018, Art. no. 37.
[48] H. T. Regenbrecht, M. T. Wagner, and G. Baratoff, “MagicMeeting: A collaborative tangible augmented reality system,” *Virtual Reality*, vol. 6, pp. 151–166, 2002.
[49] H. Jalo, H. Pirkkalainen, O. Torro, H. Kärkkäinen, J. Puhto, and T. Kankaanpää, “How can collaborative augmented reality support operative work in the facility management industry?,” in *Proc. 10th Int. Joint Conf. Knowl. Discov., Knowl. Eng. Knowl. Manag.*, 2018, pp. 41–51.
[50] P. Gurevich, J. Lanir, and B. Cohen, “Design and implementation of TeleAdvisor: A projection-based augmented reality system for remote collaboration,” *Comput. Supported Cooperative Work: Int. J.*, vol. 24, no. 6, pp. 527–562, 2015.
[51] R. Palmarini, J. A. Erkoyuncu, R. Roy, and H. Torabmostaedi, “A systematic review of augmented reality applications in maintenance,” *Robot. Comput.-Integr. Manuf.*, vol. 49, pp. 215–228, 2018.
[52] 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 Conf. Hum. Factors Comput. Syst.*, 2019, pp. 1–14.

[53] F. Lamberti, F. Manuri, A. Sanna, G. Paravati, P. Pezzolla, and P. Montuschi, "Challenges, opportunities, and future trends of emerging techniques for augmented reality-based maintenance," *IEEE Trans. Emerg. Top. Comput.*, vol. 2, no. 4, pp. 411–421, Dec. 2014.
[54] M. Augstein and T. Neumayr, "A human-centered taxonomy of interaction modalities and devices," *Interacting Comput.*, vol. 31, no. 1, pp. 27–58, 2019.
[55] R. C. Nickerson, U. Varshney, and J. Muntermann, "A method for taxonomy development and its application in information systems," *Eur. J. Inf. Syst.*, vol. 22, pp. 336–359, 2013.
[56] P. Milgram, H. Takemura, A. Utsumi, and F. Kishino, "Augmented reality: A class of displays on the reality-virtuality continuum," in *Proc. Telemanipulator Telepresence Technol.*, 1995, pp. 282–292.
[57] W. E. Mackay, "Augmented reality: Linking real and virtual worlds A new paradigm for interacting with computers," in *Proc. Working Conf. Adv. Vis. Interfaces*, 1998, pp. 13–21.
[58] R. Suomela and J. Lehikoinen, "Taxonomy for visualizing location-based information," *Virtual Reality*, vol. 8, no. 2, pp. 71–82, Jun. 2004.
[59] R. W. Lindeman and H. Noma, "A classification scheme for multi-sensory augmented reality," in *Proc. ACM Symp. Virtual Reality Softw. Technol.*, 2007, pp. 175–178.
[60] J. M. Braz and J. M. Pereira, "TARCAST: Taxonomy for augmented reality CASTing with web support," *Int. J. Virtual Reality*, vol. 7, no. 4, pp. 47–56, 2008.
[61] M. Tönnis and D. A. Plecher, "Principles in augmented reality – Classification and categorization guidelines version 1.0," Technical University of Munchen, Munich, Germany, *Tech. Rep.*, 2011.
[62] O. Hugues, P. Fuchs, and O. Nannipieri, "New augmented reality taxonomy: Technologies and features of augmented environment," in *Proc. Handbook Augmented Reality*, 2011, pp. 47–63.
[63] J.-M. Normand, M. Servières, and G. Moreau, "A new typology of augmented reality applications," in *Proc. 3rd Augmented Hum. Int. Conf*, 2012, pp. 1–8.
[64] S. Benford, C. Greenhalgh, G. Reynard, C. Brown, and B. Koleva, "Understanding and constructing shared spaces with mixed-reality boundaries," *ACM Trans. Comput.-Hum. Interact.*, vol. 5, no. 3, pp. 185–223, 1998.
[65] M. Billinghurst and H. Kato, "Collaborative augmented reality," *Commun. ACM*, vol. 45, pp. 64–70, 2002.
[66] T. Brockmann, N. Krueger, S. Stieglitz, and I. Bohlsen, "A framework for collaborative augmented reality applications," in *Proc. 19th Americas Conf. Inf. Syst.*, 2013, pp. 1–10.
[67] M. Speicher, B. D. Hall, and M. Nebeling, "What is mixed reality?," in *Proc. CHI Conf. Hum. Factors Comput. Syst.*, 2019, p. 1–15.
[68] R. A. J. de Belen, H. Nguyen, D. Filonik, D. D. Favero, and T. Bednarz, "A systematic review of the current state of collaborative mixed reality technologies: 2013–2018," *AIMS Electron. Elect. Eng.*, vol. 3, pp. 181–223, 2019.
[69] K. Halskov and N. B. Hansen, "The diversity of participatory design research practice at PDC 2002–2012," *Int. J. Hum.-Comput. Stud.*, vol. 74, pp. 81–92, 2015.
[70] C. M. Barnum, *Usability Testing Essentials: Ready, Set...Test!*, 1st ed. Burlington, MA, USA: Morgan Kaufmann Publishers Inc., 2010.
[71] J. A. Jacko, *Human-Computer Interaction Handbook: Fundamentals, Evolving Technologies, and Emerging Applications*, 3rd Ed. New York, NY, USA: CRC Press, 2012.
[72] S. Zollmann, T. Langlotz, R. Grasset, W. H. Lo, S. Mori, and H. Regenbrecht, "Visualization techniques in augmented reality: A taxonomy, methods and patterns," *IEEE Trans. Vis. Comput. Graph.*, vol. 27, no. 9, pp. 3808–3825, Sep. 2021.
[73] M. A. Teruel, E. Navarro, V. López-Jaquero, F. Montero, and P. González, "A comprehensive framework for modeling requirements of cscw systems," *J. Softw., Evol. Process*, vol. 29, no. 5, 2017, Art. no. e1858.
[74] P. Dourish and V. Bellotti, "Awareness and coordination in shared workspaces," in *Proc. ACM Conf. Comput.-Supported Cooperative Work*, 1992, pp. 107–114.
[75] M. Cidota, S. Lukosch, D. Datcu, and H. Lukosch, "Comparing the effect of audio and visual notifications on workspace awareness using head-mounted displays for remote collaboration in augmented reality," *Augmented Hum. Res.*, vol. 1, no. 1, 2016, Art. no. 1.
[76] H. H. Clark and S. E. Brennan, "Grounding in communication," in *Proc. Perspectives Socially Shared Cogn.*, 1991, pp. 127–149.
[77] H. Patel, M. Pettitt, and J. R. Wilson, "Factors of collaborative working: A framework for a collaboration model," *Appl. Ergonom.*, vol. 43, no. 1, pp. 1–26, 2012.
[78] M. Billinghurst, M. Cordeil, A. Bezerianos, and T. Margolis, "Collaborative immersive analytics," in *Proc. Immersive Anal.* Springer, 2018, pp. 221–257.
[79] S. Kim, G. Lee, N. Sakata, and M. Billinghurst, "Improving co-presence with augmented visual communication cues for sharing experience through video conference," *IEEE Int. Symp. Mixed Augmented Reality*, 2014, pp. 83–92.
[80] M. Norman, G. A. Lee, R. T. Smith, and M. Billingurst, "The impact of remote user's role in a mixed reality mixed presence system," in *Proc. Int. Conf. Virtual-Reality Continuum Appl. Ind.*, 2019, pp. 1–9.
[81] C. P. Lee and D. Paine, "From the matrix to a model of coordinated action (MoCA): A conceptual framework of and for CSCW," in *Proc. 18th ACM Conf. Comput. Supported Cooperative Work Social Comput.*, 2015, pp. 179–194.
[82] D. Stokols, S. Misra, R. P. Moser, K. L. Hall, and B. K. Taylor, "The ecology of team science: Understanding contextual influences on transdisciplinary collaboration," *Amer. J. Prev. Med.*, vol. 35, no. 2, pp. 96–115, 2008.
[83] C. A. Bolstad and M. R. Endsley, "Choosing team collaboration tools: Lessons from disaster recovery efforts," *Ergonom. Des.*, vol. 13, no. 4, pp. 7–14, 2005.
[84] J. L. Wildman, A. L. Thayer, M. A. Rosen, E. Salas, J. E. Mathieu, and S. R. Rayne, "Task types and team-level attributes: Synthesis of team classification literature," *Hum. Resour. Develop. Rev.*, vol. 11, no. 1, pp. 97–129, 2012.
[85] K. J. Ostergaard and J. D. Summers, "Development of a systematic classification and taxonomy of collaborative design activities," *J. Eng. Des.*, vol. 20, no. 1, pp. 57–81, 2009.
[86] T. Yoshioka, G. Herman, J. Yates, and W. Orlikowski, "Genre taxonomy: A knowledge repository of communicative actions," *ACM Trans. Inf. Syst.*, vol. 19, no. 4, pp. 431–456, 2001.
[87] W. Huang, S. Kim, M. Billinghurst, and L. Alem, "Sharing hand gesture and sketch cues in remote collaboration," *J. Vis. Commun. Image Representation*, vol. 58, pp. 428–438, 2019.
[88] P. Antunes, V. Herskovic, S. F. Ochoa, and J. A. Pino, "Reviewing the quality of awareness support in collaborative applications," *J. Syst. Softw.*, vol. 89, pp. 146–169, 2014.
[89] I. F. del Amo, J. A. Erkoyuncu, R. Roy, R. Palmarini, and D. Onoufriou, "A systematic review of augmented reality content-related techniques for knowledge transfer in maintenance applications," *Comput. Ind.*, vol. 103, pp. 47–71, 2018.
[90] J. Grubert, T. Langlotz, S. Zollmann, and H. Regenbrecht, "Towards pervasive augmented reality: Context- awareness in augmented reality," *IEEE Trans. Vis. Comput. Graph.*, vol. 23, no. 6, pp. 1706–1724, Jun. 2017.
[91] T. Olsson, P. Jarusriboonchai, P. Woźniak, S. Paasovaara, K. Väänänen, and A. Lucero, "Technologies for enhancing collocated social interaction: Review of design solutions and approaches," *Comput. Supported Cooperative Work*, 2019, pp. 1–55.
[92] J. E. Fischer, S. Reeves, B. Brown, and A. Lucero, "Beyond "same time, same place": Introduction to the special issue on collocated interaction," *Hum.-Comput. Interact.*, vol. 33, no. 5–6, pp. 305–310, 2018.
[93] S. D. Scott, T. Graham, J. R. Wallace, M. Hancock, and M. Nacenta, "Local remote collaboration: Applying remote group awareness techniques to co-located settings," in *Proc. 18th ACM Conf. Companion Comput. Supported Cooperative Work Social Comput.*, 2015, pp. 319–324.
[94] K. Kreijns, P. A. Kirschner, and M. Vermeulen, "Social aspects of CSCL environments: A research framework," *Educ. Psychol.*, vol. 48, no. 4, pp. 229–242, 2013.
[95] B. Yoon, H.-i. Kim, G. A. Lee, M. Billinghurst, and W. Woo, "The effect of avatar appearance on social presence in an augmented reality remote collaboration," in *Proc. IEEE Conf. Virtual Reality 3D User Interfaces*, 2019, pp. 547–556.
[96] T.-Y. Wang, Y. Sato, M. Otsuki, H. Kuzuoka, and Y. Suzuki, "Effect of full body avatar in augmented reality remote collaboration," in *Proc. IEEE Conf. Virtual Reality 3D User Interfaces*, 2019, pp. 1221–1222.
[97] M. R. Miller, H. Jun, F. Herrera, J. Y. Villa, G. Welch, and J. N. Bailenson, "Social interaction in augmented reality," *PloS One*, vol. 14, no. 5, 2019, Art. no. e0216290.

[98] A. Irlitti, T. Piumsomboon, D. Jackson, and B. H. Thomas, "Conveying spatial awareness cues in xR collaborations," *IEEE Trans. Vis. Comput. Graph.*, vol. 25, no. 11, pp. 3178–3189, Nov. 2019.

[99] J. Müller, R. Rädle, and H. Reiterer, "Remote collaboration with mixed reality displays: How shared virtual landmarks facilitate spatial referencing," in *Proc. CHI Conf. Hum. Factors Comput. Syst.*, 2017, pp. 6481–6486.

[100] D. Wang *et al.*, "Coordination breakdowns in nuclear power plant control rooms: Cause identification and behaviour-sequence analysis," *Ergonomics*, vol. 63, no. 6, pp. 660–681, 2020.

[101] J. Schmutz, F. Hoffmann, E. Heimberg, and T. Manser, "Effective coordination in medical emergency teams: The moderating role of task type," *Eur. J. Work Organizational Psychol.*, vol. 24, no. 5, pp. 761–776, 2015.

[102] M. She and Z. Li, "Design and evaluation of a team mutual awareness toolkit for digital interfaces of nuclear power plant context," *Int. J. Hum.-Comput. Interact.*, vol. 33, no. 9, pp. 744–755, 2017.

[103] P. Isenberg, N. Elmqvist, J. Scholtz, D. Cernea, K.-L. Ma, and H. Hagen, "Collaborative visualization: Definition, challenges, and research agenda," *Inf. Vis.*, vol. 10, no. 4, pp. 310–326, 2011.

[104] B. C. Kwon, H. Kim, E. Wall, J. Choo, H. Park, and A. Endert, "AxiSketcher: Interactive nonlinear axis mapping of visualizations through user drawings," *IEEE Trans. Vis. Comput. Graph.*, vol. 23, no. 1, pp. 221–230, Jan. 2017.

[105] L. Chittaro and F. Buttussi, "Assessing knowledge retention of an immersive serious game vs. a traditional education method in aviation safety," *IEEE Trans. Vis. Comput. Graph.*, vol. 21, no. 4, pp. 529–538, Apr. 2015.

[106] J. U. Kwon, J.-I. Hwang, J. Park, and S. C. Ahn, "Fully asymmetric remote collaboration system," *IEEE Access*, vol. 7, pp. 54155–54166, 2019.

[107] A. M. Fernandes and P. B. Albuquerque, "Tactual perception: A review of experimental variables and procedures," *Cogn. Process.*, vol. 13, no. 4, pp. 285–301, 2012.

[108] T. Narumi, S. Nishizaka, T. Kajinami, T. Tanikawa, and M. Hirose, "Meta cookie+: An illusion-based gustatory display," in *Proc. Int. Conf. Virtual Mixed Reality*, 2011, pp. 260–269.

[109] T. Olalere, "Methodology in accounting research: A critique of taxonomy," *Social Sci. Res. Netw.*, pp. 1–40, 2011.

[110] R. M. Carvalho, R. M. de C. Andrade, K. M. de Oliveira, I. de S. Santos, and C. I. M. Bezerra, "Quality characteristics and measures for human–computer interaction evaluation in ubiquitous systems," *Softw. Qual. J.*, vol. 25, no. 3, pp. 743–795, 2017.

[111] M. Usman, R. Britto, J. Börstler, and E. Mendes, "Taxonomies in software engineering: A systematic mapping study and a revised taxonomy development method," *Inf. Softw. Technol.*, vol. 85, pp. 43–59, 2017.

[112] D. Şmite, C. Wohlin, Z. Galviundefineda, and R. Prikladnicki, "An empirically based terminology and taxonomy for global software engineering," *Empirical Softw. Eng.*, vol. 19, no. 1, p. 105–153, 2014.

[113] 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 Trans. Vis. Comput. Graph.*, vol. 22, no. 11, pp. 2413–2422, Nov. 2016.

[114] P. Wang *et al.*, "A gesture- and head-based multimodal interaction platform for MR remote collaboration," *Int. J. Adv. Manuf. Technol.*, vol. 105, no. 7, pp. 3031–3043, 2019.

[115] S. Kim, G. Lee, W. Huang, H. Kim, W. Woo, and M. Billinghurst, "Evaluating the combination of visual communication cues for HMD-based mixed reality remote collaboration," in *Proc. CHI Conf. Human Factors Comput. Syst.*, 2019, pp. 1–13.

[116] D. Aschenbrenner *et al.*"Comparing different augmented reality support applications for cooperative repair of an industrial robot," in *Proc. IEEE Int. Symp. Mixed Augmented Reality Adjunct*, 2018, pp. 69–74.

[117] T. Teo, A. F. Hayati, G. A. Lee, M. Billinghurst, and M. Adcock, "A technique for mixed reality remote collaboration using 360 panoramas in 3D reconstructed scenes," in *Proc. ACM Symp. Virtual Reality Soft. Technol.*, 2019, pp. 1–11.

[118] T. Piumsomboon, A. Dey, B. Ens, G. Lee, and M. Billinghurst, "CoVAR: Mixed-platform remote collaborative augmented and virtual realities system with shared collaboration cues," *Int. Symp. Mixed Augmented Reality*, 2017, pp. 218–219.

[119] T. Piumsomboon, A. Dey, B. Ens, G. Lee, and M. Billinghurst, "The effects of sharing awareness cues in collaborative mixed reality," *Front. Robot. AI*, vol. 6, pp. 1–18, 2019.

[120] F. Obermair *et al.*, "Maintenance with augmented reality remote support in comparison to paper-based instructions: Experiment and analysis," in *Proc. IEEE Int. Conf. Ind. Eng. Appl.*, 2020, pp. 942–947.

Bernardo Marques received the master's degree in computer and telematics engineering in 2016 from the University of Aveiro, Portugal, where he is currently working toward the PhD degree in computer engineering. His research interests include human-centered technologies, with a focus on computer-supported cooperative work, remote collaboration, and virtual and augmented reality.

Samuel Silva received the MSc degree in electronics and telecommunication engineering and the PhD degree in informatics engineering from the University of Aveiro, in 2007 and 2012, respectively. He is currently a researcher with the Institute of Electronics and Informatics Engineering of Aveiro. His research interests include human-centered technologies, with a focus on multimodal interaction.

João Alves received the master's degree in computer and telematics engineering in 2017 from the University of Aveiro, Portugal, where he is currently working toward the PhD degree in computer engineering. His research interests include virtual and augmented reality, robotics, deep learning, and data and information visualization.

Tiago Araújo received the master's degree from the Laboratory of Visualization, Interaction, and Intelligent Systems, Federal University of Pará, where he is currently working toward the PhD degree. His research interests include data and information visualization, virtual and augmented reality, and computer vision.

Paulo Dias is currently an assistant professor with the Department of Electronics, Telecommunications, and Informatics, University of Aveiro and is involved in several works and projects within IEETA related to virtual and augmented reality, computer vision, computer graphics, 3d reconstruction, visualization, combination, and fusion of data from multiple sensors.

Beatriz Sousa Santos (Senior Member, IEEE) is currently an associate professor with the Department of Electronics, Telecommunications, and Informatics, University of Aveiro, and a researcher with the Institute of Electronics and Informatics Engineering of Aveiro. Her research interests include virtual and augmented reality and data and information visualization.

▷ **For more information on this or any other computing topic, please visit our Digital Library at** www.computer.org/csdl.