Survey of Annotations in Extended Reality Systems

Zahra Borhani, *Graduate Student Member, IEEE*, Prashast Sharma, and Francisco R. Ortega

(Survey Paper)

***Abstract*—Annotation in 3D user interfaces such as Augmented Reality (AR) and Virtual Reality (VR) is a challenging and promising area; however, there are not currently surveys reviewing these contributions. In order to provide a survey of annotations for Extended Reality (XR) environments, we conducted a structured literature review of papers that used annotation in their AR/VR systems from the period between 2001 and 2021. Our literature review process consists of several filtering steps which resulted in 103 XR publications with a focus on annotation. We classified these papers based on the display technologies, input devices, annotation types, target object under annotation, collaboration type, modalities, and collaborative technologies. A survey of annotation in XR is an invaluable resource for researchers and newcomers. Finally, we provide a database of the collected information for each reviewed paper. This information includes applications, the display technologies and its annotator, input devices, modalities, annotation types, interaction techniques, collaboration types, and tasks for each paper. This database provides a rapid access to collected data and gives users the ability to search or filter the required information. This survey provides a starting point for anyone interested in researching annotation in XR environments.**

***Index Terms*—Annotation, augmented reality, extended reality, immersive technologies, virtual reality.**

I. Introduction

The development of immersive technologies such as Augmented Reality (AR), Virtual reality (VR), and Mixed Reality (MR) offer a new way to interact with content and experience the world. AR, VR, and MR are three different technologies that provide different degrees of immersion and interactivity, from overlaying digital objects on the real world with AR, to fully immersing the user in a virtual world with VR, and combining elements of both with MR. Extended Reality (XR), which includes AR and VR, provides opportunities to overcome the limits of traditional systems such as paper and 2D screens [18]. It has also shown that using AR technology compared to 2D screens can increase work efficiency by decreasing user response time and cognitive load [106], [143]. The importance of this new technology is emphasized when a task needs users to move or to change their viewing angle [106].

XR enables the system to overlay digitally created information and instructions using annotations in a real or virtual environment. In other words, annotation is a way for the users to annotate physical or virtual entities with digital information [125] to provide more information about the world around them for the user [28]. Using annotations have many benefits, such as facilitating communication and collaboration [42], [110], [156], increasing the user's understanding of the world around them [28], [159], increasing the quality of analysis [25], and to emphasize core messages or specific object [111].

In addition, generating annotations can benefit from new advances in immersive technologies such as eye gaze, speech, and hand gestures [19], [97]. On the other hand, although many previous studies confirmed the benefits of using XR annotations, some studies have also pointed out that redundancy in AR instructions may cause a negative effect on user performance and depth perception due to an increase in user cognitive load [98], [148]. These reasons highlight the importance of having detailed knowledge of annotation in XR to avoid redundancy and cluttering scenes.

Although annotation use has been investigated in many previous XR studies, there is not currently a survey reviewing the plethora of contributions to the annotation context. A survey of annotation in XR is a valuable resource that can help researchers in five main ways: first, it provides overall knowledge and a helpful summary of previous works on annotation in XR. Second, it introduces important concepts, technologies, and methods in annotation systems for newcomers. For this purpose, we gathered definitions and classifications from different previous studies to provide comprehensive classifications and definitions. Third, this taxonomy encourages authors to report their annotation systems in a more structured way. Four, The existence of a common framework can play an essential role in facilitating the work of developers. It can help developers to design their annotation systems. Finally, it provides researchers with the opportunity to understand the scientific gaps in this area of research.

This survey aims to answer the following research questions (RQs) by conducting a systematic literature review of previous research studies that explored annotation in XR systems:

- *RQ1:* Which type of annotations were used in XR papers? (Section III-B)
- *RQ2:* What are the display technologies and input devices for viewing and generating annotation in XR systems? (Section IV)

Manuscript received 2 October 2022; revised 4 June 2023; accepted 6 June 2023. Date of publication 23 June 2023; date of current version 1 July 2024. This work was supported in part by NSF under Grants 2327569, 2238313, 2223432, 2223459, 2106590, 2016714, 2037417, and 1948254. Recommended for acceptance by Kiyoshi Kiyokawa. *(Corresponding authors: Zahra Borhani; Francisco R. Ortega.)*

Zahra Borhani and Francisco R. Ortega are with the Colorado State University, Fort Collins, CO 80523 USA (e-mail: zahra.borhani@colostate.edu; fortega@colostate.edu).

Prashast Sharma is with the University of Florida, Gainesville, FL 32611 USA (e-mail: prashast.sharma@ufl.edu).

Digital Object Identifier 10.1109/TVCG.2023.3288869

- *RQ3:* What type of collaboration and technologies were used in collaborative XR systems that used annotation? (Section V)
- *RQ4:* What were common annotation tasks used in previous studies to evaluate annotation systems? (Section VI)
- *RQ5:* Which methods have been used for generating annotations? (Section VII)
- *RQ6:* What are the gaps and challenges in the surveyed papers? (Section VIII)

The hardware of an XR annotation system is typically a combination of output and input devices. Output devices are used for displaying the XR environment and its annotations. The input devices or modalities are used for interacting with the XR environment and creating annotations. The primary factors of these systems we investigated include annotation types, commonly used annotation tasks, and various annotation techniques.

In addition to the above concepts, we explored one of the primary use cases of annotations in previous studies, which is the role of annotation in collaborative systems. Two previous surveys on collaborative MR and AR considered annotation as a complementary and key factor of XR systems that tend to improve overall communication and collaboration [27], [126]. Their results showed that around one-third of the papers in their corpus investigated annotation techniques in MR [27]. Meanwhile, about one-third of papers in our literature review used annotation in a collaborative XR setting. This shows the importance of each of these matters for one another. So, we decided to explore the use of annotation in detail in Section V.

The organizational structure of this paper is as follows: an explanation of the systematic review method we used for collecting the articles in Section II. Section III is dedicated to classifying existing digitally created annotations based on different criteria. Then, in the fourth section, we focused on the existing output technologies and their input devices for systems that used annotation. Next, an exploration of collaborative XR systems that use annotation is discussed in Section V. Section VI covers some annotation tasks there were used commonly in human-centered experiments. Then, several annotation methods used in previous research studies are examined in Section VII. Finally, Section VIII is allocated to one of our main objectives: determining existing challenges in each area of research.

II. Review Method and Publisher Search

This survey covers 103 papers related to annotation in XR, which were chosen using our systematic investigation method, which is a combination of search, selection, filtering, and classification processes [30]. Our literature corpus is restricted to the ACM Digital Library, IEEE Xplore, and SpringerLink search engines without a year limitation. The reason these three search engines were chosen is that a surveyed paper by Duenser et al. identified these publishers as leading publishers in the AR domain [30]. Further, the papers were collected in four phases based on specific criteria: keyword, subject selection, publication venue, and context search. The repeated papers were gathered using more than one search engine and subsequently removed after locating the duplicate in our corpus. Then, the collected papers from each source were classified according to their XR display technology, input devices, annotation types, annotator, collaboration types, and modalities. Fig. 1 shows the four phases of our literature review process.

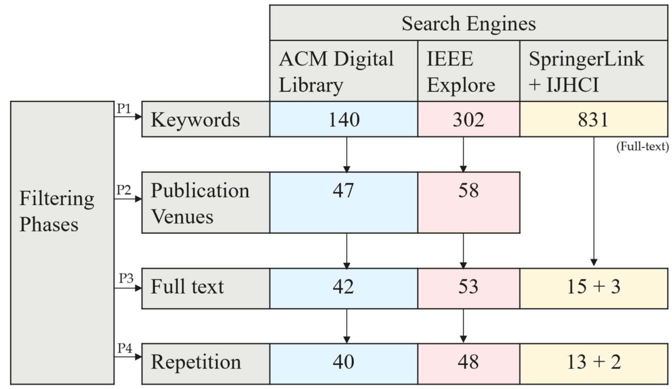

Fig. 1. Our literature review process includes four phases: Keyword selection, publication venue filtering, full-text scanning, and removing repeated publications.

TABLE I
Keyword Selection of Search Queries

Keyword Selection (only abstract and paper titles)
"annotate/tion/ing" AND "augmented reality"
"annotate/tion/ing" AND "virtual reality"
"annotate/tion/ing" AND "mixed reality"
"annotate/tion/ing" AND "extended reality"

Phase 1- Keyword Selection: The initial selection criterion for a paper to be included in this research pool was that the abstract or title must contain one of our outlined search queries. For this purpose, we searched the selected search engines using search terms covering virtual annotation in AR and VR. The search keywords for the immersive technologies included "augmented reality", "virtual reality", "mixed reality", and "extended reality". Table I shows the search queries that were used in our investigation. Papers that satisfied at least one of these search queries were included in the first step of our literature review process to be explored further.

Phase 2- Publication Venues Filtering: In the second step of the literature review process, we eliminated the research papers that were not published in XR or Human-Computer Interaction (HCI) venues while satisfying the search queries criteria. For this purpose, we selected papers only from journals, conferences, symposiums, and workshops related to HCI or XR. Our search covered various journals (IEEE TVCG, ACM TOCHI, ACM PACMHCI, IJHCS, and IJHCI), conferences (CHI, OzCHI, VRST, UIST, VR,[1] VR and 3DUI,[2] ISMAR, CSCW, UBICOMP, SUI, IUI, ICMI, 3DUI, Asian CHI, HAI, HT, ASSETS, AIVR, Web3D, and VRCAI) and workshops (ISMAR-Adjunct, CHI EA, VRW, ISMARW, and WS-REST). The distribution of surveyed papers based on their publication venue is shown in Fig. 2. As it can be seen in Fig. 2, most articles in our data pool were published in ISMAR, ISMAR-Adjunct and IEEE VR, HCI, CHI EA, UIST, CHI, HT, and OzCHI. Bar charts

[1] IEEE Virtual Reality
[2] IEEE Virtual Reality and 3D User Interfaces

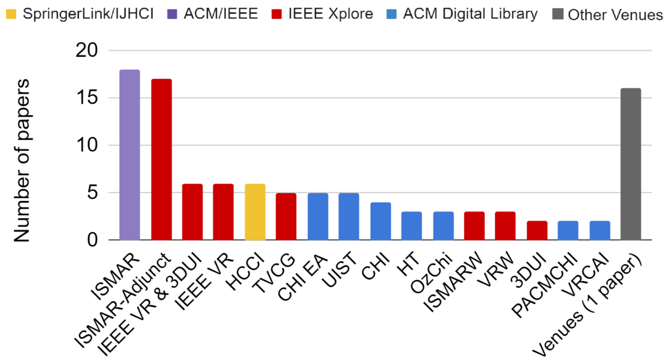

Fig. 2. Total number of publications for each venue in our corpus.

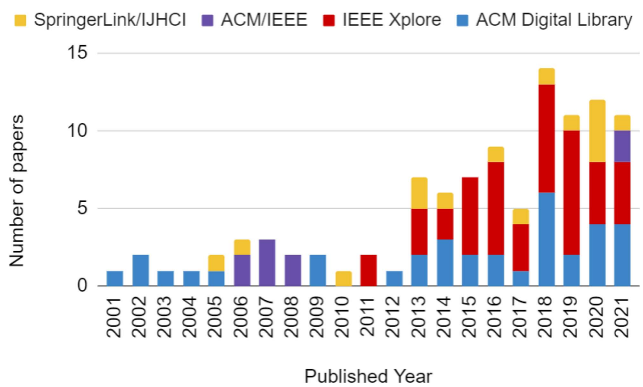

Fig. 3. Total number of publications per year in our corpus.

in Fig. 2 are distinguished with their five colors where each color shows a different search engine: ACM Digital Library (blue), IEEE VR (red), SpringerLink (yellow), both IEEE VR/ACM digital Library (purple), and papers belonging to any of three search engines (gray).

Phase 3- Full-text screening: In our third filtering phase, we searched the entire text of each paper that satisfied our criteria of search keywords and publication venues. Then, we removed articles that were not used annotation in their system or their introduced system did not fall in our XR classifications. In addition, we removed survey and taxonomy papers that did not introduce a new system (as required by our established criteria) [26].

Phase 4- Repetition Filtering: In our final step, we discarded 10 repeated papers. For example, some of the ISMAR papers were found in both the ACM Digital Library and IEEE explore. As a result, we found 103 papers in the final phase.

In total, our systematic investigation method resulted in 103 papers that are related to annotation in XR. Fig. 3 shows the distribution of these 103 surveyed papers between 2001 and 2021. Although we did not restrict our search to any specific time period, we had not found any published works before 2000 that satisfied our selection criteria. In our paper pool, most articles were published after 2013 (see Fig. 3). One of the reasons behind this phenomena is that the major XR products were first sold to the general market in 2016. Finally, We also provide a database application[3] for readers to explore the papers in more detail.

[3]https://tamgef.com/publications

III. Annotation

A typical behavior of individuals interacting with complex data and information is externalizing thoughts by producing annotations such as highlighting helpful information, note-taking, recording the hypotheses, and drawing graphical signs or lines [62]. Annotations have been employed in 2D displays vastly, and many previous studies confirmed their benefits, such as facilitating communication and collaboration [62], increasing the quality of analysis [21], [161], facilitating the review process [62], helping presenters to convey key points [112] or to emphasize core messages or specific data [68], and improving information comprehension and memorizing [153], and eventually making the process more enjoyable [22].

Although studying annotations in 2D environments can be beneficial for understanding annotation in XR environments, our research focuses only on systems that use annotation in XR. Using annotation in XR environments has offered promises across a wide variety of domains such as surgery [70], [149], repair [110], maintenance (computer) [84], [91], inspection [40], physics [91], [110], emphatic computing [155], physiotherapy [36], [52], multimedia hypertext [88], modeling of industrial systems (water treatment) [141], and 3D modeling [140].

A. Annotation Definition

Many previous studies considered "annotation" as a mere text that is added to an existing object [38]. Fonnet et al. included graphical annotations in their survey paper classification and defined annotations as any graphical or textual virtual information added to the visualization [38]. Still, these definitions are very narrow for XR environments and don't encompass a broad range of use cases. Wither et. al. defined annotation as any virtual information that is registered to an existing object and describes it. This definition implies that annotation can be represented in different forms, such as text, pictures, models, sounds, or even haptic feedback [152]. Ren et al. considered highlighting as another form of annotation [112]. Furthermore, Yamada et al. proposed a novel system that uses hand gestures as a new annotation for their system [156].

We adopted the definition by Wither et al. which mentioned that an annotation task covers interactions allowing users to add any form of virtual information to the existing virtual or real objects in order to describe and provide more information about the target [152]. The target object might vary from a real object [143], [162] or virtual object in the scene [25], [103] to a room full of objects or even the whole XR scene [155]. This target object induces a spatial dependence for the annotations describing it. However, the representation of the annotation itself is not spatially dependent. This definition allows annotations to be presented in different forms, such as text, hand drawings, images, models, audio, video, or even haptic feedback. We elaborate on different forms of annotation in previous XR studies in Section III-B1 and different target types in Section III-D.

B. Classification of Annotations

Using annotations in XR, especially in immersive environments, is a challenging problem that requires more investigation.

TABLE II
THE CLASSIFICATION OF ANNOTATION IN XR - SOME EXAMPLE REFERENCES ARE PROVIDED IN FRONT OF EACH ANNOTATION TYPE

Annotation Classifications		
Annotation Forms		**Paper Examples**
Hand Drawing	Graphical Drawing	[40, 115, 150]
	Textual Drawing	[25, 115, 129]
2D/3D Models	General Symbols	[106, 139, 144]
	Geometrical Models	[19, 124, 141]
	Stickers	[35, 72, 100]
	3D Scanned Model	[40, 90, 163]
	Networks/Links	[1, 88, 113]
Images	RGB Images	[40, 54, 113]
	Infrared Images	[40]
	Segmented Images	[88]
	Enhanced Images	[84, 103]
Audio	Spatialized Audio MSG	[66]
	Non-Spatialized Audio MSG	[46, 54, 84, 103]
Video	Imported Videos	[150]
	Captured Videos	[54, 84]
Text	Notes	[40, 54, 70, 84]
	Tags	[45, 106, 113]
	Subtitles/System Info	[34, 85, 114]
Hyperlinks/URLs		[144]
Highlights		[25, 52, 144]
Hand Gestures		[31, 84, 157]
Animation		[31, 90, 133]
Annotation Multi-Dimension		**Papers**
2D		[54, 70, 84, 107]
3D		[88, 124, 150]
2D Created - Rendered 3D (2D-3D)		[91, 96, 140]
Annotation Target Linkage		**Papers**
Screen-Fixed		[85, 106, 156]
World-Fixed		[42, 114, 144]
Annotation Creation/Selection		**Papers**
Pre-Defined		[73, 87, 98]
Free-Form		[25, 43, 139]
Annotation Generating By		**Papers**
User		[19, 40, 73]
System		[64, 90, 135]

To answer *RQ1* and have a better understanding of annotations used in current XR systems, we classified them based on different criteria. Table II shows this classification to describe annotations in a more detailed way based on previous research studies on annotations.

1) Annotation Form: From previous studies that used annotation in their systems, we classified annotation based on their form into ten main categories (see Table II): hand drawings (graphical drawing, textual drawing), graphical models (general, geometrical, and scanned models, stickers, and networks/links), images (RGB/normal, inferred, segmented, and enhanced images), audio (spatialized audio message, non-spatialized audio message), video (imported, captured), text, hand gestures, highlighting, animation, and hyperlinks. Fig. 4 shows some examples of different forms of annotation from previous works [40].

- *Hand drawing:* This form of annotation includes both textual and graphical drawings. Any non-text shape such as circle, line, arrows, free-form drawings, etc., that are usually created by the user in real-time can be considered a graphical drawing. The drawing properties such as position, scale, color, and opacity may be predefined in the application, or the application may allow the user to change the drawing properties themselves. Many previous studies that used drawing annotation allowed users to select their desired color [36], [42], [52], [98], [117], [139], [141], [155]. In [139], two types of drawing annotations (solid and outlined lines) were considered to separate annotations by the local and remote users.
- *2D/3D models:* These annotation forms involve any 2D or 3D graphical models used in previous studies to provide more information about a real or virtual object. The use of pre-existing models has been explored for collaboration and assisting another user [82]. We divided these models into four groups: (1) general graphical symbols (e.g., arrows, arcs, markers), (2) geometrical models/shapes (e.g., circles, cubes, sphere, cylindrical, etc.), (3) graphical stickers including predefined sets of 2D/3D stickers such as surgical instruments (scissors, forceps, etc.), emojis for conveying emotions, location signs, etc. (4) 3D scanned models of real objects (5) Annotation network/links are explicit links showing the connection between annotations and their respective target object.
- *Image:* In addition to sketches and graphical models, users can add an RGB image or an infrared image to the virtual/augmented environments to provide more information about a real or virtual object. Infrared photos are used to measure the temperature of an object using a sensor [40]. Users can take a screenshot of the 3D environment in real-time [102], or they can import them from a data source. Also, users may generate a new image by segmenting an image that is already provided [88] or combining them with other types of annotations. As an example, in [84], various annotation forms (e.g., drawings, notes, hand gestures, etc.) were used to create enhanced images.
- *Audio:* Users may convey information using spatial auditory cues or non-spatial auditory cues, especially in asynchronous communication. Yang et al. found that a spatialized remote expert's voice helps local workers find small occluded objects with significantly stronger spatial perception [157]. Langlotz et al. investigated spatially aligned audio annotations that were shown using visual hints are augmented in the user's mobile AR view [66]. In Fig. 4.1.j, colored dots indicate user-created audio annotations. Users can hear the audio annotation by pointing to the center of their phone's screen towards the colored dots.
- *Video:* There were several studies in AR or VR [120] that used video annotations to enrich a single user application (e.g., assessment tools for learning) [120] or a remote collaborative application. Video annotations are divided into imported videos or the recorded videos of the environment to be captured in real-time.
- *Text:* We classified the text annotations displayed in previous studies into three groups: textual notes, tags or labels, and subtitles and system info. Tags were the most common textual annotation forms used in our corpus. Wither et al.

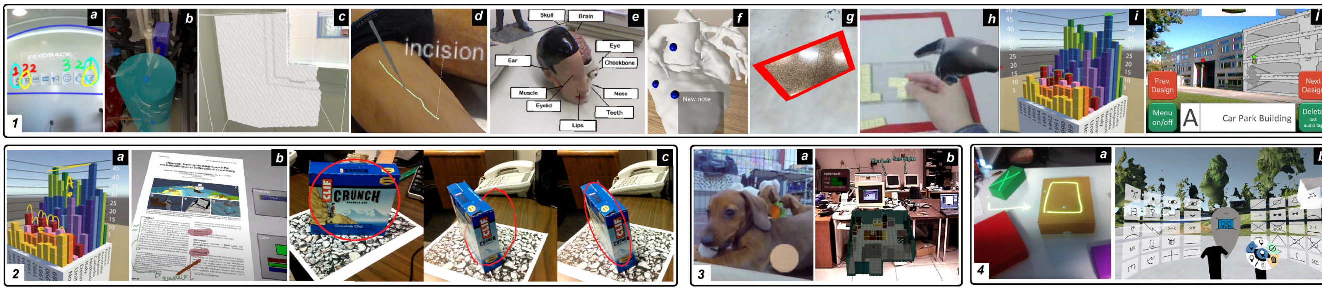

Fig. 4. Examples of different annotation types. 1.(a) textual and graphical hand drawing © 2018 ACM [129] 1.(b) graphical geometrical model © 2018 ACM [141] 1.(c) Graphical scanned model © 2019 Springer [40] 1.(d) stickers, hand drawing, and text [72] 1.(e) textual tags © 2016 IEEE [78] 1.(f) textual notes [70] 1.(g) images © 2019 Springer [40] 1.(h) hand gestures © 2018 IEEE [156] 1.(i) highlighting © 2022 IEEE [25] 1.(j) audio © 2013 ACM [66] 2.(a) 3D © 2022 IEEE [25] 2.(b) 2D © 2021 IEEE [115] 2.(c) 2D-3D © 2016 IEEE [97] 3.(a) screen-fixed © 2021 IEEE [155] 3.(b) world-fixed © 2002 ACM [11] 4.(a) free-form © 2013 ACM [2] 4.(b) pre-defined © 2019 IEEE [87].

showed textual label annotations in two ways based on the annotation viewing orientation: Edge and Perpendicular Label. In the Edge Label method, the label is oriented to match the orientation of the 3D AR object that is annotating, while in the Perpendicular Label method, the label is directly facing the user, which might look incorrect from any other location [150].

- *Link:* Link annotations can lead users to images, videos, etc. Further, adding links from social media encourages exploration of the knowledge in the social network community [160].
- *Highlighting:* This form of annotation can be authord by modifying the visual properties of a target object for annotation. This modification includes changing the size, color, and stroke to emphasize or diminish the importance of a target. In [25], VR-HMD users can highlight bars (showing by a bright green outline) of a virtual 3D bar chart one at a time by VR controllers. Another study used images to highlight buildings in an AR-HMD environment [135].
- *Hand gestures:* Yamada et al. proposed a novel system that uses hand gestures as a new annotation for their system [156]. In another study by Marques et al. the annotation system allows a remote user to point to the interest target using 3D virtual gestures in order to assist an on-site technician [84].
- *Animation:* Some previous studies augmented animated visual instructions in the 3D environment to guide the user [90], [133]. It has been shown that using navigation animation helps users to understand the details of manual assembly operations [147].
- *A combination of two or more:* Annotators may need to combine annotation forms, such as combining hand drawing and image annotations to enhance an image annotation by drawing sketches on a photograph [103].

To better understand the requirements of XR annotation systems, we extracted the common free-form hand drawing outlines and 2D/3D models used commonly in previous research studies on annotation. The free-form sketches and models are extracted from the papers' text, pictures, or the associated supplementary video in the published venue or YouTube. This provided a vocabulary that can be used as a reference for providing pre-defined sets of outlines and models, which may lead to decreasing annotation creation time and increasing user experience in 3D user interfaces. Since most previous studies did not provide any direct explanation of the generated drawings or 2D/3D models that were used, our knowledge may not be complete, and the provided dictionary may not cover all the generated drawings or used 2D/3D models.

Free-form hand drawings: Users in previous studies have generated various shapes of graphical drawings for different purposes, such as providing visual instructions in a collaborative setup or for self-review later. Some examples of these free-form drawings include: drawing a circle or ellipses outline (i.e., any kind of closed loop that resembles a circle around the object of interest [97]), square or rectangle to specify the location for placing an object [2], lines (including direct or curved lines) for immersive analysing 3D data [25] or showing the place of an action [72], or path [116], a cross-hair at objects center for the purpose of identification of an object, location [139], or as a disapproved sign [129], an arrow (including direct or curved arrows) for showing the movement directions [42] or specifying the object of interest [95], outlining the object at the target place for the placement of them, check mark signs to verify the collaborator's action [129] and handwriting the alphabet for virtual papers' review [116] or 3D data analysis [25], or numbers to determine the orders [129]. Additionally, users may use a combination of two or more graphical drawings, for example, an arrow and circle for a movement and placement task [91], or a cross-hair, arrow, and square for identifying the object of interest, movement and placement location [139]. In addition to standard drawing shapes, some drawings were allocated to a specific application such as drawing cloud-shaped outlines [15] or painting eyes and hands [107]. The right plot in Fig. 5 shows the most common graphical drawing shapes and the associated number of papers used these graphical drawings.

Graphical models: Regarding the graphical 2D/3D models that were used in previous studies, the following models were used in multiple studies: direct lines, arrows, circles/ ovals, circle dot, cubes/transparent cubes, sphere, cylindrical, a surrounding enclosed shape, numbers, alphabet, explicit links, points, and a cross-hair. In addition, some 2D/3D models were used for a specific application, such as surgical instruments (e.g., scissors,

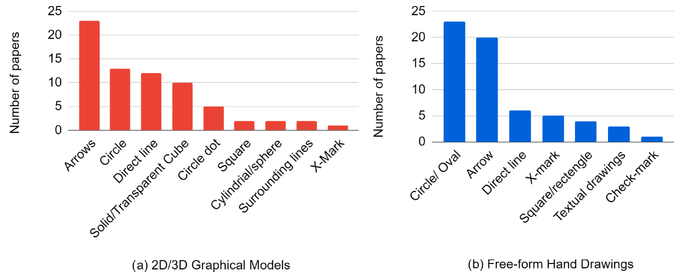

Fig. 5. Distribution of papers in our corpus for common a) 2D/3D models and b) free-form hand drawings.

TABLE III
THE VARIATION OF TEXTUAL LABELS IN OUR CORPUS | DAO AND GABBARD PAPER [26]

Billboard / Text	White	Black	Green	Red	Blue	Yellow
None	5 | 0	-	3 | 1	2 | 1	2 | 0	1 | 1
Black	3 | 1	-	-	-	1 | 0	-
Blue	3 | 1	-	-	-	-	-
Gray	1 | 1	-	2 | 0	-	1 | 0	-
Red	-	1 | 0	-	-	-	-
White	-	4 | 1	-	-	-	0 | 1
Yellow	-	3 | 0	-	-	-	-
Green	-	-	-	1 | 0	-	-
Purple	-	0 | 1	-	-	-	-

Forceps, etc.), location signs, physics signs (physics forces), and danger signs [100], [111]. The left plot in Fig. 5 shows the most common graphical models that were used in our corpus.

Textual tags: One of the methods for displaying texts in 3D AR environments is displaying the annotation as a billboard on physical 3D objects (e.g., buildings) [150], [151]. Dao and Gabbard surveyed text display variation in AR applications [26]. They found several common variations of text with billboard backgrounds, such as black, gray, or blue billboards with white text [49], white billboards with black or yellow text [77], and purple billboards with black text [99]. We surveyed text display variations for the systems that used a textual tag/label in our survey papers. Table III shows the variation of textual labels displayed in previous surveyed studies.

2) Annotation Dimension: The generated annotations are divided into three groups based on their dimension:

- *3D annotations:* 3D annotations including 3D geometric models [141], 3D audio, 3D hand drawings, etc. Langlotz et al. created the 3D audio/spatial audio annotation by adjusting the volume setting and the sound panning between the left and the right channels. The results from the user study confirmed that audio annotations are perceived as a useful source of information [66].
- *2D annotations:* 2D annotations used in our surveyed papers included 2D texts, images, hand drawings, etc. In [150], [151], 2D billboard-style text annotations (i.e., the text label is placed on the edge surface of the building) were used for annotating 3D physical buildings in an AR environment.
- *2D-3D annotations:* 2D-3D annotations refer to where the user generates annotations in 2D (e.g., drawing 2D sketches), then the system detects the 2D annotations and renders them in the XR environment with an appropriate depth and orientation as 3D annotations. This helps to minimize the perspective effect when the camera moves and shows the annotation from a novel viewpoint [96].
  One of the common cases to use these annotation types is in remote asymmetrical collaborative systems. When the remote user generates 2D annotation via a 2D screen display (e.g., tablets, smartphones, or desktop display), the local user views the generated annotations in a 3D XR environment through an AR-HMD or VR-HMD [140] or handheld AR [42]. Previous studies used different transformation methods such as "dominant plane", "median depth", "minimum depth", "Spray paint" [42], and "ellipse fitting the foreground point clouds" [71] methods. Thoravi et al. estimated scene depth using binocular disparity to render annotations at the correct depth in an asymmetrical collaboration (VR-HMD and Non-XR display) [140]. In another study that uses two symmetrical AR-HHDs for collaboration, the remote user draws 2D strokes on a smartphone's touchscreen. Then, the system maps arbitrary 2D drawings to registered 3D AR annotations, and the drawings become overlaid as 3D drawings in the AR view of the local user (smartphone) [91].

3) Annotation Target Linkage: Polys et al. has divided annotations into two major types based on the way annotations are overlaid on the real or virtual world:

- *Screen-fixed annotations:* Screen-fixed annotations are shown and anchored on display rather than on the object the information is about. These annotations provide information about the world or an object in the world and remain fixed in the user's field of view even if the user changes their point of view [28]. In [155], users provided their emotional feedback about several different 360° videos watched via a VR-HMD by two different visual annotations (dotSize and ArcShape) shown in a fixed position of the screen.
- *World-fixed annotation:* World-fixed annotations are linked to specific locations in the world or linked to an object in the world [28]. Even when the local user's view is changed, these annotations remain at a fixed spot [136]. One of the advantages of world-fixed annotations is addressing the problem of overcrowding data by not showing the information all the time [28]. We categorized world-fixed annotations into two sub-groups:
  - *Spatially-linked:* New advances in technology (e.g., wearable AR devices) provide the opportunity to overlay annotations on the real world based on the user's current position and orientation. Spatially-Linked annotations are linked to specific locations in the world. An example of such an annotation system is the system designed by Garcia et al.. All annotations are generated (e.g., the picture is taken or the scan is made) and spatially linked to the real world coordinates automatically [40].

- *Target object-linked:* These annotations are linked to specific objects in the world. The links between annotations and scene objects improve users' understanding, especially when the scene is complicated or many annotations are overlaid simultaneously. An example of using object target-links was used in [88] which displayed the linkage of different annotations to their corresponding target object by drawing lines between them.

4) Free-Form/Pre-Defined: In another classification, Fonnet et al. divided annotations into two groups of free-form and predefined annotations [38]:

- *Free-form:* Users can generate these annotations in real-time by drawing, taking pictures, scanning 3D models, and segmenting a text or picture. The most common real-time generated annotation is drawing. The properties of drawing (e.g., size, color, and opacity) can be predefined by the system or will be provided by the user while performing the task [50]. VANOTATOR is a framework designed by Mehler et al. for generating multimodal hyper-texts that allows users to segment images and texts [88].
- *Pre-defined:* For these annotations, the system allows users to choose a 3D/2D model from a predefined set of symbols/models that can be placed on the desired location or object in an XR environment. For example, in [124], the user can choose a symbol from a predefined palette of symbols (e.g., safety or maintenance area) and then put it in the geospatial model. A previous study has found using a palette of predefined symbols is very useful because it simplifies the annotation task, and the participants only need to place a specific symbol to a location instead of writing a description every time [124].

5) Generating by: Annotations in our literature review were generated or displayed during the applications by systems or users.

- *User:* users may create annotations directly in XR environments or author annotations in a non-XR environment (e.g., tablets, displays, etc.), and then the system renders the generated annotations and displays them on the XR environment. In both cases, various input modalities can be used for authoring annotations. We discussed different input devices for creating annotation in detail in Section IV-C.
- *System:* annotations that are not created directly by the user during the application are provided by the system. In such cases, the XR users are mostly only allowed to view the annotations and are not able to edit or create new annotations. However, users may be allowed to manipulate or filter the annotations. Previous studies that used system annotations mostly focused on view management, filtering, and clustering for algorithmic placement of the annotations in XR environments. We discussed these techniques in detail in Section VII.

6) Content Complexity: Content complexity which is the complexity of the information describing the object of interest, can vary greatly from low to high based on two dimensions [39], [152]:

- *Data complexity:* the amount of information that annotation provides for the user can vary greatly from one annotation to another. For example, the complexity of the amount of data that a simple textual tag [86] provides is less than an audio [66].
- *Visual Complexity:* Visual complexity of annotations can vary greatly from a simple marking of the object of interest to the ones whose content is an animated 3D model with sound [39]. Note that high data complexity doesn't necessarily imply high visual complexity. An example of high data complexity but low visual complexity is audio annotations shown with a simple dot in [66].

7) Location Complexity: Wither et al. mentioned in their AR annotation taxonomy that all the annotations must have some spatially dependent component [152]. So, a location should always be associated with the annotation in the real or virtual world. The complexity of location varies as follows:

- *A single 2D/3D point:* The simplest location complexity that an annotation can have is a single 2D/3D point that provides only the position information. Orientation information is defined arbitrarily by the application [152]. In [66], each audio annotation is spatially linked to a single 3D point in the physical world.
- *A 3D object:* If the target is a 3D object, the location complexity includes all the object's points. For example, Mosser et al. introduced an algorithm that takes a set of pixels representing an object as the input and then renders an annotation above it using a graph cut segmentation method [92].
- *A 2D/3D bounding region:* If the target object for annotation is a group of objects or a region, the location complexity includes the information of the boundary region around the target objects [39].

8) Interactivity: Annotations can be classified based on their interactivity [150]:

- *Static:* This is the case where users can only view the annotation. These annotations are static in the XR scene. They might be either created by the system/user during the offline process [78], [86], [96], [127], [159] or generated by the system/remote user and be rendered in the XR scene in real-time [73], [156], [162].
- *Interactive:* Users can only interact with annotations through manipulations, filtering, etc. Users are not allowed to edit or add any new annotations.
- *Editable:* Users are allowed to modify the existing annotations, i.e., they cannot add annotations for a new target.
- *Creatable:* This is the case where users can create new annotations for a new target location.

The distribution of the introduced annotation forms among our surveyed papers is shown in Fig. 6.

C. Annotation Creation Purpose

In each system, annotations were generated to satisfy a specific need. Matos et al. categorized annotations that were used in their system based on the annotation purpose for creation into four groups: informational (subtitles and markers), directional

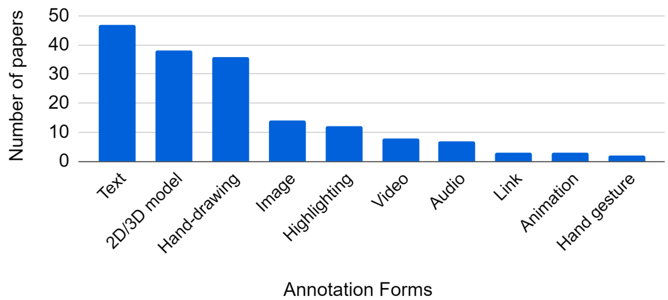

Fig. 6. The distribution of papers based on the annotation forms were reported in our literature review.

(arrow and miniature), narrative (Vignette and Lateral lights), and contextual (mini-map) [85]. Other purposes for creating annotations are as follows:

- Explaining how the annotated object works. For example, in Radu et al. authors used annotations to explain how an oscilloscope works [107].
- To guide the other user on how to act by providing information and directing the user's focus and attention to the target object [111], [124]. Most previous papers that used a worker-expert setup fell into this group [107], [140], [149]. In [98], the remote expert used annotations to guide the local worker by pointing to relevant places or specifying the contact points of two objects.
- To indicate modifications to physical objects [107]. In [107], an on-site user modifies a physical lamp into a cartoonized monster by adding virtual annotations.
- To revisit and review later [70], [116]. For example, in [70], surgeons annotated different components of a heart model using textual notes in real-time and then revisited them at a later time.
- To provide emotional feedback, Xue et al. used two annotation methods that allow users to provide emotional feedback while watching 360VR Videos [155].
- To enhance the teaching and learning experience [52]. Hoang et al. enables students of a class to author annotations on a projected skeleton in SAR to improve the learning process.
- To get the learner familiar with the complex elements and to make the reference more distinct in the collaboration [130].
- To guide to find the viewpoint [43], [85], [106]. In the AR-HHD system used by [106], When an annotation was off-screen, a virtual annotation pointed to the direction of the target annotation to guide the user where to look at.
- Labeling an object to provide more information about it [7], [135].

When designing an annotation system, it is crucial to consider the intended purposes of the annotations. However, understanding users' preferences and needs comprehensively requires further investigation. For instance, it is important to determine which annotation types are better suited for indicating users' intentions for each specific annotation. Some previous studies have demonstrated that certain annotations are particularly effective for specific purposes. For instance, it has been shown that an arrow is the most effective means of indicating off-screen content visualization [14]. Additionally, employing a direct line to guide users' attention to the next target has been identified as one of the most efficient methods [143]. Moreover, it is worth exploring whether there are any differences in the types of annotations that should be employed in AR compared to VR settings.

D. *Target Object for Annotation*

Our annotation definition covers a broad range of target objects to encompass all XR use cases. The target object for annotation might vary in size, dimension, and type. In addition, users may annotate the target object directly in the XR environment or through another device indirectly. In the indirect annotating methods, the system allows the user to annotate via another output device and then displays the generated annotations on the actual target object in the XR environment. One of the most common use cases of the indirect annotation method is in asymmetrical collaborative environments where on-site and remote users can collaborate on a task while the remote users see the local's user environment via a different device. For example, in [98], the remote user drew 2D annotations on the virtual replica of real objects through a multi-touch tablet display or VR-HMD, and the generated annotations appeared on the physical objects in the local user's environment via an AR-HMD in real-time.

The target object can range in size and number of items. It can be a single object, a group of objects, a semantic region of a scene [29], a room full of objects, or the entire AR/VR scene [154]. The target objects can be 2D, 3D, or a combination of 2D and 3D. We classified the target objects were annotated based on previous annotation studies that we found as follows:

- *Physical objects in AR:* Physical objects were the most common targets annotated in AR environments [7], [69], [90], [130], [139], [143]. Some examples of physical objects that were annotated in AR environments include physical buildings [125], industrial systems for water treatment [141], physical puzzles [156], and physical books [160]. Physical objects in AR might be annotated directly in the AR environment by the AR user or be created by the remote user via an external display and reflected in real-time on a real object by superimposing them on the live streaming video [156].
- *Virtual objects in AR:* Other target objects for annotating in AR environments are virtual objects. These virtual objects may be 3D objects such as virtual human body parts [70], previously generated annotations, etc.
- *Virtual objects in VR:* In several previous studies, users annotated the virtual objects of a VR environment to provide more information or analysis data. Danyluk et al. allows users to annotate 3D virtual bar charts via sketches and highlighting annotations [25]. Two other VR studies enabled their users to annotate 3D virtual environments while navigating simulated indoor or outdoor environments. For example, users could annotate the virtual objects of the scene, virtual roads, etc., by adding location mark models, inserting text notes, drawing paths, etc. Although virtual

3D model objects were the most common objects for annotation, some studies have investigated annotation for 2D objects in a 3D environment. In [116], users annotated 2D articles in VR.

- *The projected AR on the real object:* In [36], [52], users were asked to annotate a virtual 3D model projected on an actual human body in Spatial Augmented Reality (SAR) environments. The system enabled users to draw directly on the virtual 3D model projected on a volunteer's body [36] or draw annotations on a 2D display touch that renders the generated annotations on the augmented skeleton projected indirectly [36], [52].
- *360-degree VR video:* Several previous studies overlaid visual annotation cues in a live 360° VR video through VR-HMDs [85], [120], [136], [155]. In this case, the entire VR scene is considered the target object for annotating.
- *360-degree VR image:* Several previous studies allowed the VR user to annotate 360° panoramic images through a VR-HMD [15], [32]. The system introduced by Emerson et al. enables the user to explore a sequence of 360° images (e.g., street art, urban community gardens) through a VR-HMD [32] and annotate them.
- *360-degree non-immersive video:* The system introduced in [129], 360Anywhere, allowed remote users to draw sketches, place images, and write text into the 360° streaming display, which is displayed on a desktop display. Then, the generated annotations are augmented and projected into the physical environment to guide the local user in SAR.
- *2D video:* The virtual annotations may be overlaid onto the live 2D video [42]. Users need to freeze the video while annotating the 2D video [69].

TABLE IV
THE DISTRIBUTION OF ANNOTATION FORMS IN SURVEYED PAPERS - EACH COLOR SHOWS A DIFFERENT NUMBER OF ANNOTATION FORMS (1: PURPLE, 2: YELLOW, 3: RED, 4: GREEN, 6: GRAY)

Hand-drawing	Text	2D/3D Model	Image	Audio	Video	Hand gestures	Highlight	Link	animation	Num of papers
✓	-	-	-	-	-	-	-	-	-	24
-	✓	-	-	-	-	-	-	-	-	14
-	-	✓	-	-	-	-	-	-	-	15
-	-	-	-	-	-	✓	-	-	-	1
✓	✓	-	-	-	-	-	-	-	-	2
✓	-	✓	-	-	-	-	-	-	-	1
✓	-	-	-	-	-	-	✓	-	-	3
-	✓	✓	-	-	-	-	-	-	-	20
-	✓	-	✓	-	-	-	-	-	-	1
-	✓	-	-	✓	-	-	-	-	-	1
-	-	✓	-	✓	-	-	-	-	-	1
-	-	✓	-	-	-	-	✓	-	-	4
✓	✓	✓	-	-	-	-	-	-	-	2
✓	✓	-	✓	-	-	-	-	-	-	1
-	✓	✓	✓	-	-	-	-	-	-	3
-	✓	✓	-	-	-	-	-	-	✓	1
-	✓	-	✓	-	✓	-	-	-	-	1
-	✓	-	✓	-	-	-	✓	-	-	1
✓	✓	✓	✓	-	-	-	-	-	-	1
✓	✓	-	✓	✓	-	-	-	-	-	1
✓	-	✓	✓	-	-	-	-	✓	-	1
-	✓	✓	✓	-	✓	-	-	-	-	1
-	✓	-	✓	-	✓	-	-	✓	-	1
-	✓	-	✓	✓	✓	-	-	-	-	1
-	✓	✓	✓	✓	✓	-	✓	-	-	1

E. Discussion

Previous research studies have employed various annotation types in their systems based on factors such as input and output technologies and the type of target object being annotated. Therefore, researchers and developers can make informed decisions regarding the most suitable annotation types for their XR systems by considering these factors.

Our findings from previous XR annotation systems indicate that AR-HMD users commonly generate free-form hand drawings as annotations [19], [20], [107], [110], [139]. Conversely, text inputs are less frequently generated using these devices. On the other hand, AR-HHD users heavily rely on text annotations, benefiting from the ease of typing through multi-touch hand gestures [34], [45], [59], [59], [72], [84], [119], [160].

The characteristics of target objects for annotating, including their types and size, may play a role in determining the appropriate annotation type. For example, when the target object is the entire XR scene, screen-fixed annotations are more advantageous [154]. Furthermore, the use of some annotation forms was more popular for some targets. As an example, textual tags were commonly used for annotating real objects in AR [78], [86].

Finally, with the advancements in XR technologies, offering a more comprehensive range of customizable options and diverse annotation forms may positively affect embracing XR technologies by a broader range of users. However, it is noteworthy that many previous studies have focused on employing only one or two forms of annotations in their systems. Table IV shows the distribution of surveyed papers on the combination of different annotation forms. Our finding shows that most previous studies used single annotation forms (free-form drawn lines, 2D/3D modes, and textual notes) in their systems (shown in purple). However, the combination of textual notes and graphical models was also commonly used.

IV. TECHNOLOGY

To answer *RQ2*, we classified previous XR studies that used annotation in their system based on their output and input devices. Although this section focuses on the hardware, we first listed some of the software used in previous XR systems to provide a general idea of the required software in these systems. Microsoft Remote Assist,[4] Spatial,[5] Tilt Brush,[6] and Gravity

[4]https://docs.microsoft.com/en-us/dynamics365/mixed-reality
[5]https://www.spatial.io/
[6]https://store.steampowered.com/app/327140/Tilt_Brush/

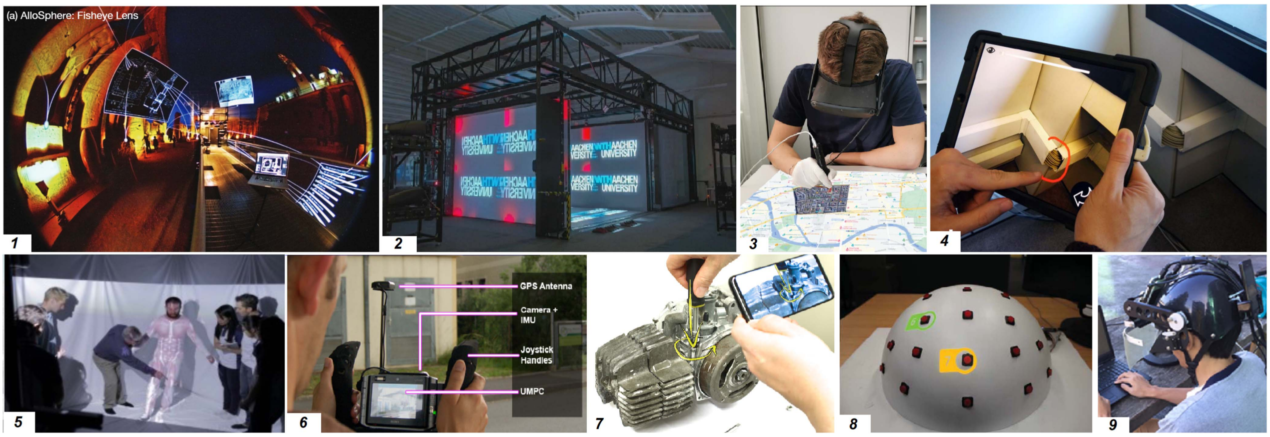

Fig. 7. Examples of different XR display technologies: (1) SAR (AlloSphere) © 2016 IEEE [113] (2) VR-SSD (CAVE) [103] (3) VR-HMD (Oculus Quest) © 2021 IEEE [116] (4) AR-HHD (tablet) [40] (5) SAR (3-sided projected- based) © 2017 ACM [52] (6) AR-HHD (custom-built) © 2008 IEEE © 2020 ACM [124] (7) AR-HHD (smartphone) [91] (8) SAR (Demo project-based) © 2018 IEEE [143] (9) AR-HMD (HHMPD) © 2014 EEE [63].

Sketch[7] are some examples of industry tools that support collaborative AR and drawing or 3D modeling in VR. However, since most industry tools are not free, open-source, or compatible with the device or software of the study, only some previous studies used these industry tools [41], and many previous XR studies developed their own annotation systems. Most previous XR systems utilized the Unity engine for developing their XR environment [28], [40], [84], [88], [119]. Additional software and libraries that were used for developing XR environments are: Vuforia [84], Microsoft's Mixed Reality ToolKit (MRTK) [25], [76], Virtual Reality Toolkit (VRTK) [28], Visualization Toolkit (VTK) [2], OpenGL plugin [54].

A. XR Output Devices

Output devices or display technologies refer to the type of visual output device that is used for viewing the environment [126]. Over previous years, virtual annotations have been used in various XR display technologies. Hoang et al. considered three types of display technologies in their AR technology classification: Head Mounted Display (HMD), screen-based display, and projection-based AR [52]. In a more comprehensive classification in a survey paper on 65 papers on XR technologies by Sereno et al. [126], output devices were classified into six groups: CAVEs, VR Head-Mounted Display (VR-HMD), AR Head-Mounted Display (AR-HMD), AR Hand-Held Displays (AR-HHD), Spatial Augmented Reality devices (SAR), and traditional screens.

From previous research studies on annotation in XR, all of the XR display technologies and their corresponding input devices were extracted. The classification that was used by [52], [126] was considered for the categorization of used output devices in our papers pool. In addition to existing display outputs in [52], [126], several other display technologies (e.g., workbench) were found in the surveyed papers and added to this classification. Table V shows the classification of surveyed papers based on their display technology into seven groups: (1) AR-HMDs including Microsoft HoloLens, MagicLeap, etc., (2) AR-HHDs, including smartphones, tablets (with/without pen), and custom-built AR handheld devices, (3) SAR or projected-based AR (3-sided, 2-sided, and 1-sided), (4) VR-HMDs including Oculus Rift, Oculus Quest, HTC Vive/Pro Eye, (5) VR Surrounded Screen Displays (VR-SSDs) including CAVE, and AlloSphere, (6) VR Table-like (VR-TBL) display including workbench-like 3-D display [37], and (7) external screens (non-XR displays) including desktop displays, multi-touch displays, and tablets. Fig. 7 shows some examples of different XR display technologies that were collected from surveyed papers.

It is important to note that our study only focuses on categorizing the XR output display technologies used in our collected papers. However, there exist other XR output devices that are not included in our classification, such as glasses-free 3D autostereoscopic displays that provide immersive visual experiences with quality depth of various image applications without the need for specialized eyewear like 3D glasses or goggles [56].

- *AR Head Mounted Display (AR-HMD):* Two main methods for displaying AR content on smart glasses are: Optical See-Through (e.g., HoloLens) and Video See Through (e.g., HTC Vive) AR-HMDs, which enable users to perform tasks freely and manipulate objects in a real environment by leaving their hands free. Many previous studies confirmed improvement of performance by using AR-HMDs, especially when the task requires manipulation of the physical environment [48], [55]. One of the most popular AR-HMDs is Microsoft HoloLens which has been used in several AR studies [107], [136], [149]; however, most AR-HMDs have a limited field of view. In [63], the authors used an adapted version of the wearable Hyperboloidal Head-Mounted Projective Display (HHMPD) which was developed by kishishita et al. to investigate the effects of a wide field of view AR-HMD on the perception of augmentations in terms of search performance and mental workload.

[7] https://www.gravitysketch.com/

TABLE V
OUTPUT DEVICES AND THEIR ASSOCIATED INPUT MODALITIES

	Output Devices	Input Modalities
AR-HMDs	1. Microsoft HoloLens [150] 2. Magic Leap One [121] 3. Canon HM-A1 [98] 4. Google's ARCore [88] 5. HHMPD [63] 6. SVGA Sony [152] 7. Pixel opaque [138] 8. DAQRI Glasses [163] 9. Cy-VisorDH-4400VP [58] 10. Not Specified (NS) [88]	1. Bare hand gestures [110] 2. Head gaze [135] 3. Eye gaze [121] 4. Voice input [136] 5. Joystick [156] 6. Handed tracked mouse [138] 7. Handed keyboard (twiddler2) [58] 8. Spatially tracked tablet [76]
AR-HHDs	1. Windows tablet [159] 2. Android tablet [45] 3. iPad tablet [40] 4. Smartphone [66] 5. Customized HHD [124]	1. Display touch gesture [141] 2. Digital Pen [69] 3. Joystick [124] 4. Voice input [119] 5. Tracking sensors [40]
SAR	1. AlloSphere (+AR-HHD) [113] 2. SAR (3-sided) [36] 3. SAR (1-sided) [144]	1.Smartphone [113, 114] 2. Maxell Pen [44] 3. Customized pen [98] 3. Customized projected palette [98] 4. Projected display touch [44]
VR-HMDs	1. HTC Vive [25] 2. HTC Vive Pro [156] 3. Oculus Quest [116] 4. Oculus Rift [140] 5. Sony HMZ-T3W [98] 6. Not specified [46]	1. VR Stylus [116] 2. VR Controllers [139] 3. Vive trackers [25] 4. Joystick [156] 5. Bare Hand gestures (LeapMotion) [137] 6. Eye Gaze [121] 7. Hi5 VR Gloves [65] 8. Voice input [32] 9.Tracked Mouse [98] 10. Lazy susan turntable [98] 11. Tracked Leonar3Do bird controller [98] 12. Physical keyboard [32, 60] 13. Tangible user drawing (TUD) [145]
VR-SSDs	1. CAVE (5-sided) [103]	1. ART Flystick 2 [103] 2. Smartphone [103] 3.Voice input [103]
VR-TBLs	1.Workbench [37]	1. Tracked Pen [37] 2. Tracked palette [37] 3. Tracked Navigator [37]
External	1. Desktop display [157] 2. Multi-touch display [42] 3. Tablet [139] 4. Large display Wall [32] 5. Table-like Multi-touch display [72]	1. Display touch gestures [42] 2. Mouse [32] 3. Keyboard [32] 4. Digital pen [36] 5. Gestural commands [130] 6. Vocal commands [4]

A referenced example of each output/input is provided in front of each of them.

- *AR Hand-held Display (AR-HHD):* While AR handheld devices might not be as immersive as AR-HMDs, they are more practical and accessible [83]. AR handhelds are primarily used for tasks that require high mobility or outdoor navigation [106]. Video-see-through mobile devices are common standard handheld displays that have been used in AR environments (smartphones and tablets). Smartphones can be used as handheld magic lenses or video streams in near-eye displays as a stereo binocular display through a mobile viewer (e.g., VR2GO mobile viewer) [114]. In addition, some previous studies developed a custom-built handheld AR system for their applications. An example of customized handheld AR is the system designed by Schall et al. for outdoor network planning and the inspection of underground infrastructure. The authors developed a two-handed shell around an Ultra Mobile PC (Sony Vaio UX) that holds a GPS receiver, an inertial orientation tracker, and a high-quality industrial camera connected to the UMP [124].
- *Spatial Augmented Reality (SAR):* Another form of AR system is called Spatial Augmented Reality (SAR), which uses projectors to display spatially aligned augmented information directly onto the surfaces of real objects [16], [109]. SAR, or project-based AR, does not have the limitations and technical issues caused by other AR display technologies (e. g., AR HMD and AR Handheld) such as a small field of view, tracking issues, or registration error [123]. Also, users don't need to wear or hold display equipment [108]. SAR technologies differ in the shape of the surface for projection, number of sides, number of projectors, resolution, and their immersive degree. Previous studies in our literature review used three types of surfaces for projection-based AR: spherical display [113], [114], wall-like display [129], and a demo display [10], [36], [52], [80], [143].
  Ren et al. used AlloSphere as the simulator of AR devices consisting of a three-story high full-surround display environment driven by 26 active stereo projectors. The authors used this technology to investigate wide-field-of-view annotations that link objects far apart in the visual field [113]. Fig. 7.1 shows this display technology. Another study developed a system called Augmented Studio consisting of a three-sided stage with two white walls and a floor for projection mapping. Augmented Studio projects anatomical structures and annotations over moving physical bodies for physiotherapy education [52] (Fig. 7.5).
- *VR Head Mounted Display (VR-HMD):* VR-HMDs are small displays mounted on a helmet and designed for a single user [23]. VR-HMDs provide a high degree of immersion and recreate a 360-degree Field OF View (FOV) in VR. VR-HMDs allow users to move their hands while exploring the immersive virtual environment freely. A virtual environment might be simulated using 3D virtual objects [25], a 360-degree image [15], or 360-degree video recording of a physical environment [120], [155], or a combination of these options. Oculus Quest/Rift and HTC

Vive/Pro Eye were common VR-HMDs used in previous VR-HMD annotation systems.

- *VR Surround-Screen Display (VR-SSD):* Two well-known high-end surround-view VR environments are CAVE and AlloSphere [12]. Pic et al. used a 5-sided CAVE (no ceiling) that consists of 24 active-stereo projectors (16 for four walls and 8 for the floor) and an optical infrared tracking system [103]. Fig. 7 shows a CAVE immersive environment. AlloShere, is a large-scale full-surround immersive VR system equipped with high-resolution active stereo projectors [53]. Unlike VR-HMDs, VR-SSDs are designed for multiple users. Previous studies have shown the positive effect of these displays on perception and navigation for visual tasks [9].
- *VR-Table like (VR-TBL)/WorkBench:* A workbench is a large table-like display monitor under which virtual objects appear three-dimensional, as floating in the air, projected stereoscopically. A workbench system similar to a CAVE allows multiple users to view the same range projected simultaneously. The workbench setup includes a VR table, tracking cameras, and a pair of tracked shutter glasses that allows users to experience a semi-immersive environment.. [37]. Fiorentino et al. used a Workbench for free-form drawing, surfacing, and engineering visualization purposes. Users were able to interact with the system (e.g., draw) through a 6D input device. [37]

Note that although almost all studies used only AR or VR environments in their system, some previous systems gave their user the ability to switch between AR-VR modes [139]. This allowed users to choose their preferred environment. For example, VR can be used when the user needs to focus on a task without distracting from the physical environment [139]. A more advanced example is ThirdEye's X1 Smart Glasses, which enable users to easily switch between AR and VR interfaces.

Another categorization that can be considered is based on the immersive degree of the display technology. Slater et al. considered the feeling of being present as a determiner of immersive technologies [128]. Fonnet et al. classified systems that offer 3D graphics, stereo vision, and head tracking as immersive technologies [38]. We considered both definitions and grouped the display technologies in this literature review into two groups: immersive and semi-immersive technologies. Both VR and AR HMDs, which offer 3D graphics, head tracking, and stereo vision, were placed into the immersive technologies group. Another example of immersive technologies is CAVE, a well-known projection-based display, which is a multi-sided immersive projection room [13]. The next group is semi-immersive technologies, including hand-held AR displays, SAR display technologies with less than three projection sides, and workbench-like 3D display.

B. Input Devices

Interacting with each XR output display requires its own input devices and modalities. Input devices refer to the type of input used to interact with a virtual or augmented environment [126]. These input devices might be used to interact with the system or with another user. Fig. 8 shows examples of different input devices used in previous studies. We collected the main input devices that were used for interacting with each type of display technology and summarized them in Table V. Although we collected all the input devices for the system, the primary focus was on the input devices used to author annotations.

TABLE VI
MODALITIES CLASSIFICATION - MODALITIES ARE SHOWN IN BOLD FONT WERE USED FOR ANNOTATIONS

Input Modalities			Paper Examples
NonVerbal	Gestures	**Hand gestures**	[19, 110, 157]
		Foot gestures	[6]
	Gaze	**Head gaze**	[135, 137]
		Eye gaze	[147]
Verbal	Audio MSG	**Non-spatialized audio**	[54]
		Spatialized audio	[32, 66, 81]
	Speech	Non-spatialized speech	[43, 107, 137]
		Spatialized speech	[110]
	Voice recognition		[32, 102]
Tangible	Controllers	**VR controllers**	[87, 139, 140]
		Joystick / flystick	[103, 124, 155]
	VR Stylus	**Standard VR pen**	[115]
		Custom-built VR pen	[37, 150]
	Traditional	**Keyboard**	[32, 58, 63]
		Handed Mouse	[98, 138, 152]
		Digital Pen	[69]
	Display touch gestures		[72, 127, 161]
	Custom- Built		[36]
Physical	Heart rate (HR)		[155, 156]
	Accelerometer (ACC)		[155, 156]
	Blood Volume Pulse (BVP)		[155, 156]
	SKin Temperature (SKT)		[155, 156]
Virtual	**Avatars**		[107, 139, 150]
	Virtual pointer		[98, 130, 136]
	Virtual menu		[87, 102, 103]
Haptic	**Tangible User Drawing (TUD)**		[145]
	Physical keyboard in VR		[60]

Hertel et al. classified the interaction techniques into five main groups based on a literature review of 44 papers that used immersive AR: brain-computer interfaces (BCI), gaze (head gaze, eye gaze), voice, gestures (hand, head, foot), tactile interactions including tangible (controller, custom-built, everyday objects), generic input devices (clicker, stylus/pen, mouse), and touch [50]. In another surveyed paper by Sereno et al. on 68 papers in collaborative AR, input modalities were categorized into hand tracking, tracked controllers, hand mid-air gestures (which involves hand tracking), touch, head gaze/orientation, eye gaze, tangible, non-tracked controller, speech, and regular keyboard and mouse input modalities.

We adopted the classification by [50], [126] and modified it based on our literature review on previous XR studies. In total, we identified the following input modalities: non-verbal, verbal, tangible, trackers, physical, virtual, and haptic inputs. Table VI shows the classification of modalities in our literature review.

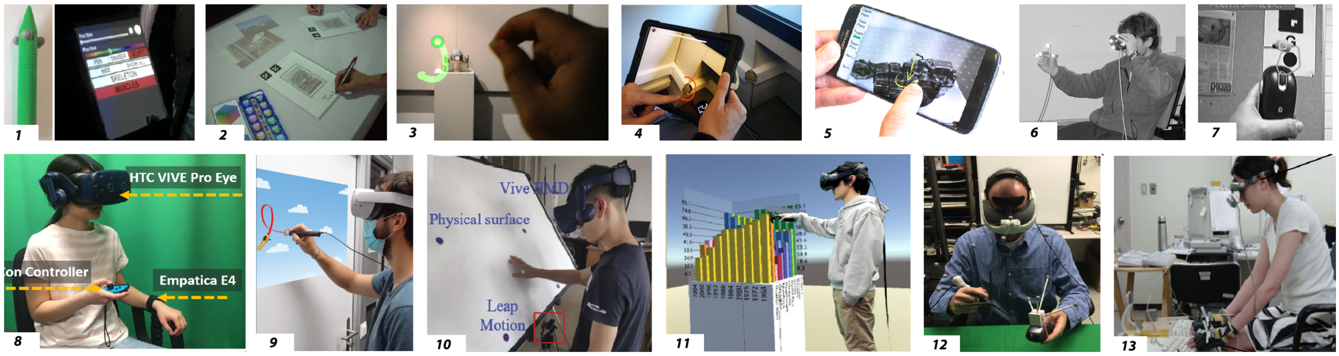

Fig. 8. Examples of different input devices for creating annotations and intercting in XR interfaces (the XR display technology that associated with each input device): (1) customized 3D printed pen and projected palette (SAR) © 2019 IEEE [36] (2) Maxcell pen (SAR) [44] (3) mid-air gestures (AR-HMD) [19] gestures (AR-HMD) (4) display-touch gestures (AR-HHD: tablet) © 2019 Springer [40] (5) display-touch gestures (AR-HHD: smartphone) WorkBench [91] (6) AR stylus (AR-HMD) [37] (7) handed mouse (AR-HMD) [151] (8) joy-controller and Ematica E4 [155], (9) Customized VR-stylus (VR-HMD) [116] (10) Leapmotion and TUD (VR-HMD) [144] (11) VR controllers (VR-HMD) [25] (12) Leonar3Do bird controller and mouse in pint3D (AR-HMD) [98] (13) data gloves and physical keyboard (VR-HMD) [60].

- *Non-verbal cues:* Hand gestures and eye gaze are two natural non-verbal cues that can be used to increase the degree of being co-present and to decrease user's workload [8]. The surveyed papers in our literature review used hand gestures to interact with the system or another user (e.g., by generating annotations). Different tracking devices can be used to track users' hand gestures, such as Leap Motion hand tracker [121], [136], [137], HoloLens sensors [20], [110], camera tracking hand gestures [107], and data gloves [65].
  Bare-hand or mid-air gestures have been used to draw annotation annotations [20], [110] or manipulate an object. Several previous studies used gestural commands to perform a simple task such as manipulating the perspective or an object [130]. Another study enabled users to rotate the perspective horizontally when their fist moves left/right or zoom in/out the perspective when the fist moves forward/back [130]. In another study, the authors used a set of 16 distinct gestures (rotate Right(R)/Left(L)/RL/LR, circle R/L, swipe R/L. wirl R/L, tap, double-tap, Eartouch R/L, chesttouch) for performing singular tasks [4].
- *Verbal cues:* Verbal cues are divided into speech and voice recognition. Where speech refers to the act of producing audible sounds and words using the human voice, while voice recognition refers to the technology that enables computers to recognize and interpret spoken language. Pick et al. used voice recognition via a microphone as the input method for typing a text [103]. Some previous studies used simple vocal commands as the main input for interacting with their system, such as for scrolling or confirmation of a task [4].
- *Tangible inputs:* VR Controller, Joystick, XR stylus, keyboard, mouse, data glove, and display touch are some tangible devices have been used in previous studies. Joystick has been used as the input device of a VR-HMD [154], [155] and AR-HHD devices [124] to interact with the XR system and create annotations. In [103], a wireless 6-DOF ART Flystick 2 was used as the input device of a CAVE. This input device enabled users to interact with CAVE through point-and-click interactions.
  Mice, physical keyboards, and digital pens were mainly used as input devices in the non-XR (e.g. desktop displays) side of a collaborative XR task. For example, in [43], a remote user can pan, zoom in/out, and change the view by moving the mouse while pressing the right button, scroll wheel, and right-click. Also, users can save the current viewpoint or revisit a saved view by pressing Alt plus any number key or pressing the respective number key alone from the keyboard. Display touch gestures have been mainly used for interacting with AR handheld devices that act as magic lenses. Furthermore, display touch gestures that are mentioned in this literature review may be used in a collaborative XR setup, but on the non-immersive side, such as a multi-touch desktop [42]. Various gestures have been developed and used for multi-touch displays such as one-finger (tapping) and two-finger (panning, pinch, and swiveling) [141].
- *Physical cues:* The users' physiological signals can be captured using different input devices to better understand participants' emotional states. The system used in [155] measures the Blood Volume Pulse (BVP) and EDA of the user using an Empatica E4 wristband.
- *Haptic inputs:* Some previous studies provided haptic feedback by using Tangible User Interfaces (TUI) to improve user performance [5], [89] and user preference [144]. TUI uses suitable physical objects to provide haptic interaction in human-computer interfaces. For example, in [144], the authors used a Tangible Physical Drawing (TPD) that enables VR-HMD users to sketch on a real physical surface. In another study by Kim et al. haptic feedback is provided using a physical keyboard for VR-HMD users to generate a text input [60].
- *Virtual cues:* Virtual avatars, pointers, keyboards, and menus can be used in an XR environment to interact with the system and create annotations. In [136], [137], users can make a virtual ray pointer using an index finger hand

gesture pointing at a target or draw annotations at the tip of the virtual ray pointer and erase all generated annotations by making an 'OK' hand sign. Pic et al. allows users to interact with the system via a virtual pie menu. In addition, users can create text inputs using a virtual keyboard [103].

C. Input Devices for Authoring Annotation

As we discussed in the previous section, different input modalities may be used for generating annotations in an XR system, such as fingertip and hand direction, stylus [149], gestural commands [4], [130], bare hand gestures using Leap Motion or HoloLens trackers [130], [136], [144].

Here we listed some of the input devices and modalities used to author two primary annotation forms (hand-drawn lines and text). Furthermore, we briefly mentioned some modalities for manipulating existing annotations.

Users drew in a 3D XR space using various inputs such as tracked bare hands [107], VR stylus [149], VR controllers [139], mouse [138], [151] and TUD [144]. Radu et al. allows users to draw in a 3D space through their fingertips or hand directions, which are tracked by the XR headsets [107]. Users can draw strokes in mid-air using VR controllers [25] or bare hand gestures [19], [20], [95] through VR-HMDs or AR-HMDs.

Weibel et al. equipped the VR user with a VR stylus that enables the user to interact with the VR interface and to draw annotations [149]. In another study, the authors used a 3D printed physical pen to draw directly on a virtual model of a skeleton projected on a physical human body. They also designed a projected palate that allows users to select the brush size, color, eraser, etc [36].

In [42], multi-touch gestures were used for navigation and drawing annotations: one-finger tap to go to respective viewpoint, one-figure free-form to draw annotations, two-finger pinch gesture for zooming, two-finger tap gesture to freeze the view or go to the point of interest, two-finger swipe gesture for rotating the view, two-finger panning (i.e., moving two fingers in parallel) to rotate the virtual camera around its optical center, two-finger snap gesture (i.e., keeping one finger static at a specific point and move the second finger in an arc around it) for orbiting around a point.

Another common form of annotation is textual annotation. In our surveyed papers, we found different methods to type a note, such as voice recognition [32], [103], physical keyboards [138], VR controllers through a virtual keyboard [32], display touch [103], [113], and air-writing [4]. A system might offer a combination of two or more text input methods. For example, the system introduced by Pick et al. enables users to enter a text using speech recognition via a microphone and Microsoft's Speech API (SAPI) or through a display touch of a smartphone [103]. Another study allowed users to type via voice recognition and VR controllers accompanying a virtual keyboard [32].

Although the physical keyboard was primarily used as an input device for entering text inputs in non-immersive technologies, Kim and Kim [60] used a physical keyboard for generating alphanumeric input for HMD-based VR systems. Since the VR user wears the VR-HMD, they can not see the physical keyboard, and the physical keyboard is used as a tangible interface to the virtual keyboard in the virtual space.

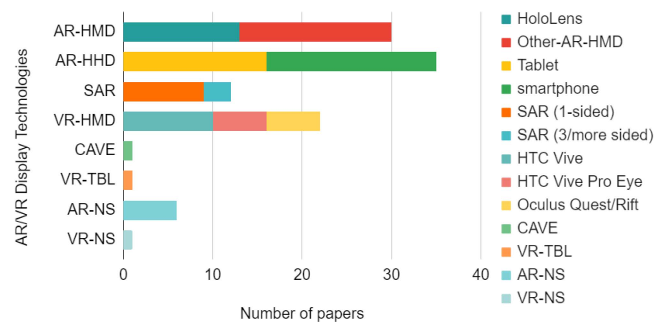

Fig. 9. Distribution of papers based on their XR output technology (AR/VR-NS means AR/VR technology was used, but the authors did not specify the device).

In addition to input techniques for generating free-form annotations, users may manipulate or generate pre-defined annotations using different modality inputs such as display touch gestures and gestural commands. Tomlein et al. used two different display touch gestures (single-finger and two-finger) for moving graphical annotations within a 3D AR environment: single-finger for automatically placing annotations on the nearest detected surface, and two-finger gestures for freely moving annotations in case surface couldn't detect [141]. Sun et al. used gesture commands to bring out or remove the annotations. They used "Tap" to add tags on the corresponding optical elements or removed all of the augmented annotations by performing the gesture of swipe [130].

D. Discussion

Depending on the target application, different types of XR displays are more appropriate to use. For example, if the application is 3D data visualization, using HHDs might not be as valuable as other display technologies due to multiple reasons [126]. First, the non-stereoscopic nature of HHDs have limited its value compared to HMDs. On the other hand, HHDs are usually less powerful for rendering and have smaller screens compared to traditional screens [126]. We found three papers in our literature survey review that focused on data analysis applications, and all three used VR-HMD [25], [103] and AR-HMD [76] in their immersive system.

Fig. 9 shows the distribution of surveyed papers based on their output display. As can be seen from the plots, AR-HHDs (34%), AR-HMDs (29%), and VR-HMDs (22%) were the most commonly used XR devices that used virtual annotations in their systems.

V. Annotation in Collaborative XR

In this section, we focused on collaboration types, technologies, and the annotator in a collaborative setting of annotation systems to answer *RQ3*. Previous studies have shown that virtual annotations can be used to improve collaboration and remote assistance applications in XR environments [33], [63], [156].

TABLE VII
Paper Distribution Along Their Collaboration Type in Terms of Time, Space, and Output Technology Dimensions

Location of Use	Time	Display	num of papers
Co-located	synchronous	symmetrical	2
Co-located	synchronous	asymmetrical	6
Co-located	asynchronous	symmetrical	2
Co-located	asynchronous	asymmetrical	2
Remote	synchronous	symmetrical	3
Remote	synchronous	asymmetrical	21
Remote	asynchronous	symmetrical	3
Remote	asynchronous	asymmetrical	3

The use of annotation in XR environments has many benefits, such as helping users to perform collaborative tasks faster, easier, and with fewer errors [137], enhancing alertness, awareness, and a better understanding of the situation [81], and helping to convey spatial information [27], [61].

The growth of highly connected businesses has made the urge for remote collaborations critical in their need to overcome barriers such as geographic restrictions and different time zones [101]. One of the most common applications for annotation in collaborative mixed reality (XR) environments is local-expert guidance or mentor-mentee use cases. When the local user lacks specific expertise, they rely on the help of the remote expert through tele-mentoring systems while being geographically separated. In a tele-monitoring system, the expert user guides the local user through various modalities such as audio, gestures, gaze, annotation, etc. The use of annotations allows remote collaborators to annotate the user's view by overlaying information, which improves communication between collaborators [27]. The use-case of annotation in various previous tele-mentoring systems has been explored in various contexts such as repair [110], maintenance [84], inspection [40], surgery training [73], [149], and physics [91]. Our survey data shows that 33 out of 103 papers that used annotation in their systems were collaborative.

A. Collaborative Systems Classification

Computer-supported cooperative work (CSCW) has classified any form of collaborative computer-based tasks, including XR tasks in a 2×2 setup for spatial and temporal dimensions [123]. Based on their classification, the spatial dimension of the users (the location of use) can either be co-located or remote, and the temporal dimension (time aspect) can be synchronous or asynchronous. We also considered the display technology that collaborators use for connecting to the environment (symmetrical and asymmetrical) as the third dimension.

Table VII shows the distribution of papers based on their collaboration type. Our finding shows the most common collaboration types used in previous annotation studies were remote, synchronous, and asymmetrical collaboration, and only a few exceptions supported asynchronous annotation. These results were not surprising for us since the results from a previous study on collaborative MR by Ens et al. reported that the vast majority of papers in their corpus focused primarily on synchronous collaboration, and only three studies used asynchronous annotations, usually for AR systems in a co-located and symmetrical collaborative setup [33].

B. Annotator in Collaborative XR

Our focus on collaborative papers is only on papers that used annotation in their collaborative setup. The initial criterion for a paper to be included in our collaborative setup was that at least one of the display technologies used in the collaborative system must be an XR output device. Collaborative setups between two users based on their output device can be classified in a {7 x 7} - 1 setup including a combination of AR-HMD, AR-HHD, SAR, VR-HMD, VR-SSD, VR-TBL, or external display technologies except if all the output technologies used in the system were non-XR display.

We extracted all the collaborative papers from our corpus, which resulted in eleven collaborative setups, including two user collaborations (16 articles), asynchronous collaboration among multiple AR-HHDs users (5 articles), two co-located synchronous collaborations among SAR users and a hybrid synchronous asymmetrical cooperation between three users. Table VIII shows the XR collaboration setup, display technologies, and the annotator in our corpus. Our finding shows most previous collaborative studies were designed for only two collaborators. However, most asynchronous collaborative studies did not mention any limitation on the number of collaborators (NL) that can benefit from the collaborative setup. We reported the annotator side for each collaborative configuration. Our finding shows the annotator in more than half of the collaborative papers was an external user (gray cells in the table), not an XR user. This highlights the need for more investigation of the challenges that XR users face while creating annotations for AR/VR environments.

In addition to annotation, XR technologies can support various interaction methods to enrich collaboration by sharing other non-verbal cues such as virtual pointer [104], View Frustums or head gaze [142], eye gaze [24], [79], [104]. These non-verbal cues have an important role in improving awareness of remote collaborator's status [145]. Furthermore, video streaming [118], speech [118], [142], [157], avatars [47], [67], [94], [105] or even face-to-face interactions in co-located collaboration setup are other interaction methods that have been used in a collaborative configuration.

To better understand the use of modalities alongside annotation in a collaboration setup, we summarized the combination of collaboration modalities in our literature review paper in Table IX. Our findings showed video streaming and speech, in addition to annotations, were commonly used in remote synchronous collaboration systems. However, due to the nature of asynchronous collaboration, annotation was used as the only collaboration modality in asynchronous studies.

VI. Annotation Tasks

In this section, we investigated *RQ4*, which were common tasks designed to evaluate the effect of annotation in XR studies. In our literature review, users used various actions to interact

TABLE VIII
THE COLLABORATIVE SETUPS OF SURVEYED PAPERS IN OUR CORPUS - COLORED CELLS HIGHLIGHTS THE WORKS THAT, ALTHOUGH ARE CONSIDERED AS ANNOTATION IN COLLABORATIVE XR, THE ANNOTATOR CREATED THE ANNOTATION IN A NON-XR DISPLAY

Collaborative Setup	**Display 1 (D1)**	**Display 2 (D2)**	**D1- Annotator**	**D2- Annotator**	**Both**
AR-HMD & AR-HMD	Microsoft HoloLens	Microsoft HoloLens	-	-	[110]
AR-HMD & VR-HMD	Magic Leap one Microsoft HoloLens Canon HM.A1 VST	HTC Vive Pro Eye HTC Vive Sony HMZ-T3W	- - -	[121] [150], [136, 137] [98]	- - -
XR-HMD & XR-HMD	HTC Vive	HTC Vive	-	-	[139]
AR-HMD & VR-Table	Sony glasses	Workbench	[37]	-	-
AR-HMD & External	Microsoft HoloLens Microsoft HoloLens DAQRI Smart Glasses Canon HM,A1 vST OSTHMD	Multi-touch table display Desktop display Desktop display Multi-touch tablet Desktop display	- - - - -	[72, 73] [157] [163] [98] [130]	- - - - -
AR-HHD & AR-HHD	Smartphone Tablet	Smartphone (NL) Tablet	- [95]	- -	[161], [119], [54], [66] [40](NL)
AR-HHD & External	Smartphone (AR) Tablet	Smartphone (AV) Multi-touch desktop display	- -	[91] [42, 43]	- -
SAR & External	SAR (1-side) SAR (3-sided) SAR (1-side)	Desktop display Tablet/Desktop display Multi-touch desktop display	- - -	[147], [129] [36, 52] [2]	- - -
SAR & SAR	SAR(3-sided)	SAR(2/3-sided) (NL)	[36]	-	[44]
VR-HMD & External	Oculus Quest Oculus Rift/HTC Vive Oculus Rift	Large-scale display (NL) Desktop display Tablet	- - -	- - [140]	[32] [87] -
Collaborative Setup	Display 1 (D1)	Display 2 (D2)	Display 3 (D3)	Annotator: All	
AR-HMD & VR-HMD & External	Microsoft HoloLens	Oculus Quest	Desktop display	[107]	

TABLE IX
THE COMBINATIONS OF COLLABORATION MODALITIES IN OUR LITERATURE REVIEW PAPERS

Annotation	Video Streaming	Speech	Virtual pointer	View Frustrum	Gaze	Avatar	Face-to-face	num of papers
✓	-	-	-	-	-	-	-	6
✓	✓	-	-	-	-	-	-	6
✓	-	-	✓	-	-	-	-	1
✓	-	-	-	✓	-	-	-	1
✓	-	-	-	-	-	-	✓	4
✓	✓	✓	-	-	-	-	-	5
✓	✓	-	-	-	-	-	✓	1
✓	-	✓	✓	-	-	-	-	1
✓	✓	✓	-	✓	-	-	-	2
✓	✓	✓	-	-	-	✓	-	1
✓	✓	-	✓	-	-	✓	-	1
✓	✓	-	-	-	-	✓	✓	1
✓	✓	✓	-	-	-	✓	-	1
✓	✓	✓	✓	✓	-	-	-	2
✓	✓	✓	-	-	✓	✓	-	2

with the system, collaborate with other users, or perform tasks. Scene navigation [37], [52], [98], panning [72], zooming [72], rotation [17], [98], [156], scaling [17], [156], drawing [52], selection, and deleting annotations [98] were the most common actions used to complete a task.

- *Search tasks:* One of the most common tasks for evaluating annotations is the search task [85], [113], [143]. In search tasks, users are asked to search for a specific scene object or annotation and then report the found object via audio feedback, pressing a physical button [113], [143], distinguishing the object by annotating it, placing the found object in a determined location [113], etc. Search tasks are usually a sub-task of other tasks such as procedural tasks [10] and question answering tasks [113]. Ren et al. asked users to answer some simple questions involving search and object manipulation, such as "Where is statue No. 62? Locate the statue in the scene." or "Which statue is tagged with the letter A? Locate the statue in the scene." User provided their feedback by pressing a button on the phone, which was the input device of an AlloSphere environment [113].
- *Single Procedural tasks:* In procedural tasks, participants are required to complete a procedure which is a sequence of activities in a defined order. Madson et al. asked their participants to select a sequence of labels indicated by highlighting the labeling number on the left to explore the label placement problem through an AR-HHD [78]. Three other papers in our literature review used button-pressing tasks to evaluate the effect of SAR [80] subliminal annotations [10], and predictive annotations [143]. The task required users to press a series of buttons in a defined order that is highlighted by SAR annotations (a procedural task). Their results showed better performance (faster task completion speed or/and fewer errors or/and less mental effort) for SAR annotations compared to monitor based

instructions [80], using subliminal SAR annotations over the use of only SAR annotations [10], and using predictive line annotations compared to other types of predictive cues [143]. Users may be asked to assemble the parts of a component by following the provided patterns, animations, and paper instructions in a single user setup [90], [144].In a study by Oda et al. users were asked to assemble the physical pieces of an engine [98] by following animation annotation is AR.

- *Annotation creation tasks:* In annotation creation tasks, users were asked to generate an annotation in real-time using different methods and input devices. Annotation creation tasks include drawing a determined pattern such as geometric shapes [69], [115], [138], typing [34], selecting an a pre-defined annotation [87], [151], taking screenshots of the scenes [103], entering a textual note annotation [32], adding a label, and recording [120] a voice annotation [66]. In [69], participants were asked to trace a square AR marker using two inputs (multi-touch gestures and pen) for a handheld AR device. In another study, by Ericson et al. participants were asked to create 3D-registered labels in order to add English translations to a Japanese rice cooker using multi-touch handheld AR [34]. In [115], participants reproduced free-form hand drawings of four given patterns (WAVE, CIRCLE, TRIANGLE, and RECTANGLE) with different levels of smoothness. In another study, participants were asked to create a virtual room described in the input text by selecting, placing, texturing objects within the room and then annotating the 3D objects by assigning them to corresponding text segments for generating spatial hypertexts [1]. Taking screenshots is another type of annotation creation task. In [103] users were generating enhanced image annotations by first taking a screenshot of interesting information in the scene, then marking it using drawn sketches, and finally describing it via a text label.
- *Providing Assistance:* A common use-case of annotation in XR environments is complex physical tasks where a remote expert may need to assist a local user in performing the physical tasks. The remote expert guides actions on objects in the local user's environment through various modalities such as speech [73], annotations, gaze, avatar, pointer, and gestures [145]. A survey on physical tasks that investigated 215 papers in AR/MR remote collaboration reported the use of five main physical tasks in their corpus: providing assistance, assembly, training, maintenance, repair, and Co-design.

  Several previous studies used assembly tasks in their annotation system to explore the effect of different annotation forms on user performance. In their systems, local users were asked to assemble the parts of a component by following remote users' instructions in a collaborative setup [98], [131], [156]. Another physical task that required assistance was maintenance. In [84], participants were asked to replace a physical component that was connected to several others by following the remote user's guidelines on how to perform the task (e.g., augmented annotations).

VII. Annotation Rendering Techniques

In this section, we explored *RQ5*, which shows some main techniques that have been used for annotation in XR. We should note that many previous studies that have used annotation in their system did not specify or focus on any specific method for their annotation systems [70], [107]. This shortcoming in communicating the specifics of used annotation forms raises the need for more research on finding annotation methods and classifying them for future studies. The following shows these main techniques in our reviewed corpus.

- *Beautification process:* Several previous studies used a beautification method to make hand drawing annotations perfect. In this method, after the drawing annotation is completed, the completed drawing will go through a process of beautification and will be transformed. For example, a free-form hand drawn arrow annotation will be replaced with parametrized straightened standard arrows with corresponding orientation and position [19], [95], [96], [97].
- *Placement management:* View placement techniques are commonly used to control the annotation placement of objects in 3D environments and to avoid overlaps between each other and the world [132], [134]. Label placement for 3D environments is a NP-hard problem that was the subject of several previous studies in our literature review [7], [78], [132], [134]. Finding the optimum position of a label depends on several factors such as field of view [63], motion, separation distance, and readability of annotations (size, font, overlap, crowding/density) [7], [132], and avoiding occluding important real-world landmarks [132], [133].

  Azuma et al. compared four placement algorithms (Adaptive Simulated Annealing (ASA), Greedy (G), Gradient Descent (GD), Clustering (C), and None (N)) to find the best algorithm that automatically places annotations in a way that allows users to read AR data tags more quickly and accurately. Their results showed the cluster-based method had the best average placement accuracy with relatively moderate computation time [7]. Another study by Tatzgern et al. used two techniques for placement of virtual labels in AR: center-based, where labels move along a 3D pole stuck into the anchor of an annotated object, and plane-based, where labels are placed in a dominant plane at run-time [132]. Each approach has its pros and cons. The center-based approach is suggested to be used only if anchor points are well distributed around the object to avoid clutter and long poles. The plane-based approach that enforces a certain spacing between annotations is suggested to be used when there are few enough labels that they do not occlude each other, but they are non-uniformly distributed around the object. In the second case, changing the viewpoint might lead to occlusion, in which freezing the orientation of labels can help. The latter solution should be used only when the viewpoint changes with small angles; otherwise, it leads to perspective distortion problems.
- *Network annotation:* Mehler et al. introduced VANOTATOR, which is an annotation system that used an annotation network connecting related information units (texts, images, and their segments 3D models) via direct lines as

network edges, the cube as network nodes. When the user pointed to a single node N, the sets of nodes directly linked to N were highlighted [88].

- *Filtering:* Filtering techniques aim to reduce the amount of data presented in user view [133]. This causes reducing cognitive load and avoiding cluttering [124]. Wither et al. classified strategies for filtering annotations into four groups: user-controlled, spatially-controlled, time-controlled, and information-filtered [152]. Spatial and user-controlled (knowledge-based) filtering are two common filtering methods that discard the data points when they are beyond a certain distance or by selecting information based on user preference in a respective order [133]. Each technique has pros and cons. Spatial filtering may lead to information loss, and knowledge-based filtering doesn't guarantee that the amount of data is sufficiently reduced [133]. Time-controlled filtering allows annotations to be visible for a certain amount of time and are often temporally sorted [152]. Temporal sorting, which means arranging events in time and via a step-by-step technique, helps reviewers avoid reviewing complex annotations that display all annotations at once by visiting a larger set of simpler annotations step by step [84]. Marques et al. used temporal sorting to display annotation during the time to assist an on-site technician during asynchronous maintenance tasks [84]. Finally, the information-controlled method filters the visible annotations based on each application [133]. Lindlbauer et al. introduced an approach that automatically controls how much information should be displayed based on the application. For example, when the application requires a low cognitive load of the user (tasks such as doodling and cutting fruits), the system displays more information [74].
  In another study, the annotation system filtered information based on various criteria such as annotator, the region of interest [124], and pre-group categories [124]. The AR handheld system is designed by schall et al. allows users to filter unwanted information and select a region of interest, and then turn on 3D features based on pre-grouping into asset categories (gas, water, buildings, etc.) [122].
- *Clustering:* Hierarchical clustering is a technique that selects annotations and groups them based on different attributes, then represents data based on the created hierarchy. Two common clustering methods group data based on spatial and semantic attributes. An example of spatial clustering is the annotation system used by García-Pereira et al. which shows the combination of annotations for each real-world location as a single virtual sphere using a handheld AR device. Each annotation includes the type, the annotator's name, and the date and time of creation [40]. In another study, authors used a hierarchical clustering that allows users to select and represent the appropriate level of information based on user-defined preferences [133].
- *Predictive annotations:* Volmer et al. investigated the effectiveness of four predictive annotation forms on user performance in a SAR environment. Their method highlighted the current and next target object for manipulation. The next target was highlighted using four different predictive cueing forms: BLINK around the target annotation, COLOR that highlights the target object with a different color, ARROW heads to the next target object), and LINE that connects the current object to the next object. Their results confirmed the usefulness of predictive cues on user performance and mental load. The study's outcomes revealed that utilizing a line as a visual cue to indicate the next target proved to be the most efficient cue in improving task completion time and reducing mental effort [143].

VIII. Challenges

In this section, we focused on *RQ6* and explored the challenges of annotation in XR, which is a growing field of research (see Fig. 3).

- *Outdoor use case:* Although 81 out of 103 XR display technologies used in our literature review fell within XR-HMDs or AR-HHDs which could be used outside, especially mobile wearable and mobile handheld devices; only 18 of them mentioned or used their prototype in outdoor environments [43], [88], [124], [135], [151]. This gets worse if we consider only the annotator side environment, which limits us to three studies [66], [122], [151]. One reason for the lack of outdoor experiments in AR annotation systems could be the challenges that current XR systems are facing, such as lighting, heat, mobility, battery, safety issues, noise, and the problem that input devices might cause [57], [95], [151]. Users in a study by Wither et al. explored outdoor annotation through a non-mobile AR-HMD display, by wearing an Alien-ware laptop on their back [151]. In addition, three out of four papers were prototype based papers and did not conduct an experiment. One possible reason, in addition to mentioned reasons for the lack of outdoor XR studies, could be a lack of education on how to evaluate outdoor XR experiments, how to properly design tasks and control external factors.
  Annotator: Out of 33 collaborative papers in our corpus, 13 papers only supported the non-XR user to generate annotations. In such annotation systems, the remote user authored annotations on a non-XR side (e.g., a desktop display), and then generated annotations were rendered and displayed on the XR display for the local user. Almost half of the papers that supported authoring annotations for XR users (9 papers) provided the ability for the VR user to generate annotations. Regarding AR environments, three allowed AR-HMD users to annotate, five supported annotating by AR-HHD users, and two generated annotations for SAR environments. Although 11 out of 33 papers used a system that allowed both users to annotate, only 3 were the subject of XR-HMDs (see Table VIII). This highlights the need for more research on XR-HMD annotation systems, allowing both collaborators to author annotations.
- *Interpreting user's behaviors in 3D interfaces:* A previous study by [97] investigated user behaviors while drawing sketches to help a collaborator in 2D environments. They

found that arrows and circles were the most commonly generated annotations. Understanding user behaviors and preferences is a challenging problem, and 3D environments such as XR make it even more difficult. Some of the possible research questions that require more investigation are: which types of 2D/3D models are preferred by the users in a 3D user interface? What are the most effective annotation types for indicating an object in XR? Do user preferences for annotation types change if it is a single-user system compared to a collaborative XR system? Do the preferred annotations change depending on the task? As an example, Nuernberger et al. observed user behavior while drawing arrows, and their results showed that most of their users anchored the arrows on their heads for the referencing tasks, while for the action tasks, arrows were anchored on their tails.

Another interesting challenge that requires further research is how to render the 2D annotations in 3D in a way that better matches the original intention of the user. As an example, Nuernberger et al. studied how to present circles and arrows that are drawn in a 2D space in a 3D environment when the line of sight changes for the annotated object. In their approach, they used the normal of the 2D scene at the anchor point of the arrow to make a perspective visualization of that in 3D [97].

- *View management:* Although many previous studies investigated textual label placement and view management, the research on the placement of other forms of annotations in a 3D environment is minimal. Also, it will be interesting to explore view management, filtering, and clustering techniques for a combination of different annotation forms (e.g., labels, 3D models, images, audio, etc.).
- *Wide FOV AR:* Most AR devices provide only a small FOV for the users. However, with growing advances in AR technologies achieving wide FOV would be eminent. This makes it important to study this concern. Our investigation showed only two previous studies in our corpus examined annotation for wide FOV AR [63], [113]. Kishishita et al. investigated annotation placement in a wide field of view AR using a customized AR device and found that the wide field of view affects the annotation tasks. In the second study, authors compared annotations in three FOV setups (Full FOV and Small FOV: $45° * 30°$ FOV and $30° * 17.5°$ FOV) using tracked glasses and AlloShere setup. Their results showed Full FOV helped participants locate annotations in less time than small FOV; however, the accuracy decreased in Full FOV. These studies confirm the effect of the FOV on virtual annotations and the need for further investigation since the existing research and solutions for annotations may behave differently in a wide FOV.
- *Cooperative gestures:* Cooperative gestures refer to the combined gestural inputs of multiple users that can be used to initiate various actions [75], such as annotating target objects over large distances. Previous studies have shown that manipulating objects over large distances raises precision and fatigue problems [3], [51]. Although many previous studies used gestures for authoring annotations, none investigated cooperative gestures for generating annotations. However, two previous studies used cooperative gestures for well-sized multi-touch displays to investigate data manipulation and navigation while avoiding the fatigue problem problem [75], [93]. In addition, although some gestures, such as air-tapping and pinch, were used in previous XR studies for authoring annotations, an elicitation study that explores the effect of various gestures on creating annotations can be beneficial.
- *Asynchronous in immersive XR:* How people can collaborate asynchronously by using annotation in AR/VR-HMDs needs further investigation. This survey found 33 papers that dealt with collaboration, where only 7 were asynchronous (five used AR-HHDs). This modality provides an important paradigm where people would work in different time zone, at their own pace, and without assistance from the user that initially added annotations. It is interesting and tempting to use the same options that google docs provide in a 3D environment, where data can be compromised in multiple dimensions (e.g., 3D data visualization); however, the way a user will annotate and another user will later add or comment to those annotations to complete tasks is under-explored.

IX. Conclusion

This article presented a literature survey of annotation in selected XR research publications between 2001 and 2021. We reviewed a total of 103 papers through a systematic review and categorized them based on their output device, input device, annotation types, annotation targets, collaboration type, and annotator. Based on this review, we extracted a current research agenda and discussed challenges and remaining research areas that have yet to be covered. Our surveyed paper can provide important guidelines for developing future XR systems using annotation. First, it can help authors develop their annotation systems or collaborative systems in a more structured and standard format. Second, the discussed challenges and provided statistics can help researchers identify research gaps and possible research areas for further investigation.

Acknowledgments

We would like to thank Mr. Hamilton Chevez for the database site and Ms. Lauren Mangus for her editing help.

References

[1] G. Abrami, A. Henlein, A. Kett, and A. Mehler, “Text2SceneVR: Generating hypertexts with VAnnotatoR as a pre-processing step for text2scene systems,” in *Proc. 31st ACM Conf. Hypertext Soc. Media*, 2020, pp. 177–186.

[2] M. Adcock, S. Anderson, and B. Thomas, “Remotefusion: Real time depth camera fusion for remote collaboration on physical tasks,” in *Proc. 12th ACM SIGGRAPH Int. Conf. Virtual-Reality Continuum Appl. Ind.*, 2013, pp. 235–242.

[3] S. Al-Megren, A. Kharrufa, J. Hook, A. Holden, S. Sutton, and P. Olivier, “Comparing fatigue when using large horizontal and vertical multi-touch interaction displays,” in *Proc. 15th IFIP Int. Conf. Hum.-Comput. Interact.*, Bamberg, Germany, 2015, pp. 156–164, Springer.

[4] C. Amma, M. Georgi, T. Lenz, and F. Winnen, “Kinemic wave: A mobile freehand gesture and text-entry system,” in *Proc. CHI Conf. Extended Abstr. Hum. Factors Comput. Syst.*, 2016, pp. 3639–3642.

[5] R. Arora, R. H. Kazi, F. Anderson, T. Grossman, K. Singh, and G. W. Fitzmaurice, “Experimental evaluation of sketching on surfaces in VR,” in *Proc.CHI Conf. Hum. Factors Comput. Syst.*, 2017, pp. 5643–5654.

[6] C. R. Austin, B. Ens, K. A. Satriadi, and B. Jenny, “Elicitation study investigating hand and foot gesture interaction for immersive maps in augmented reality,” *Cartogr. Geogr. Inf. Sci.*, vol. 47, no. 3, pp. 214–228, 2020.

[7] R. Azuma and C. Furmanski, “Evaluating label placement for augmented reality view management,” in *Proc. IEEE 2nd ACM Int. Symp. Mixed Augmented Reality*, 2003, pp. 66–75.

[8] H. Bai, P. Sasikumar, J. Yang, and M. Billinghurst, “A user study on mixed reality remote collaboration with eye gaze and hand gesture sharing,” in *Proc. CHI Conf. Hum. Factors Comput. Syst.*, 2020, pp. 1–13.

[9] R. Ball and C. North, “Effects of tiled high-resolution display on basic visualization and navigation tasks,” in *Proc. CHI Extended Abstr. Hum. Factors Comput. Syst.*, 2005, pp. 1196–1199.

[10] J. Baumeister, M. R. Marner, R. T. Smith, M. Kohler, and B. H. Thomas, “Visual subliminal cues for spatial augmented reality,” in *Proc. IEEE Int. Symp. Mixed Augmented Reality Workshops*, 2015, pp. 4–11.

[11] B. Bell, T. Höllerer, and S. Feiner, “An annotated situation-awareness aid for augmented reality,” in *Proc. 15th Annu. ACM Symp. User Interface Softw. Technol.*, 2002, pp. 213–216.

[12] O. Bimber and R. Raskar, *Spatial Augmented Reality: Merging Real and Virtual Worlds*. Boca Raton, FL, USA: CRC Press, 2005.

[13] O. Bimber and R. Raskar, *Spatial Augmented Reality: Merging Real and Virtual Worlds*. Natick, MA, USA: AK Peters/CRC Press, 2019.

[14] S. Burigat, L. Chittaro, and S. Gabrielli, “Visualizing locations of off-screen objects on mobile devices: A comparative evaluation of three approaches,” in *Proc. 8th Conf. Hum.- Comput. Interact. Mobile Devices Serv.*, 2006, pp. 239–246.

[15] N. R. Caluya and M. E. C. Santos, “Kantenbouki VR: A virtual reality authoring tool for learning localized weather reporting,” in *Proc. IEEE Conf. Virtual Reality 3D User Interfaces*, 2019, pp. 866–867.

[16] S. Casas, J. Gimeno, P. Casanova-Salas, J. V. Riera, and C. Portalés, “Virtual and augmented reality for the visualization of summarized information in smart cities: A use case for the city of Dubai,” in *Smart Systems Design, Applications, and Challenges*. Hershey, PA, USA: IGI Global, 2020, pp. 299–325.

[17] N. Chaconas and T. Höllerer, “An evaluation of bimanual gestures on the microsoft hololens,” in *Proc. IEEE Conf. Virtual Reality 3D User Interfaces*, 2018, pp. 1–8.

[18] T. Chandler et al., “Immersive analytics,” in *Proc. Big Data Visual Analytics*, 2015, pp. 1–8.

[19] Y. S. Chang, B. Nuernberger, B. Luan, and T. Höllerer, “Evaluating gesture-based augmented reality annotation,” in *Proc. IEEE Symp. 3D User Interfaces*, 2017, pp. 182–185.

[20] Y. S. Chang, B. Nuernberger, B. Luan, T. Höllerer, and J. O’Donovan, “Gesture-based augmented reality annotation,” in *Proc. IEEE Virtual Reality*, 2017, pp. 469–470.

[21] G. Chin Jr, O. A. Kuchar, and K. E. Wolf, “Exploring the analytical processes of intelligence analysts,” in *Proc. SIGCHI Conf. Hum. Factors Comput. Syst.*, 2009, pp. 11–20.

[22] E. K. Choe, B. Lee, and M. C. Schraefel, “Characterizing visualization insights from quantified selfers’ personal data presentations,” *IEEE Comput. Graph. Appl.*, vol. 35, no. 4, pp. 28–37, Jul./Aug. 2015.

[23] M. Cordeil, T. Dwyer, K. Klein, B. Laha, K. Marriott, and B. H. Thomas, “Immersive collaborative analysis of network connectivity: Cave-style or head-mounted display?,” *IEEE Trans. Visual. Comput. Graph.*, vol. 23, no. 1, pp. 441–450, Jan. 2017.

[24] S. D’Angelo and D. Gergle, “Gazed and confused: Understanding and designing shared gaze for remote collaboration,” in *Proc. CHI Conf. Hum. Factors Comput. Syst.*, 2016, pp. 2492–2496.

[25] K. Danyluk, T. T. Ulusoy, W. Wei, and W. Willett, “Touch and beyond: Comparing physical and virtual reality visualizations,” *IEEE Trans. Vis. Comput. Graph.*, vol. 28, no. 4, pp. 1930–1940, Apr. 2022.

[26] B. Dao and J. L. Gabbard, “Early steps towards understanding text legibility in handheld augmented reality,” in *Proc. IEEE Virtual Reality*, 2013, pp. 159–160.

[27] R. A. J. de Belen, H. Nguyen, D. Filonik, D. Del Favero, and T. Bednarz, “A systematic review of the current state of collaborative mixed reality technologies: 2013–2018,” *AIMS Electron. Elect. Eng.*, vol. 3, no. 2, pp. 181–223, 2019.

[28] J. Dominic and A. Robb, “Exploring effects of screen-fixed and world-fixed annotation on navigation in virtual reality,” in *Proc. IEEE Conf. Virtual Reality 3D User Interfaces*, 2020, pp. 607–615.

[29] S. Dow, M. Mehta, E. Harmon, B. MacIntyre, and M. Mateas, “Presence and engagement in an interactive drama,” in *Proc. SIGCHI Conf. Hum. Factors Comput. Syst.*, 2007, pp. 1475–1484.

[30] A. Dünser, R. Grasset, and M. Billinghurst, “A survey of evaluation techniques used in augmented reality studies,” Human Interface Technology Laboratory New Zealand, Tech. Rep. TR-2008-02, 2008.

[31] L. Edlin, Y. Liu, N. Bryan-Kinns, and J. Reiss, “Exploring augmented reality as craft material,” in *Proc. Int. Conf. Hum.- Comput. Interact.*, Springer, 2020, pp. 54–69.

[32] L. Emerson, R. Lipinski, H. Shirey, T. Malloy, and T. Marrinan, “Enabling collaborative interaction with 360° panoramas between large-scale displays and immersive headsets,” in *Proc. IEEE Int. Symp. Mixed Augmented Reality Adjunct*, 2021, pp. 183–188.

[33] B. Ens et al., “Revisiting collaboration through mixed reality: The evolution of groupware,” *Int. J. Hum.- Comput. Stud.*, vol. 131, pp. 81–98, 2019.

[34] M. Ericson, T. Taketomi, G. Yamamoto, G. Klinker, C. Santos, and H. Kato, “[poster] towards estimating usability ratings of handheld augmented reality using accelerometer data,” in *Proc. IEEE Int. Symp. Mixed Augmented Reality*, 2015, pp. 196–197.

[35] R. Fedorov, D. Frajberg, and P. Fraternali, “A framework for outdoor mobile augmented reality and its application to mountain peak detection,” in *Proc. Int. Conf. Augmented Reality, Virtual Reality Comput. Graph.*, Springer, 2016, pp. 281–301.

[36] H. S. Ferdous et al., “What’s happening at that hip?” evaluating an on-body projection based augmented reality system for physiotherapy classroom,” in *Proc. CHI Conf. Hum. Factors Comput. Syst.*, 2019, pp. 1–12.

[37] M. Fiorentino, R. de Amicis, G. Monno, and A. Stork, “Spacedesign: A mixed reality workspace for aesthetic industrial design,” in *Proc. Int. Symp. Mixed Augmented Reality*, 2002, pp. 86–318.

[38] A. Fonnet and Y. Prie, “Survey of immersive analytics,” *IEEE Trans. Vis. Comput. Graph.*, vol. 27, no. 3, pp. 2101–2122, Mar. 2021.

[39] I. García-Pereira, J. Gimeno, P. Morillo, and P. Casanova-Salas, “A taxonomy of augmented reality annotations,” in *Proc. 15th Int. Joint Conf. Comput. Vis., Imag. Comput. Graph. Theory Appl.*, 2020, pp. 412–419.

[40] I. García-Pereira, C. Portalés, J. Gimeno, and S. Casas, “A collaborative augmented reality annotation tool for the inspection of prefabricated buildings,” *Multimedia Tools Appl.*, vol. 79, no. 9, pp. 6483–6501, 2020.

[41] D. Gasques et al., “ARTEMIS: A collaborative mixed-reality system for immersive surgical telementoring,” in *Proc. CHI Conf. Hum. Factors Comput. Syst.*, 2021, pp. 1–14.

[42] S. Gauglitz, B. Nuernberger, M. Turk, and T. Höllerer, “In touch with the remote world: Remote collaboration with augmented reality drawings and virtual navigation,” in *Proc. 20th ACM Symp. Virtual Reality Softw. Technol.*, 2014, pp. 197–205.

[43] S. Gauglitz, B. Nuernberger, M. Turk, and T. Höllerer, “World-stabilized annotations and virtual scene navigation for remote collaboration,” in *Proc. 27th Annu. ACM Symp. User Interface Softw. Technol.*, 2014, pp. 449–459.

[44] M. Haller, P. Brandl, D. Leithinger, J. Leitner, T. Seifried, and M. Billinghurst, “Shared design space: Sketching ideas using digital pens and a large augmented tabletop setup,” in *Proc. Int. Conf. Artif. Reality Telexistence*, Springer, 2006, pp. 185–196.

[45] D. Handa, H. Ishii, and H. Shimoda, “Enhancing metric perception with RGB-D camera,” in *Proc. Int. Conf. Virtual, Augmented Mixed Reality*, Springer, 2013, pp. 23–31.

[46] M. Haouach, G. Venturini, and C. Guinot, “A 3D hypermedia with biomedical stereoscopic images: From creation to exploration in virtual reality,” in *Proc. 20th ACM Conf. Hypertext Hypermedia*, 2009, pp. 337–338.

[47] J. D. Hart, T. Piumsomboon, G. A. Lee, and M. Billinghurst, “Sharing and augmenting emotion in collaborative mixed reality,” in *Proc. IEEE Int. Symp. Mixed Augmented Reality Adjunct*, 2018, pp. 212–213.

[48] S. J. Henderson and S. Feiner, “Evaluating the benefits of augmented reality for task localization in maintenance of an armored personnel carrier turret,” in *Proc. IEEE 8th Int. Symp. Mixed Augmented Reality*, 2009, pp. 135–144.

[49] S. J. Henderson and S. K. Feiner, “Augmented reality in the psychomotor phase of a procedural task,” in *Proc. 10th IEEE Int. Symp. Mixed Augmented Reality*, 2011, pp. 191–200.

[50] J. Hertel, S. Karaosmanoglu, S. Schmidt, J. Bräker, M. Semmann, and F. Steinicke, "A taxonomy of interaction techniques for immersive augmented reality based on an iterative literature review," in *Proc. IEEE Int. Symp. Mixed Augmented Reality*, 2021, pp. 431–440.

[51] J. D. Hincapié-Ramos, X. Guo, P. Moghadasian, and P. Irani, "Consumed endurance: A metric to quantify arm fatigue of mid-air interactions," in *Proc. SIGCHI Conf. Hum. Factors Comput. Syst.*, 2014, pp. 1063–1072.

[52] T. Hoang, M. Reinoso, Z. Joukhadar, F. Vetere, and D. Kelly, "Augmented studio: Projection mapping on moving body for physiotherapy education," in *Proc. CHI Conf. Hum. Factors Comput. Syst.*, 2017, pp. 1419–1430.

[53] T. Höllerer, J. Kuchera-Morin, and X. Amatriain, "The allosphere: A large-scale immersive surround-view instrument," in *Proc. Workshop Emerg. Displays Technol.: Images Beyond: Future Displays Interacton*, 2007, pp. 3–es.

[54] A. Irlitti, S. Von Itzstein, L. Alem, and B. Thomas, "Tangible interaction techniques to support asynchronous collaboration," in *Proc. IEEE Int. Symp. Mixed Augmented Reality*, 2013, pp. 1–6.

[55] H. Ishii et al., "Augmented reality applications for nuclear power plant maintenance work," in *Proc. CD-ROM Int. Symp. Symbiotic Nucl. Power Syst.*, 2007, pp. 262–268.

[56] D. Kang, J.-H. Choi, and H. Hwang, "Autostereoscopic 3d display system for 3d medical images," *Appl. Sci.*, vol. 12, no. 9, pp 4288, 2022.

[57] V. Kasapakis, D. Gavalas, and D. Elena, "Robust outdoors marker-based augmented reality applications: Mitigating the effect of lighting sensitivity," in *Proc. Int. Conf. Augmented Reality, Virtual Reality Comput. Graph.*, Springer, 2018, pp. 423–431.

[58] C. Ke, B. Kang, D. Chen, and X. Li, "An augmented reality-based application for equipment maintenance," in *Proc. Int. Conf. Affect. Comput. Intell. Interact.*, Springer, 2005, pp. 836–841.

[59] H. Kim, G. Reitmayr, and W. Woo, "Interactive annotation on mobile phones for real and virtual space registration," in *Proc. IEEE 10th Int. Symp. Mixed Augmented Reality*, 2011, pp. 265–266.

[60] S. Kim and G. J. Kim, "Using keyboards with head mounted displays," in *Proc. ACM SIGGRAPH Int. Conf. Virtual Reality Continuum Appl. Ind.*, 2004, pp. 336–343.

[61] S. Kim, G. A. Lee, and N. Sakata, "Comparing pointing and drawing for remote collaboration," in *Proc. IEEE Int. Symp. Mixed Augmented Reality*, 2013, pp. 1–6.

[62] Y.-S. Kim et al., "Inking your insights: Investigating digital externalization behaviors during data analysis," in *Proc. ACM Int. Conf. Interactive Surfaces Spaces*, 2019, pp. 255–267.

[63] N. Kishishita, K. Kiyokawa, J. Orlosky, T. Mashita, H. Takemura, and E. Kruijff, "Analysing the effects of a wide field of view augmented reality display on search performance in divided attention tasks," in *Proc. IEEE Int. Symp. Mixed Augmented Reality*, 2014, pp. 177–186.

[64] M. Koeda, Y. Matsumoto, and T. Ogasawara, "Annotation-based assistance system for unmanned helicopter with wearable augmented reality environment," in *Proc. IEEE 3rd ACM Int. Symp. Mixed Augmented Reality*, 2004, pp. 288–289.

[65] V. Kühn, G. Abrami, and A. Mehler, "WikNectVR: A gesture-based approach for interacting in virtual reality based on wiknect and gestural writing," in *Proc. Int. Conf. Hum.- Comput. Interact.*, Springer, 2020, pp. 299–312.

[66] T. Langlotz, H. Regenbrecht, S. Zollmann, and D. Schmalstieg, "Audio stickies: Visually-guided spatial audio annotations on a mobile augmented reality platform," in *Proc. 25th Australian Comput.- Hum. Interact. Conf.: Augmentation, Appl., Innov., Collaboration*, 2013, pp. 545–554.

[67] B. Lee, X. Hu, M. Cordeil, A. Prouzeau, B. Jenny, and T. Dwyer, "Shared surfaces and spaces: Collaborative data visualisation in a co-located immersive environment," *IEEE Trans. Vis. Comput. Graph.*, vol. 27, no. 2, pp. 1171–1181, Feb. 2021.

[68] B. Lee, R. H. Kazi, and G. Smith, "Sketchstory: Telling more engaging stories with data through freeform sketching," *IEEE Trans. Visual. Comput. Graph.*, vol. 19, no. 12, pp. 2416–2425, Dec. 2013.

[69] G. A. Lee and M. Billinghurst, "Assistive techniques for precise touch interaction in handheld augmented reality environments," in *Proc. 11th ACM SIGGRAPH Int. Conf. Virtual-Reality Continuum Appl. Ind.*, 2012, pp. 279–285.

[70] J. Leo et al., "Interactive cardiovascular surgical planning via augmented reality," in *Proc. Asian CHI Symp.*2021, 2021, pp. 132–135.

[71] K.-C. Lien, B. Nuernberger, M. A. Turk, and T. Höllerer, "2D-3D cosegmentation for AR-based remote collaboration," in *Proc. IEEE Int. Symp. Mixed Augmented Reality*, 2015, pp. 184–185.

[72] C. Lin et al., "A first-person mentee second-person mentor AR interface for surgical telementoring," in *Proc. IEEE Int. Symp. Mixed Augmented Reality Adjunct*, 2018, pp. 3–8.

[73] C. Lin et al., "How about the mentor? Effective workspace visualization in AR telementoring," in *Proc. IEEE Conf. Virtual Reality 3D User Interfaces*, 2020, pp. 212–220.

[74] D. Lindlbauer, A. M. Feit, and O. Hilliges, "Context-aware online adaptation of mixed reality interfaces," in *Proc. 32nd Annu. ACM Symp. User Interface Softw. Technol.*, 2019, pp. 147–160.

[75] C. Liu, O. Chapuis, M. Beaudouin-Lafon, and E. Lecolinet, "CorEach: Cooperative gestures for data manipulation on wall-sized displays," in *Proc. CHI Conf. Hum. Factors Comput. Syst.*, 2017, pp. 6730–6741.

[76] W. Luo, E. Goebel, P. Reipschläger, M. O. Ellenberg, and R. Dachselt, "Exploring and slicing volumetric medical data in augmented reality using a spatially-aware mobile device," in *Proc. IEEE Int. Symp. Mixed Augmented Reality Adjunct*, 2021, pp. 334–339.

[77] B. MacIntyre, A. Hill, H. Rouzati, M. Gandy, and B. Davidson, "The argon ar web browser and standards-based ar application environment," in *Proc. 10th IEEE Int. Symp. Mixed Augmented Reality*, 2011, pp. 65–74.

[78] J. B. Madsen, M. Tatzqern, C. B. Madsen, D. Schmalstieg, and D. Kalkofen, "Temporal coherence strategies for augmented reality labeling," *IEEE Trans. Visual. Comput. Graph.*, vol. 22, no. 4, pp. 1415–1423, Apr. 2016.

[79] T. Mahmood, W. Fulmer, N. Mungoli, J. Huang, and A. Lu, "Improving information sharing and collaborative analysis for remote geospatial visualization using mixed reality," in *Proc. IEEE Int. Symp. Mixed Augmented Reality*, 2019, pp. 236–247.

[80] M. R. Marner, A. Irlitti, and B. H. Thomas, "Improving procedural task performance with augmented reality annotations," in *Proc. IEEE Int. Symp. Mixed Augmented Reality*, 2013, pp. 39–48.

[81] B. Marques, S. Silva, J. Alves, T. Araujo, P. Dias, and B. S. Santos, "A conceptual model and taxonomy for collaborative augmented reality," *IEEE Trans. Visual. Comput. Graph.*, vol. 28, no. 12, pp. 5113–5133, Dec. 2022.

[82] B. Marques, S. Silva, J. Alves, A. Rocha, P. Dias, and B. S. Santos, "Remote collaboration in maintenance contexts using augmented reality: Insights from a participatory process," *Int. J. Interactive Des. Manuf.*, vol. 16, pp. 419–438, 2022.

[83] B. Marques, S. Silva, P. Dias, and B. S. Santos, "Do hand-held devices have a future in augmented reality real-life remote tasks? refections on impact/acceptance versus head-mounted displays," in *Proc. Eur. Conf. Comput.-Supported Cooperative Work*, 2022, pp. 1–9.

[84] 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.

[85] T. Matos, R. Nóbrega, R. Rodrigues, and M. Pinheiro, "Dynamic annotations on an interactive web-based 360° video player," in *Proc. 23rd Int. ACM Conf. 3D Web Technol.*, 2018, pp. 1–4.

[86] A. McNamara and C. Kabeerdoss, "Mobile augmented reality: Placing labels based on gaze position," in *Proc. IEEE Int. Symp. Mixed Augmented Reality*, 2016, pp. 36–37.

[87] M. L. Medeiros, "Marking the city: Interactions in multiple space scales in virtual reality," in *Proc. IEEE Int. Symp. Mixed Augmented Reality Adjunct*, 2019, pp. 465–469.

[88] A. Mehler, G. Abrami, C. Spiekermann, and M. Jostock, "VannotatoR: A framework for generating multimodal hypertexts," in *Proc. 29th Hypertext Soc. Media*, 2018, pp. 150–154.

[89] R. R. Mohanty, U. H. Bohari, and E. Ragan, "Kinesthetically augmented mid-air sketching of multi-planar 3D curve-soups," in *Proc. Int. Des. Eng. Tech. Conf. Comput. Inf. Eng. Conf.*, 2018, pp. 1–11.

[90] P. Mohr, B. Kerbl, M. Donoser, D. Schmalstieg, and D. Kalkofen, "Retargeting technical documentation to augmented reality," in *Proc. 33rd Annu. ACM Conf. Hum. Factors Comput. Syst.*, 2015, pp. 3337–3346.

[91] P. Mohr-Ziak, S. Mori, T. Langlotz, B. H. Thomas, D. Schmalstieg, and D. Kalkofen, "Mixed reality light fields for interactive remote assistance," in *Proc. ACM CHI Conf. Hum. Factors Comput. Syst.*, 2020, pp. 1–12.

[92] J. Mooser, S. You, and U. Neumann, "Real-time object tracking for augmented reality combining graph cuts and optical flow," in *Proc. 6th IEEE ACM Int. Symp. Mixed Augmented Reality*, 2007, pp. 145–152.

[93] M. R. Morris, A. Huang, A. Paepcke, and T. Winograd, "Cooperative gestures: Multi-user gestural interactions for co-located groupware," in *Proc. SIGCHI Conf. Hum. Factors Comput. Syst.*, 2006, pp. 1201–1210.

[94] S. Noh, H.-S. Yeo, and W. Woo, "An HMD-based mixed reality system for avatar-mediated remote collaboration with bare-hand interaction," in *Proc. Int. Conf. Artif. Reality Telexistence Eurographics Symp. Virtual Environ.*, 2015, pp. 61–68.

[95] B. Nuernberger, K.-C. Lien, L. Grinta, C. Sweeney, M. Turk, and T. Höllerer, "Multi-view gesture annotations in image-based 3D reconstructed scenes," in *Proc. 22nd ACM Conf. Virtual Reality Softw. Technol.*, 2016, pp. 129–138.

[96] B. Nuernberger, K.-C. Lien, T. Höllerer, and M. Turk, "Anchoring 2D gesture annotations in augmented reality," in *Proc. IEEE Virtual Reality*, 2016, pp. 247–248.

[97] B. Nuernberger, K.-C. Lien, T. Höllerer, and M. Turk, "Interpreting 2D gesture annotations in 3D augmented reality," in *Proc. IEEE Symp. 3D User Interfaces*, 2016, pp. 149–158.

[98] O. Oda, C. Elvezio, M. Sukan, S. Feiner, and B. Tversky, "Virtual replicas for remote assistance in virtual and augmented reality," in *Proc. 28th Annu. ACM Symp. User Interface Softw. Technol.*, 2015, pp. 405–415.

[99] T. Olsson and M. Salo, "Online user survey on current mobile augmented reality applications," in *Proc. IEEE 10th Int. Symp. Mixed Augmented Reality*, 2011, pp. 75–84.

[100] A. M. Peña, E. D. Ragan, and J. Kang, "Designing educational virtual environments for construction safety: A case study in contextualizing incident reports and engaging learners," in *Proc. Int. Conf. Hum.- Comput. Interact.*, Springer, 2019, pp. 338–354.

[101] D. N. E. Phon, M. B. Ali, and N. D. Abd Halim, "Collaborative augmented reality in education: A review," in *Proc. Int. Conf. Teach. Learn. Comput. Eng.*, 2014, pp. 78–83.

[102] S. Pick, S. Gebhardt, B. Hentschel, and T. W. Kuhlen, "Scalable metadata in-and output for multi-platform data annotation applications," in *Proc. IEEE Virtual Reality*, 2015, pp. 261–262.

[103] S. Pick, B. Weyers, B. Hentschel, and T. W. Kuhlen, "Design and evaluation of data annotation workflows for cave-like virtual environments," *IEEE Trans. Visual. Comput. Graph.*, vol. 22, no. 4, pp. 1452–1461, Apr. 2016.

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

[105] T. Piumsomboon, G. A. Lee, and M. Billinghurst, "Snow dome: A multi-scale interaction in mixed reality remote collaboration," in *Proc. Extended Abstr. CHI Conf. Hum. Factors Comput. Syst.*, 2018, pp. 1–4.

[106] J. Polvi et al., "Handheld guides in inspection tasks: Augmented reality versus picture," *IEEE Trans. Visual. Comput. Graph.*, vol. 24, no. 7, pp. 2118–2128, Jul. 2018.

[107] I. Radu, T. Joy, and B. Schneider, "Virtual makerspaces: Merging AR/VR/MR to enable remote collaborations in physical maker activities," in *Proc. Extended Abstr. CHI Conf. Hum. Factors Comput. Syst.*, 2021, pp. 1–5.

[108] R. Raskar and K.-L. Low, "Interacting with spatially augmented reality," in *Proc. 1st Int. Conf. Comput. Graph., Virtual Reality Visualisation*, 2001, pp. 101–108.

[109] R. Raskar, G. Welch, and H. Fuchs, "Spatially augmented reality," in *Proc. Int. Workshop Augmented Reality: Placing Artif. Objects Real Scenes*, 1999, pp. 64–71.

[110] M. Rebol et al., "Remote assistance with mixed reality for procedural tasks," in *Proc. IEEE Conf. Virtual Reality 3D User Interfaces Abstr. Workshops*, 2021, pp. 653–654.

[111] G. Reitmayr, E. Eade, and T. W. Drummond, "Semi-automatic annotations in unknown environments," in *Proc. IEEE 6th ACM Int. Symp. Mixed Augmented Reality*, 2007, pp. 67–70.

[112] D. Ren, M. Brehmer, B. Lee, T. Höllerer, and E. K. Choe, "ChartAccent: Annotation for data-driven storytelling," in *Proc. IEEE Pacific Visual. Symp.*, 2017, pp. 230–239.

[113] D. Ren, T. Goldschwendt, Y. Chang, and T. Höllerer, "Evaluating wide-field-of-view augmented reality with mixed reality simulation," in *Proc. IEEE Virtual Reality*, 2016, pp. 93–102.

[114] M. Rodrigue, A. Waranis, T. Wood, and T. Höllerer, "Mixed reality simulation with physical mobile display devices," in *Proc. IEEE Virtual Reality*, 2015, pp. 105–110.

[115] H. Romat, A. Fender, M. Meier, and C. Holz, "Demonstrating high-precision and high-fidelity digital inking for virtual reality," in *Proc. IEEE Conf. Virtual Reality 3D User Interfaces Abstr. Workshops*, 2021, pp. 749–750.

[116] H. Romat, A. Fender, M. Meier, and C. Holz, "Flashpen: A high-fidelity and high-precision multi-surface pen for virtual reality," in *Proc. IEEE Virtual Reality 3D User Interfaces*, 2021, pp. 306–315.

[117] H. Romat, A. R. Fender, M. Meier, and C. Holz, "Demonstration of flashpen: A high-fidelity and high-precision multi-surface pen for virtual reality," in *Proc. Extended Abstr. CHI Conf. Hum. Factors Comput. Syst.*, 2021, pp. 1–4.

[118] H. Romat et al., "Activeink: (th) inking with data," in *Proc. CHI Conf. Hum. Factors Comput. Syst.*, 2019, pp. 1–13.

[119] B. Ryskeldiev, T. Igarashi, J. Zhang, Y. Ochiai, M. Cohen, and J. Herder, "Spotility: Crowdsourced telepresence for social and collaborative experiences in mobile mixed reality," in *Proc. Companion ACM Conf. Comput. Supported Cooperative Work Soc. Comput.*, 2018, pp. 373–376.

[120] N. K. Sankaran et al., "Efficacy study on interactive mixed reality (IMR) software with sepsis prevention medical education," in *Proc. IEEE Conf. Virtual Reality 3D User Interfaces*, 2019, pp. 664–670.

[121] P. Sasikumar, L. Gao, H. Bai, and M. Billinghurst, "Wearable remotefusion: A mixed reality remote collaboration system with local eye gaze and remote hand gesture sharing," in *Proc. IEEE Int. Symp. Mixed Augmented Reality Adjunct*, 2019, pp. 393–394.

[122] G. Schall, E. Mendez, and D. Schmalstieg, "Virtual redlining for civil engineering in real environments," in *Proc. IEEE/ACM 7th Int. Symp. Mixed Augmented Reality*, 2008, pp. 95–98.

[123] D. Schmalstieg and T. Hollerer, *Augmented Reality: Principles and Practice*. Reading, MA, USA: Addison-Wesley, 2016.

[124] G. Schall, E. Mendez, and D. Schmalstieg, "Virtual redlining for civil engineering in real environments," in *Proc. IEEE/ACM 7th Int. Symp. Mixed Augmented Reality*, 2008, pp. 95–98.

[125] P. Selonen, P. Belimpasakis, and Y. You, "Developing a restful mixed reality web service platform," in *Proc. 1st Int. Workshop ReSTful Des.*, 2010, pp. 54–61.

[126] M. Sereno, X. Wang, L. Besançon, M. J. McGuffin, and T. Isenberg, "Collaborative work in augmented reality: A survey," *IEEE Trans. Vis. Comput. Graph.*, vol. 28, no. 6, pp. 2530–2549, Jun. 2020.

[127] P. Skinner, J. Ventura, and S. Zollmann, "Indirect augmented reality browser for GIS data," in *Proc. IEEE Int. Symp. Mixed Augmented Reality Adjunct*, 2018, pp. 145–150.

[128] M. Slater, "A. note on presence terminology," *Presence Connect*, vol. 3, no. 3, pp. 1–5, 2003.

[129] M. Speicher, J. Cao, A. Yu, H. Zhang, and M. Nebeling, "360anywhere: Mobile ad-hoc collaboration in any environment using 360 video and augmented reality," *Proc. ACM Hum.-Comput. Interact.*, vol. 2, pp. 1–20 2018.

[130] H. Sun, Z. Zhang, Y. Liu, and H. B. Duh, "Optobridge: Assisting skill acquisition in the remote experimental collaboration," in *Proc. 28th Australian Conf. Comput.- Hum. Interact.*, 2016, pp. 195–199.

[131] A. Tang, C. Owen, F. Biocca, and W. Mou, "Comparative effectiveness of augmented reality in object assembly," in *Proc. SIGCHI Conf. Hum. Factors Comput. Syst.*, 2003, pp. 73–80.

[132] M. Tatzgern, D. Kalkofen, R. Grasset, and D. Schmalstieg, "Hedgehog labeling: View management techniques for external labels in 3D space," in *Proc. IEEE Virtual Reality*, 2014, pp. 27–32.

[133] M. Tatzgern, D. Kalkofen, and D. Schmalstieg, "Dynamic compact visualizations for augmented reality," in *Proc. IEEE Virtual Reality*, 2013, pp. 3–6.

[134] M. Tatzgern, V. Orso, D. Kalkofen, G. Jacucci, L. Gamberini, and D. Schmalstieg, "Adaptive information density for augmented reality displays," in *Proc. IEEE Virtual Reality*, 2016, pp. 83–92.

[135] R. Tenmoku, M. Kanbara, and N. Yokoya, "Annotating user-viewed objects for wearable AR systems," in *Proc. IEEE 4th ACM Int. Symp. Mixed Augmented Reality*, 2005, pp. 192–193.

[136] T. Teo, G. A. Lee, M. Billinghurst, and M. Adcock, "Hand gestures and visual annotation in live 360 panorama-based mixed reality remote collaboration," in *Proc. 30th Australian Conf. Comput.- Hum. Interact.*, 2018, pp. 406–410.

[137] T. Teo, G. A. Lee, M. Billinghurst, and M. Adcock, "Supporting visual annotation cues in a live 360 panorama-based mixed reality remote collaboration," in *Proc. IEEE Conf. Virtual Reality 3D User Interfaces*, 2019, pp. 1187–1188.

[138] B. H. Thomas, "Evaluation of three input techniques for selection and annotation of physical objects through an augmented reality view," in *Proc. IEEE/ACM Int. Symp. Mixed Augmented Reality*, 2006, pp. 33–36.

[139] B. Thoravi Kumaravel, F. Anderson, G. Fitzmaurice, B. Hartmann, and T. Grossman, “Loki: Facilitating remote instruction of physical tasks using bi-directional mixed-reality telepresence,” in *Proc. 32nd Annu. ACM Symp. User Interface Softw. Technol.*, 2019, pp. 161–174.

[140] B. Thoravi Kumaravel, C. Nguyen, S. DiVerdi, and B. Hartmann, “TransceiVR: Bridging asymmetrical communication between VR users and external collaborators,” in *Proc. 33rd Annu. ACM Symp. User Interface Softw. Technol.*, 2020, pp. 182–195.

[141] M. Tomlein and K. Grønbæk, “Augmented reality supported modeling of industrial systems to infer software configuration,” *Proc. ACM Hum.-Comput. Interact.*, vol. 2, pp. 1–17, 2018.

[142] S. Utzig, R. Kaps, S. M. Azeem, and A. Gerndt, “Augmented reality for remote collaboration in aircraft maintenance tasks,” in *Proc. IEEE Aerosp. Conf.*, 2019, pp. 1–10.

[143] B. Volmer et al., “A comparison of predictive spatial augmented reality cues for procedural tasks,” *IEEE Trans. Visual. Comput. Graph.*, vol. 24, no. 11, pp. 2846–2856, Nov. 2018.

[144] P. Wang et al., “Haptic feedback helps me? A VR-SAR remote collaborative system with tangible interaction,” *Int. J. Hum.–Comput. Interact.*, vol. 36, no. 13, pp. 1242–1257, 2020.

[145] P. Wang et al., “AR/MR remote collaboration on physical tasks: A review,” *Robot. Comput.- Integr. Manuf.*, vol. 72, 2021, Art. no. 102071.

[146] P. Wang et al., “Do you know what i mean? an MR-based collaborative platform,” in *Proc. IEEE Int. Symp. Mixed Augmented Reality Adjunct*, 2018, pp. 77–78.

[147] Z. Wang et al., “SHARIdeas: A visual representation of intention sharing between designer and executor supporting ar assembly,” in *Proc. SIGGRAPH Asia Posters*, 2020, pp. 1–2.

[148] Z. Wang et al., “M-AR: A visual representation of manual operation precision in ar assembly,” *Int. J. Hum.–Comput. Interact.*, vol. 37, no. 19, pp. 1799–1814, 2021.

[149] N. Weibel et al., “ARTEMIS: Mixed-reality environment for immersive surgical telementoring,” in *Proc. Extended Abstr. CHI Conf. Hum. Factors Comput. Syst.*, 2020, pp. 1–4.

[150] J. Wither, C. Coffin, J. Ventura, and T. Hollerer, “Fast annotation and modeling with a single-point laser range finder,” in *Proc. IEEE/ACM 7th Int. Symp. Mixed Augmented Reality*, 2008, pp. 65–68.

[151] J. Wither, S. DiVerdi, and T. Hollerer, “Using aerial photographs for improved mobile ar annotation,” in *Proc. IEEE/ACM Int. Symp. Mixed Augmented Reality*, 2006, pp. 159–162.

[152] J. Wither, S. DiVerdi, and T. Höllerer, “Annotation in outdoor augmented reality,” *Comput. Graph.*, vol. 33, no. 6, pp. 679–689, 2009.

[153] W. Wright, D. Schroh, P. Proulx, A. Skaburskis, and B. Cort, “The sandbox for analysis: Concepts and methods,” in *Proc. SIGCHI Conf. Hum. Factors Comput. Syst.*, 2006, pp. 801–810.

[154] T. Xue, A. El Ali, G. Ding, and P. Cesar, “Annotation tool for precise emotion ground truth label acquisition while watching 360° VR videos,” in *Proc. IEEE Int. Conf. Artif. Intell. Virtual Reality*, 2020, pp. 371–372.

[155] T. Xue, A. El Ali, T. Zhang, G. Ding, and P. Cesar, “RCEA-360VR: Real-time, continuous emotion annotation in 360 VR videos for collecting precise viewport-dependent ground truth labels,” in *Proc. CHI Conf. Hum. Factors Comput. Syst.*, 2021, pp. 1–15.

[156] S. Yamada and N. P. Chandrasiri, “Evaluation of hand gesture annotation in remote collaboration using augmented reality,” in *Proc. IEEE Conf. Virtual Reality 3D User Interfaces*, 2018, pp. 727–728.

[157] J. Yang, P. Sasikumar, H. Bai, A. Barde, G. Sörös, and M. Billinghurst, “The effects of spatial auditory and visual cues on mixed reality remote collaboration,” *J. Multimodal User Interfaces*, vol. 14, no. 4, pp. 337–352, 2020.

[158] S. Yonemoto, “A pen based tool for annotating planar objects,” in *Proc. Int. Conf. Virtual, Augmented Mixed Reality*, Springer, 2014, pp. 418–427.

[159] Q. Zhang, X. Zhu, H. Yu, and Y. Jiang, “Enhancing rock painting tour experience with outdoor augmented reality,” in *Proc. IEEE Int. Symp. Mixed Augmented Reality Adjunct*, 2019, pp. 118–121.

[160] Y. Zhang, L. Tao, Y. Lu, and Y. Li, “Design of paper book oriented augmented reality collaborative annotation system for science education,” in *Proc. IEEE Int. Symp. Mixed Augmented Reality Adjunct*, 2019, pp. 417–421.

[161] J. Zhao, M. Glueck, S. Breslav, F. Chevalier, and A. Khan, “Annotation graphs: A graph-based visualization for meta-analysis of data based on user-authored annotations,” *IEEE Trans. Visual. Comput. Graph.*, vol. 23, no. 1, pp. 261–270, Jan. 2017.

[162] J. Zillner, E. Mendez, and D. Wagner, “Augmented reality remote collaboration with dense reconstruction,” in *Proc. IEEE Int. Symp. Mixed Augmented Reality Adjunct*, 2018, pp. 38–39.

Zahra Borhani (Graduate Student Member, IEEE) received the MS degree in computer science from Colorado State University, in 2020. She is working toward the PhD degree with Natural User Interface Lab (NuiLab), Colorado State University. Her research centers on annotation in asynchronous collaborative immersive analytic environments using AR.

Prashast Sharma received the B Tech degree in computer science and communication engineering from the Kalinga Institute of Industrial Technology, India, in 2021. He is working toward the master’s degree with the University of Florida, USA. His work revolves around mixed reality development and video game design.

Francisco R. Ortega received the PhD degree in computer science in the field of human-computer interaction and 3D user interfaces from Florida International University (FIU). He is an assistant professor with Colorado State University (CSU) and director with Natural User Interaction Lab (NUILAB). He also held the position of post-doc and visiting Assistant Professor position with FIU between 2015 and 2018. Broadly speaking, his research has focused on multimodal and unimodal interaction (gesture-centric), augmented reality notifications, information access tradeoffs for visual search, and virtual reality forest bathing.