Using Online Videos as the Basis for Developing Design Guidelines: A Case Study of AR-Based Assembly Instructions

NIU CHEN*, University of British Columbia, Canada and Cornell University, USA
FRANCES JIHAE SIN*, University of British Columbia, Canada
LAURA MARIAH HERMAN, Adobe, USA
CUONG NGUYEN, Adobe Research, USA
IVAN SONG, University of British Columbia, Canada
DONGWOOK YOON, University of British Columbia, Canada

Design guidelines serve as an important conceptual tool to guide designers of interactive applications with well-established principles and heuristics. Consulting domain experts is a common way to develop guidelines. However, experts are often not easily accessible, and their time can be expensive. This problem poses challenges in developing comprehensive and practical guidelines. We propose a new guideline development method that uses *online public videos* as the basis for capturing diverse patterns of design goals and interaction primitives. In a case study focusing on AR-based assembly instructions, we apply our novel Identify-Rationalize pipeline, which distills design patterns from videos featuring AR-based assembly instructions (N=146) into a set of guidelines that cover a wide range of design considerations. The evaluation conducted with 16 AR designers indicated that the pipeline is useful for generating comprehensive guidelines. We conclude by discussing the transferability and practicality of our method.

CCS Concepts: • **General and reference** → *Computing standards, RFCs and guidelines*; • **Human-centered computing** → **Mixed / augmented reality**.

Additional Key Words and Phrases: design guidelines, guideline development, online public videos, methodology, augmented reality, assembly instruction

ACM Reference Format:
Niu Chen, Frances Jihae Sin, Laura Mariah Herman, Cuong Nguyen, Ivan Song, and Dongwook Yoon. 2023. Using Online Videos as the Basis for Developing Design Guidelines: A Case Study of AR-Based Assembly Instructions. *Proc. ACM Hum.-Comput. Interact.* 7, ISS, Article 428 (December 2023), 23 pages. https://doi.org/10.1145/3626464

428

1 INTRODUCTION

Tech companies, standards organizations, and researchers develop and release *design guidelines* when a sector of technology starts to gain traction. When designing experiences for applications in such emerging domains, design guidelines serve as an important conceptual tool that help

*Both authors contributed equally to this research.

Authors' addresses: Niu Chen, jxc5@cornell.edu, University of British Columbia, Vancouver, Canada and Cornell University, Ithaca, USA; Frances Jihae Sin, frances.sin@alumni.ubc.ca, University of British Columbia, Vancouver, Canada; Laura Mariah Herman, lherman@adobe.com, Adobe, San Francisco, USA; Cuong Nguyen, cunguyen@adobe.com, Adobe Research, San Francisco, USA; Ivan Song, ivansong3@alumni.ubc.ca, University of British Columbia, Vancouver, Canada; Dongwook Yoon, yoon@cs.ubc.ca, University of British Columbia, Vancouver, Canada.

2573-0142/2023/12-ART428 $15.00
https://doi.org/10.1145/3626464

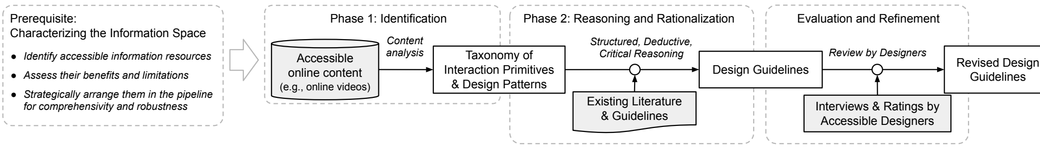

Fig. 1. Our two-phase method pipeline for developing design guidelines in application domains with limited access to experts: the process begins with identifying design patterns by analyzing online public videos, which are an easily accessible but often noisy dataset (Phase 1). Then, it progressively transforms the patterns into *guidelines* by triangulating them with peer-reviewed literature (Phase 2). "Review by Designers" serves as an evaluation of our method.

guide designers by providing well-established principles, heuristics, and strategies. Development of existing design guidelines have been primarily dependent upon two types of resources: expert consultation [17, 41, 60] and the research literature [1, 10, 27, 35, 68]. Nonetheless, consulting domain experts for guidelines can be expensive and, in some instances, may not be feasible. The literature tends to focus on topics that suit researchers' academic agenda, such as a particular theoretical framework or a specific empirical standpoint. The limitations of these resources pose challenges in developing guidelines that are comprehensive and practical. This work introduces a new guideline development method that uses *online public videos* as the basis for a comprehensive representation of design goals and interaction primitives, and assesses the method's strengths and weaknesses. For a visual summary of the entire process, please refer to Figure 1.

Online public videos are an untapped source of data, providing significant analytical opportunities to identify prevailing patterns in interaction design. In HCI, digital ethnography on YouTube is a particularly popular approach to understand how people use emerging technologies [2, 33, 43]. There exist many online videos on emerging technical systems created and posted by indie developers, hardware manufacturers, and research labs. These videos can offer a practical and user-centered perspective, as they allow the guideline developers to observe and analyze real-world-facing interactions and provide a comprehensive view of the diverse patterns of interactions in that specific domain. This approach can promise design guidelines that are effective and usable in real-world scenarios, taking into account the needs and preferences of users. Additionally, online videos provide up-to-date information on the latest advancements, making them relevant and current. However, as a data source, such videos are noisy and unreliable as they are. This is predominately because organizations and individuals create and post such videos with different intentions: promotional, academic, personal, etc. Incorporating online videos in the guideline generation process in a way that can compensate such caveats remains an open question.

To showcase and evaluate our guideline development approach, we apply the method and develop design guidelines for an exemplary domain of AR assembly instruction. This domain is growing in both capabilities and consumer adoption [6]. An oft-referenced application for AR-based assembly includes the provision of instructions for consumer appliance assembly [13, 36], such as replacing or supplementing an IKEA instruction manual, and for industrial maintenance and repair [65] (see Figure 2). Previous empirical studies have repeatedly demonstrated the benefits of AR-based instructions; in-situ spatial guidance is known to help operators complete physical assembly faster and more accurately than traditional 2D manuals by offering lower cognitive and physical overload [8, 28, 62, 63]. Building upon decades of research on AR-based guidance, a few corporate entities have recently been launching new frameworks and platforms [39, 46].

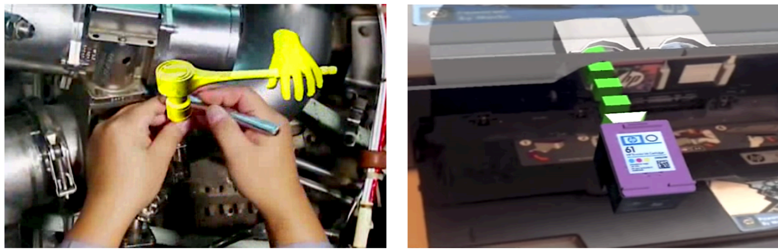

Fig. 2. Exemplary AR-based assembly instructions in an industrial setting [32] (left) and in consumer product assembly [29] (right).

Developing guidelines for such a niche domain poses an *information scarcity* problem, which we define as the shortage of information resources that are conducive to the guideline development process. AR-guided assembly is an emerging technology that is yet to be widely adopted in the consumer market or in industrial settings [16]. While a few companies are using AR training/manuals in industrial applications, access to their expert designers[1] is limited to specific AR platform vendors. Although AR designers have become more accessible in recent years, it is difficult to find those who are experienced in the particular domain of AR assembly instruction.

The literature serves as an alternative information resource for generating guidelines when access to experts is infeasible. Academic literature offers valuable insights based on empirical evidence and theoretical foundations; however, relying solely on it can be limiting due to its restricted perspective and coverage [25], which focuses on specific technical, theoretical, or methodological frames. Furthermore, perspectives of researchers and institutions with existing recognition and resources may be overrepresented [45]. Another limitation of academic literature is the time-consuming publication process, which may not always keep pace with rapidly evolving advancements in AR technology. As a result, guidelines generated from such literature might not encompass the full range of factors, such as user behavior, market trends, cultural influences, and recent advancements. This can lead to an overemphasis on certain aspects of AR experiences while overlooking others, ultimately limiting the scope of the derived guidelines.

When it comes to the specific area of AR-based assembly instruction, the literature currently lacks comprehensive and robust design guidelines. While numerous guidelines are available for various aspects of technological design, including generic AR interfaces [18], 3D [37], and visual UIs and web interfaces [3, 55], these guidelines often do not easily transfer across different technological scopes [14]. For example, referring to empirical studies on generic AR design may lead to recommend bimanual mid-air interactions for resizing virtual objects [9], but this approach may not be appropriate for assembly operators whose hands are typically occupied by assembly objects.

Our work proposes video analysis as a means to create a new guideline development method that addresses the potential downsides of relying solely on literature when access to expert opinions is difficult, rather than replacing them. Designing a robust guideline development *framework* is at the heart of this methodological problem. The concern goes beyond merely incorporating videos as part of information resources. Existing studies on design guideline generation [40, 55] illustrate the caveats of developing guidelines without an established framework: informants can suffer from narrow or skewed interpretations due to biases or preconceptions, identifying which information

[1]It is worth noting here that *developers* are not designers.

resources to appropriate can be challenging, and the generated guidelines may not be practically applicable to the target context of use.

To address these challenges, we designed a novel methodological pipeline (see Figure 1) for guideline development that does not require inputs from the domain experts. Our approach features the following three characteristics: (a) it begins with an exploratory and inductive process of online video analysis for the comprehensive coverage of design variables, (b) it triangulates information from the literature to critically assess each guideline for their validity and relevance, and (c) it is structured in a systematic step-wise process with well-defined sub-goals, each of which can be achieved by using a variety of different methods. We used our methodological pipeline to develop and present the comprehensive and robust guidelines for AR-based assembly instructions.

This paper consequently delivers three primary contributions: (1) a thorough description of our methods pipeline that capitalizes on online public videos to mitigate the challenges posed by scarce information resources when developing guidelines, (2) empirical evaluation results demonstrating the strengths and weaknesses of our method, and (3) a robust set of nine comprehensive design guidelines for AR-based assembly instructions.

2 RELATED WORK

This study builds on previous literature on guideline development methods, online video analysis, and AR-based assembly guidelines. Although there exist several working definitions of AR and Mixed Reality (MR), we follow Milgram's definition of AR for this study [48].

2.1 Approaches to Developing Design Guidelines

Design guidelines are specific and actionable design suggestions that practitioners can adopt for real-world applications. Guidelines have been developed through a variety of methods [55] depending on goals, context, and available resources. There are many different types of guidelines, ranging from universal design/ergonomic principles to concrete and domain-specific design conventions [50]. In this study, our definition of guidelines pertains to "(design) rules" [40, 55] that are generalizable but still bounded within a specific technological domain area.

In industry, guidelines are often regarded as the organization's accrued design knowledge. Corporate entities rarely discuss or disclose how they generated such guidelines [3, 22, 47]. Additionally, guidelines generated from specific companies tend to biased, as they are focused on the company's own technological platforms and typically based upon the personal experiences of a handful of designers. In academic publications, the most common method used to generate guidelines is meta analysis over a set of previous studies. An extensive literature review—typically more than one hundred articles—affords deeper probing and factor prioritization [1, 35, 68]. But this is only possible when the topic area is not niche and a large body of literature exists on the topic. We suggest using online video data as a seed, based upon which the literature can be used to infer broader design patterns. Often, guidelines emerge as a generalization of findings from a series of empirical studies [27] or an instantiation of Grand Theories (e.g. Gestalt theory) in a specific domain [10]. However, these methods tend to generate guidelines that are too specific or too abstract by nature. Some guidelines are generated by synthesizing multiple sources of data [17, 18, 35, 38, 60]. For example, as shown in Lee et al.'s report [38], urban design guidelines can be generated through a compilation of data from expert working groups, a literature review, and a design charrette as well as engagement with the public and private sectors.

Our methodology is designed to ensure both the comprehensiveness and validity of produced guidelines by conducting an exploratory pattern analysis as a basis for a subsequent literature survey and designer reviews. In this way, our methodology is particularly well-suited to overcome the information scarcity problem of emerging technologies, especially the limited access to domain experts.

2.2 Finding Patterns from Online Videos

The design of our online video analysis methodology is informed by several recent HCI studies that have used online videos as a data source for digital ethnography [43]. However, most of these previous studies were focused on identifying users' interaction patterns (i.e. how do people use it?) [2, 7, 33]. Our video analysis methods—which includes search term generating, video sampling, filtering, and thematic analysis—are geared toward identifying *design patterns* (i.e., what do designers do and care about?). Additionally, to compensate for the limitations of online videos as a data source [21], we supplement our video analysis with a literature survey.

2.3 Design Guidelines for AR-based Assembly Instruction

Assembly instruction refers to an artifact or medium describing how to put target objects together. A key challenge in conveying assembly instruction is the cognitive effort of operators in mapping the instructions into physical actions [12]. This challenge has motivated the development of AR-based assembly systems that can overlay instructions and guide users toward completing tasks directly in the user environment. In fact, in their seminal work, Caudell and Mizell coined the term "Augmented Reality" to describe a system developed to deliver instructions to aid physical tasks in a Boeing factory [8]. From then on, many studies have been conducted to demonstrate the benefits of AR-based assembly instruction. Studies conducted by Tang et al. [63] and Roberson et al. [52] have demonstrated the performance advantages of AR-based instruction as compared with printed documents or LCD screens. Subsequent work by Henderson et al. [28] again confirmed that AR was faster and more accurate than a stationary display in helping participants complete repair tasks.

Although the benefits of AR-based assembly instruction are clearly justified, these systems and studies do not provide design guidelines. The development of guidelines for new technologies has been an active area of research within and around the HCI field, as detailed in next subsection. However, design guidelines for AR-based assembly instruction are still under-explored. A number of guidelines for AR-based assembly have been proposed outside of the HCI field, such as manufacturing science [11] and human factors [20]. These studies, however, recommended guidelines that are more applicable to their own fields, rather than to designers and application developers. The set of guidelines developed by Rolim et al. [53] is most closely related to the present study. These guidelines were derived only through literature analysis and focused primarily on body movement instruction. The present work proposes design guidelines that are both more comprehensive and more robustly evaluated. Our design guidelines were developed through a novel pipeline that synthesizes relevant information from multiple data points.

3 METHODS

In this section, we illustrate the principles, prerequisites, and specifics of our guideline generation methods.

3.1 Guiding Principles

Through iterative discussions between our research team members, we came up with three guiding principles for the design of our methods pipeline.

Comprehensiveness: For technologies in an emerging domain, such as AR-based instruction, there is a lack of widely accepted standard design features. Without knowing the constituent elements of the application, it is difficult to prescribe how it should be designed. To ensure that our guidelines cover a comprehensive set of possible design dimensions and variables of the target application domain, our pipeline begins with a data-driven exploration stage, taking online public videos as accessible large-scale resources from which we generate a taxonomy of interaction primitives and design patterns.

Usefulness: We aim to generate guidelines that designers will find helpful in enhancing user experience of their systems. To this end, we prioritized guidelines that are feasible, meaning that they are easy to understand and apply in practice, and have clear benefits to users. In section 4.1, we interview AR designers to evaluate the generated guidelines against different aspects of this usefulness criteria: feasibility, utility, etc.

Validity: The problem of using videos as a data source is that it reveals "common" design practices, which are not necessarily "good" or "effective". In the second stage of our pipeline, we triangulate the videos with the research literature to progressively validate the patterns identified in the first stage and refine them into into a set of prescriptive guidelines.

The following sections document a case study of using our pipeline to develop design guidelines for AR-based assembly instruction. The goal is to generate a set of guidelines at the intended level of generalizability within the target domain (i.e. not being too specific nor abstract). Following the pipeline maximizes the benefits of the available resources, while simultaneously compensating for their limitations.

3.2 Prerequisite: Characterizing the Information Space

The first step of guideline generation is to identify accessible information resources and critically examine their limitations. For our target domain, we identified three data sources: online videos, literature, and designers.

Online Videos: Videos of AR-based assembly can capture a wide spectrum of design practices used by corporations, researchers, and independent developers. Videos excel at demonstrating the visual and experiential nature of the technology and are highly accessible via online search engines. However, online videos are a noisy data source; videos are created to serve various intentions, such as promoting a particular system, and may contain extraneous information that is misleading or irrelevant. Our approach puts this video analysis process into a systematic structure.

Literature: The HCI and AR literature offers substantive empirical evidence and theories. Yet, academic research is driven by the motivation to generate novel and generalizable knowledge, rather than enhancing users' experience [61]. Thus, there are mismatches between the (level of) concerns in academic studies and the end-users of the guidelines—designers in our target domain. Our approach compensates for this shortcoming of literature by running the video analysis first to explore the wide range of design considerations before narrowing down the focal points.

Domain Experts: In general, experienced designers are scarce resource, and involving them in guideline generation is costly. Although gaining access AR interaction designers are becoming increasingly easier as the whole AR application domain expands, most of accessible AR designers are experienced in designing applications for entertainment, *not for assembly instruction*. Hence, using perspectives of these generic AR designers as a starting point is unlikely to result in guidelines that *comprehensively* capture design considerations specific to the target domain of assembly instruction. Our method does not require expert inputs and is thus robust to the scarcity of domain experts as information resources.

3.3 Phase 1: Identifying Design Patterns (Analyzing Online Videos)

To identify commonly-followed design patterns, we analyzed online videos in the target domain by: (a) creating a diverse, information-rich video dataset (N = 146) by composing search queries, (b) generating a taxonomy of interaction primitives (Figure 3, first column), and (c) identifying

Table 1. The final search terms organized into three categories. These sets yield 749 unique combinatorial search queries.

Tech-Descriptor Search Terms *(N = 17)*
AR or Augmented Reality, VR or Virtual Reality, MR or Mixed Reality, Virtual, Interactive, Contextual, Workspace, Mediated Reality, Intelligent, Extended Reality, Diminished Reality, HoloLens or Lens, Magic Leap, Google Glass, ARKit, ARCore, Vuforia
Purpose-Descriptor Search Terms *(N = 10)*
Directions or Directional, Set-Up, Assembly, Maintenance or Repair, Manufacturing, DIY, Construction, Assistance, Training, Step-by-step or manual or tutorial or guide or Instruction or Instructional
Media-Descriptor Search Terms *(N = 7)*
Video, Manual, System or Experience or Application or App, Directions or Explanation, Class or Teaching, Assistant, Prototype or Demo

design patterns (Figure 3 second column) by conducting a co-occurrence analysis over the videos annotated based on the taxonomy.

3.3.1 Sampling Online Videos. We executed a breadth-first, query-based video selection method as a variant of its kind from previous studies covered in Related Work. We found videos from two popular video-sharing platforms: YouTube and Vimeo.

Composing Search Queries: To ensure that our samples capture patterns from a diverse selection of applications, domains, and hardware types, we constructed search queries as compositions of topical keywords from three orthogonal categories: Technology, Purpose, and Media-descriptor search terms. Each category represents an aspect of AR-based assembly instructions: Technology terms capture the platform for which the experience was designed; Purpose terms describe the primary goal of the experience; Media terms refer to the resulting experience channeled through a specific mode of interactions. Each search query is comprised of one term from each category (e.g., "AR" "Assembly" "Assistance"). Prior to starting the video collection process we pruned out search queries that did not return a single video on AR-based assembly within the first 150 results. The final 34 terms used for video collection are presented in Table 1.

Collecting Target Videos: For each search query, two lead investigators inspected the first 20 video results (or all if fewer than 20) and selected items based on two inclusion criteria: (1) the video clearly depicts an assembly task or an assembly/repair training procedure and (2) the instructions involve in-situ AR overlays. To ensure the comprehensive coverage of different types of visual overlays, we used a broad definition of AR, following Milgram's seven classes of MR displays [48], for the second criteria. After this step, we collected 146 videos (the supplementary material: 1.videos.xlsx).

3.3.2 Generating a Taxonomy of Interaction Primitives. In our methodological framework, *interaction primitives* refers to common constituents of user interactions in the target application domain. We established a taxonomy of interaction primitives specific to AR-based assembly to characterize and discover the repeating elements of interaction design (for the full taxonomy, see the supplementary material, 2.taxonomy.pdf). We identify primitives from videos and structured them into a taxonomy through open-coding, followed by axial coding [59]. Since not all 146 videos were

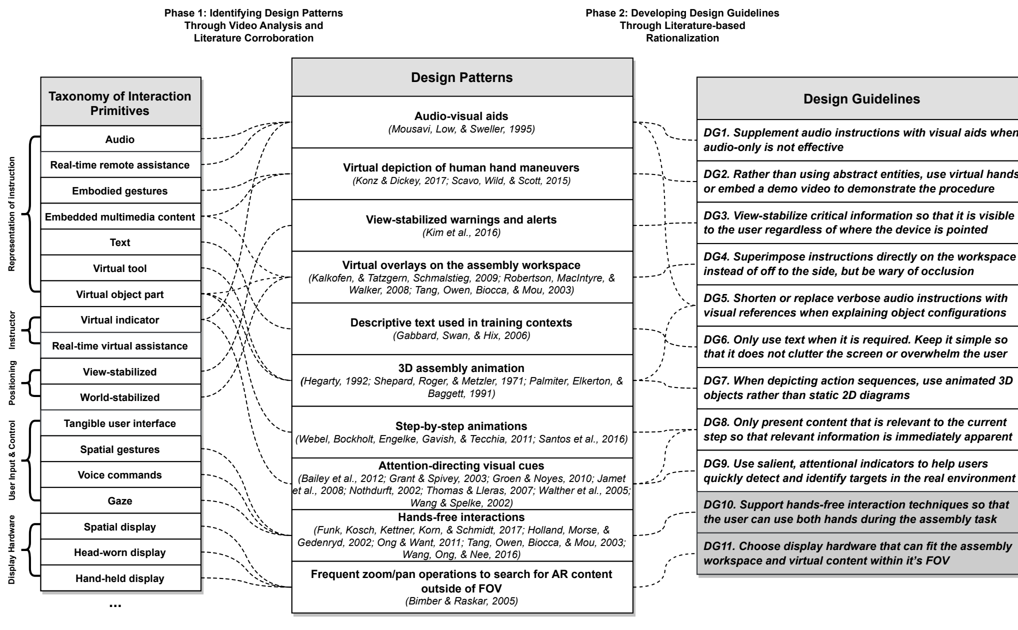

Fig. 3. Relationship between step-wise outputs in our guideline development pipeline. The presented design patterns are a subset of many others that eventually evolved into design guidelines. Guidelines 10 and 11 have been greyed out, as we chose to omit them following evaluation. Please refer to section 4.2 for the final design guidelines and their corresponding rationales.

"information-rich", we prioritized analyzing videos that showcased a variety of instructional design elements (i.e. interaction primitives). Before coding the videos, two investigators watched and sorted the videos in the order of prospective richness. The coding process began with the most information-rich video and stopped at the 50th video, when we reached the data saturation of found primitives.

After generating the taxonomy, the two investigators watched the remaining 96 videos with a focus on finding primitives that were not already included the taxonomy, but no new primitives were found. 36% of the videos were coded collaboratively and discussed by two investigators to ensure that there were no major discrepancies in the coding scheme or conflicting interpretations. The remaining two-thirds of the videos were coded independently 50/50. All ambiguities were resolved through discussion with the entire team. In addition, to ensure the reliability of our taxonomy, we recruited an independent coder, trained them by using the taxonomy as a codebook, and conducted an inter-coder reliability testing, which yielded Cohen's kappa score of 0.795, indicating strong agreement between the coder and the investigators.

3.3.3 Extracting Design Patterns. To identify prevalent patterns of design in the target domain, we coded each video using our interaction primitive taxonomy and examined the frequency of each primitive to identify. We also calculated the co-occurrence rates between different primitives to determine if there were any pairs or triplets that occurred frequently together. Prevalence or frequency, however, is not necessarily sufficient to constitute an insightful design pattern. We wanted to understand *why* these primitives were being used so often and to achieve *what* result/goal. Thus, the two lead investigators analyzed the video dataset, whilst asking questions such as: "*Why* are these [interaction primitives] being used frequently?" and "*Why* are people using these [interaction primitives] for this [specific task type/context]?" After deeper analysis, we hypothesized possible rationales underlying the patterns and searched the literature to find theoretical scaffolding and empirical support. We corroborated each design pattern with at least one published study (and in most cases, more than one). As shown in Figure 3, our design patterns comprehensively cover our interaction taxonomy. It's worth nothing that, at this stage, the objective is to find a set of common design practices, rather than generating prescriptive suggestions (i.e. how it should be designed).

3.4 Phase 2: Developing Guidelines (Triangulating Video Data, Design Patterns, and Literature)

The purpose of this phase is to rationalize the aforementioned design patterns and transform them into prescriptive and succinct design recommendations, which we refer to as Design Guidelines (DGs, listed on the right side of Figure 3).

Eliciting guidelines from design patterns requires bridging the gap between current practices and the ideal. This necessitates a critical value judgment of patterns discovered in Phase 1, evaluating whether a design practice is "right" or "wrong." Approaching this as a specific and well-defined pseudo-research question, we conducted a systematic literature review by following the process illustrated in [70]. We delved into the literature to find relevant studies that could provide rationales for these design practices or offer persuasive arguments against them. This process comprises approximately three steps: (1) searching the literature with descriptive keywords about the design pattern and broader technical keywords, such as "AR"; (2) assessing the quality of the search results and determining inclusion based on relevance, rigor of methods used, and credibility of the publication; (3) analyzing and synthesizing the data to make a judgment call, identify contextual factors, and encode the descriptions of patterns and context into a prescriptive statement of the design guideline. This approach allowed us to supplement the design patterns with a more reliable data source: theories and empirical findings from existing studies in research domains such as AR/HCI, cognitive psychology, and engineering.

In addition, to refine the DGs, we re-analyzed videos that contained interaction primitives that pertained to a DG, paying close attention to nuanced patterns and comparing them to recommendations from previous research. For example, we measured the temporal co-occurences of certain design patterns, which provided insights into which the design patterns might be united to form a single coherent guideline. Additionally, this process highlighted the gaps between researchers and creator-practitioners in this field, thereby helping us identify the feasibility and utility of the proposed design guidelines. In some cases, the videos were artifacts of research projects and may therefore be supported by the same researchers that produced the videos.

We discussed the DGs amongst the five investigators and evaluated the guidelines based on feasibility, utility, and relevance. We excluded extraneous guidelines that were irrelevant to AR assembly instructions, too specific, or too general. Finally, we corroborated each design pattern with research to create a set of 11 DGs (see Figure 3, the third column).

4 EVALUATING AND REFINING DESIGN GUIDELINES (DESIGNER REVIEW)

We evaluated the generated DGs counting on AR designers' perspectives and then refined them into a set of Revised Design Guidelines (RDGs). The objective is twofold: (1) to validate if our method can generate comprehensive and robust guidelines based on the video analysis and literature review, and (2) to understand how the inputs from designers, if available, can be used to refine the generated DGs. Formally, this evaluation phase is not part of our proposed method. But, optionally, for the

Table 2. Interview study participant demographics; YoE = Years of experience with AR. Project = Number of projects worked on.

ID	Age Range	Gender	Occupation	YoE	AR projects	VR projects	MR projects
P1	40-49	Man	University Associate Vice-President	3	2-4	2-4	1
P2	50-59	Man	Professor	19	2-4	10+	2-4
P3	30-39	Man	Post-doctoral Researcher	7	10+	10+	10+
P4	30-39	Man	Research Lead	3	2-4	5-10	2-4
P5	40-49	Man	Immersive Media Artist	5	5-10	2-4	0
P6	25-29	Man	Senior Program Manager	5	10+	10+	10+
P7	25-29	Woman	Product Designer	2	2-4	2-4	2-4
P8	40-49	Man	Interactive Producer, Professor	2	5-10	0	5-10
P9	25-29	Woman	Senior Experience Researcher	4	10+	10+	10+
P10	N/A	N/A	Post-doctoral Researcher	2	2-4	2-4	2-4
P11	40-49	Man	Experience Designer	6	5-10	2-4	1
P12	40-49	Woman	Senior User Experience Researcher	10	10+	10+	10+
P13	40-49	Woman	Professor	3	2-4	2-4	1
P14	40-49	Man	Senior Experience Designer	8	5-10	0	5-10
P15	40-49	Woman	Senior Design Researcher	9	2-4	2-4	5-10
P16	40-49	Man	Senior Experience Developer	2	1	2-4	0

domain where the expert consultation is available, the pipeline can be extended to include the designer review as the third phase.

4.1 Methods and Results

For our evaluation, we had 16 AR designers to examine the 11 DGs generated in Phase 2. In essence, the results suggested that using videos as the foundation for broadening the design considerations enabled us to generate comprehensive and robust guidelines.

4.1.1 Recruiting AR Designers. Our definition of the term "AR designer" includes industry specialists and academic researchers with at least two years of experience in designing AR systems and contributions to at least two AR projects[2]. We recruited a total of 16 participants (see Table 2). Participants were required to complete a screening questionnaire to verify their eligibility. All of them had considerable contributions to AR/MR applications and domain expertise. Among them, three participants had experience in our target domain of AR-based assembly. The rest were experienced in other AR-related domains, such as educational AR and immersive data visualization. Six participants were industry professionals and ten were from academia. Occupations included the following: AR UX/UI designer, program manager, industry researcher, university professor, and post-doc researcher. Ten self-identified as man, five self-identified as woman, and one preferred not to state their gender. Participants' ages ranged from 25 to 59 years (Median = 44.5, SD = 9.2).

4.1.2 Apparatus: AR Gallery. Many guidelines, particularly those published in academia, are typically presented as plain text. However, the dynamic and visual nature of AR creates specific interaction paradigms such as the positioning of virtual content (i.e. view-stabilized vs. world-stabilized)

[2]Participant P16 marginally meets the inclusion criteria based on our definition, as they have worked on only one AR project and served as a developer. However, we chose to include their data due to the relevance of their experience in the AR project, the depth of their collaboration with designers, and the quality of the evaluations they contributed during the review session.

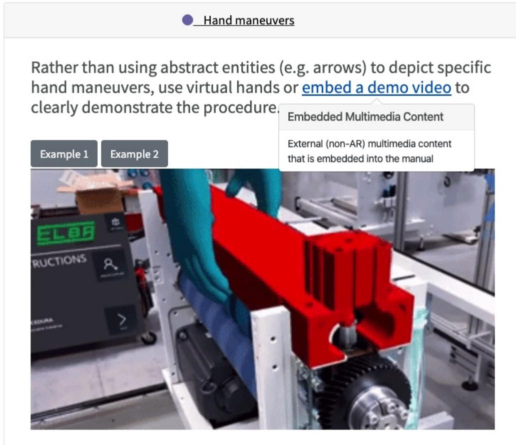

Fig. 4. A screenshot of the AR Guideline Gallery. Each guideline was supplemented with up to two video examples. Domain-specific terminology was clarified with popovers (e.g., "Embedded Multimedia Content").

that can be hard to conceptualize with textual descriptions, even when they are supplemented with images. Thus, we built *AR Gallery*, a video gallery web application, to present our guidelines. In the gallery, guidelines are presented along with video clips from our video dataset that showcase practical examples of how the guideline can be applied in an assembly setting. This helps viewers understand the guideline statement better. To ensure impartiality and provide a comprehensive understanding, we have included multiple video examples that represent various types of applications and use contexts relevant to the respective guideline. For guidelines that contain domain-specific terminology (e.g. superimposition), we embedded definitions as popovers (see Figure 4).

4.1.3 Procedure. The study sessions consisted of (1) a short introductory interview on general design background and AR experience, (2) a think-aloud exploration of the DGs in the AR Gallery while completing a structured questionnaire, and (3) a longer semi-structured interview discussing the proposed guidelines.

For the introductory interview, we asked our participants to share their AR design background. This included details about their design process, the projects they contributed to in the past, and any guidelines that they follow in their current practice. These questions later served to situate the participants' responses to the guidelines in the context of their personal experiences and practices.

The primary task of the study session was to examine and rate the DGs with regards to their feasibility, utility, novelty, and agreement. The guidelines were presented through the aforementioned AR Gallery and the order was balanced for each participant to counter primacy and recency effects. As participants examined each guideline along with example videos, we instructed them to *think-aloud* [58] while acquainting themselves to the design guideline at hand and also while rating it. Therefore, we would hear the participants' thoughts as they first read and interpreted the guidelines and also would glean the decision-making process of their rating choices.

Finally, we conducted a semi-structured interview to probe retrospectively about participants' thoughts on each DG. We asked participants to state the most and least useful guidelines ("Which guideline do you consider most/least valuable and why?"), potential improvements to the guidelines ("How can we improve the overall set of guidelines?"), and to explain any implications that the

Guideline	DG1	DG2	DG3	DG4	DG5	DG6	DG7	DG8	DG9	DG10	DG11
Utility	4.56	4.06	4.19	4.31	4.31	4.25	4.31	4.25	4.56	4.31	3.94
Feasibility	3.69	3.81	4.31	3.62	3.69	4.12	4.00	3.69	3.69	3.25	3.12
Agreement	4.44	4.19	4.06	4.31	4.19	4.25	4.31	4.25	4.62	4.12	3.81
Novelty	2.56	3.12	2.56	2.81	2.81	2.37	2.37	2.75	2.00	2.31	2.81
Average	3.81	3.80	3.78	3.76	3.75	3.75	3.75	3.74	3.72	3.50	3.42

Fig. 5. Designers' ratings (mean) for each of our design guidelines. We reflect on the low novelty scores in the Discussion section.

guidelines may have for their own work ("How do you see yourself applying these guidelines to your current design project?").

4.1.4 Analysis. First, we transcribed all audio recorded think-aloud and interview data. Then, we highlighted relevant quotes for each DG and generated a list of thematic categories using open coding, followed by axial and selective coding [59]. We grouped the categories under high-order headings and noted any trends that found across participants.

Next, we computed the average and standard deviation for each Likert-type questions in the questionnaire (Figure 5); the answers ranged from "strongly disagree" (1) to "strongly agree" (5). Additionally, we investigated any significant differences between individual participants, such that we could contextualize their responses accordingly. For instance, one participant consistently rated all guidelines as low in "feasibility" due to their lack of technical expertise.

Overall, our analytic process was primarily guided by the deductive reasoning based on our guideline evaluation criteria, such as comprehensiveness, utility, and feasibility. During the analysis period the research team members closely discussed the codes to resolve disagreements on interpretation of the coded data.

4.1.5 Results. In this section, we present the results of running our pipeline and the ratings on our design guidelines. To view the designers' detailed comments and how they were integrated into our guideline development process, please refer to the "Rationale and Designer Input" of each guideline in section 4.2.

Comprehensiveness: As we aimed to cover comprehensive design considerations related to AR-based assembly instruction, we used the video data to generate a taxonomy of interaction primitives with 28 primitives categorized into seven design dimensions. In Figure 3, the connecting lines between the Taxonomy items and Design Patterns illustrate the broad coverage of our guidelines with respect to domain-specific design issues. To further ensure comprehensiveness, we *explicitly* asked designers during the interview study whether there were any aspects of AR experience missing in our RDGs. Their collective responses did not indicate any significant omissions.

Usefulness and Validity: Overall, designers rated our guidelines favourably, especially in regards to utility (M = 4.28, SD = 0.18) and overall agreement (M = 4.23, SD = 0.21). This is a promising finding that indicates the potential benefits of the guidelines for improving user experience. For our final set of guidelines, we elected to exclude DGs 10 and 11. These two guidelines received lower feasibility scores compared to the others because they prescribe the use of specific hardware. Our experts pointed out that hardware choices are often made prior to UI design and are beyond their control. For example, one participant stated, *"Actually, we can't choose the device"* (P9), and another stated, *"[Y]our client has already enforced the device constraint"* (P16). For the full ratings, see Figure 5.

Dropping DGs 10 and 11: The expert feedback allowed us to pinpoint the feasibility problems associated with DGs 10 and 11. Several participants highlighted that these guidelines recommend specific hardware, such as platform or display types. However, in reality, hardware decisions are typically made early in the product development process or are simply assigned to the designers. Consequently, participants expressed concerns about applying these guidelines in their design practices, where they lack control over hardware choices. As P1 aptly noted, "it's not easy to implement in practice because it's actually a hardware challenge and not a software challenge." Given the relatively low feasibility ratings and well-rationalized designer comments, we decided to remove DGs 10 and 11. The remaining nine guidelines were retained, as their evaluations from the designers were consistently positive.

We present the final guidelines in the next section (4.2) and discuss the implications of these evaluation results in 5.3.

4.2 Revised Design Guidelines

By triangulating the designer inputs with the design patterns from our video analysis and existing recommendations from the literature, we could revise the nine DGs. We have included a "Rationale and Designer Input" section for each guideline that elaborates on the evolving considerations at different phases of guideline development, including prevalent patterns identified in video analysis (Phase 1), relevant prior studies (Phase 2), and designers' commentary (Evaluation). The guidelines are sorted by descending average rating. To help the readers understand the guidelines, we accompany them with visual examples through the AR Gallery (see the supplementary material: 4.ar-guideline-gallery.zip).

RDG 1. Supplement audio instructions with visual aids: Audio instructions are easy to implement and can be useful in AR manuals for a variety of reasons. However, when deciding to use audio it is crucial to consider factors such as environmental conditions, such as noise in an industrial setting. Whenever audio is used, we recommend adding visual aids as a supplement.

Rationale and Designer Input: Audio is a low-cost, easy-to-implement modality that most AR systems can support. However, in assembly contexts, there are several cases where audio-only instruction may be insufficient. For instance, in industrial settings, workers often carry out assembly procedures in noisy, high-pressure workspaces. Such environments can make it difficult to hear or follow audio instruction. Despite recommendations from the literature that advise against incorporating redundant modalities [44], we advocate for the inclusion of supplementary visual guidance (e.g. text subtitles for instructional narration) whenever audio instructions are used.

Overall, this guideline received the highest average rating from the designers amongst all other guidelines. Their remarks stress the significance, fairness, and practicality of this rule in actual design work. The design pattern identified in our video analysis was in agreement with this guideline. In total, 12 out of 50 videos analyzed (24%) contained some form of auditory instruction. 11/12 (92%) of those videos supplemented the audio with visual aids, such as closed captioning or 3D animations.

RDG 2. Demonstrate hand maneuvers using virtual hands: Represent hand maneuvers with first-person embodied gestures so that users can “fit their hands” in the virtual hands and reproduce their configuration and movement. This method is beneficial for kinesthetic learning and ergonomic safety in tasks that involve manipulating hazardous assembly components or that require excessive physical forces like push, pull, and lift. To successfully apply this guideline, designers should consider the end-users' varying physical capabilities (e.g., strength, hand size).

Rationale and Designer Input: When designers use abstract symbols, such as arrows and words, to demonstrate hand maneuvers, users must translate these symbols into mental 3D images and then match their hands with that mental image [34]. In total, 25 videos (50%) contained at least one instance of embodied gestures (i.e. a representation of a human or part of a human with a virtual model). Often, these were virtual hands superimposed on the real assembly object from a first-person perspective. This guideline received the highest novelty rating out of all the guidelines.

Our designers evaluated this guideline positively when they perceived an explicit need for specific hand placement, finger positioning, or manual action: *"If your fingers have to be in a certain position, for example, I think it would be a useful design guideline"* (P1). However, P6 commented that designers should keep in mind that there may be multiple ways of completing the same task. Different users may *"use their bodies in very different ways and unexpected ways [...] [they] often don't follow things the way you expect them to"* (P6).

RDG 3. Ensure critical information about the workspace is always in view: In AR assembly, the user is often required to change their positioning and gaze. As a result, virtual content may be temporarily "lost" if they are fixed to a real location. To overcome the challenge, critical information should be fixed to the screen or made tag-along (i.e. moves with user's gaze).

Rationale and Designer Input: We examined all the videos with text instructions: the majority (73%) fixed the text to the screen. In terms of implementation, designers agreed that it was *"reasonably easy to use"* (P2). However, several were wary of having view-stabilized content because it can block the line of sight to task objects and environments. P3 and P14's comments were in agreement with P15's point: *"having something constantly in [the user's] eye is annoying"*. Designers must carefully prioritize content as view-stabilized messages can clutter the screen, reduce the FOV, and (ironically) occlude the intended hazard or assembly component.

RDG 4. Superimpose instructions directly on the workspace instead of off to the side, but be wary of occlusion: Superimposing digital instructions onto real objects clarifies the assembly process and minimizes errors such as misalignment, skipping over a required action, and picking the wrong tool or component. For instance, target elements can be outlined with high-contrast colours to help users quickly locate them. For tasks involving multiple steps, designers can animate superimposed virtual models to demonstrate the correct way of performing a task. Although useful, superimposition can occlude the view of real-world elements, particularly potential hazards, with virtual content. A potential work-around is to use semi-transparent representations, such as "ghostings" [31].

Rationale and Designer Input: Several empirical studies have shown that AR systems can improve user task performance during assembly tasks by integrating virtual information into the real environment and anchoring relevant cues to physical objects [52, 63]. In accordance with the literature, many AR systems in our video dataset superimposed virtual content on real-world elements, rather than "off-to-the-side" (25/60, 52% vs. 8/50, 16%).

Many designers (P4, P5, P6, P7, P12, P13) explicitly stated that direct superimposition, rather than off-to-the-side positioning, is useful for enhancing user experience, because direct visual overlays enhance *contextual information*, which can increase spatial comprehension of the action to perform. However, despite the benefits of superimposition, P9, P12, P14, P15, and P16 emphasized that designers must be careful of occluding real-world information. Occlusion, according to P2, is the *"biggest problem"* with superimposition. In our video analysis, we observed several strategies to avoid occlusion. For instance, the use of translucent CAD model materials and the removing of virtual content from the display after a fixed amount of time.

RDG 5. Use visual instructions to promote spatial understanding rather than text or audio: Text and audio are common modalities that are used in AR-based instructions. However, for assembly instructions that involve complex spatial relationships and sophisticated motions, textual and auditory descriptions can be verbose, confusing, and ineffective. Instead, we recommend using visual representations for instructing highly spatial assembly tasks. Potential solutions include 3D diagrams and step-by-step animations.

Rationale and Designer Input: Existing studies on multimedia learning suggest to use visual representations for instructions involving spatial manipulations [15, 69]. Most designers agreed that visual references, rather than other modalities, are crucial for explaining complex object configurations and motions because visuals facilitate the understanding of spatial relationships. Due to the inherent complexity of assembly tasks and the novelty of AR interaction paradigms, P13 explained that the cognitive load of following instructions in AR is *"quite high, especially as we're not used to it yet"*. Thus, designers should aim to *"go visually whenever possible"* (P5) because *"visual aids are always more effective than the alternative"* (P14).

Interestingly, the design patterns from our video analysis go against the suggestions from the literature and designers. In our video dataset, audio and text were common modalities used to explain and describe object configurations, spatial relationships, and complex motions. Textual descriptions were found in 17/50 videos (34%) and auditory descriptions were found in 12/50 videos (24%). This can be attributed to the relatively high cost and effort required in creating or obtaining high-quality visual content (e.g. detailed 3D models), contrary to text and audio instruction, which can be sourced from an existing paper manual or easily generated,

RDG 6. Only use text instructions when language is a key component of the assembly task: For the majority of assembly tasks, designers should carefully evaluate whether text is effective for the task. The substitution of lengthy text for structural and action diagrams is already regarded as a best practice for 2D instructions, but in AR, text is even less effective due to legibility concerns. However, there are task contexts where language is a key component of the assembly task (e.g., training for a specific assembly, common terminology between coworkers, etc.). In such cases, *selective* use of text-based assets, such as instructions and labels, may leverage positive transfer from the AR experience to their work practices for enhanced learning.

Rationale and Designer Input: Legibility concerns of text in AR due to "uncontrollable environmental conditions" [19] is supported by a rich body of research. To designers, this guideline appeared obvious: *"Everyone who works in the AR field knows not to use so much text"* (P13), *"If they wanted to read [text], they could read the pamphlet* (P3). Contrary to designers' opinions, 41 videos (82%) contained text-based assets, primarily in the form of lengthy instructions. Although in most assembly contexts, designers should keep text minimal, *"there are times when you want more verbose text [...] to aid the assembler in understanding more deeply"* (P16).

RDG 7. Use 3D animations to depict action sequences: AR manual designers should strive to use 3D animations to demonstrate actions. Animations should be repeatable and viewable from various angles. Further, they must convey the necessary direction and motion of the task at hand. For example, an instruction such as "rotate the screwdriver to turn the screw counter-clockwise three times" can be best explained visually through a 3D animation.

Rationale and Designer Input: Research shows that mentally transforming 2D diagrams into dynamic 3D information imposes a high cognitive load on users [26, 57]. Animations, on the other hand, allow users to rehearse their future actions while viewing the demonstration, which in turn, reduces cognitive load [51]. To designers, 3D animations were viewed as *"the whole point [of AR]"* (P5). They believed that animation could better engage users' attention (P3, P7) and improve learning outcomes for both simple (P2) and complex (P3, P10) tasks. In our video analysis, actions, such as joining components or lifting equipment, were typically prescribed using a combination of text (30/50, 60%) and animated virtual objects (36/50, 72%), indicating that this guideline is implicitly followed by many content creators.

RDG 8. Only show virtual content that is relevant to the current step of the assembly process: Like most paper manuals, AR assembly instructions should provide users with manageable step-by-step instructions. To reduce the complexity of lengthy procedures and to reduce the search time for instructions, designers must ensure that all virtual guidance materials rendered on the display are relevant to the current step being performed. The ideal interface would display requisite virtual instructions and subsequently update them (e.g. remove virtual models that are no longer needed) once the user has completed the step.

Rationale and Designer Input: Previous studies have shown that cluttered displays can hamper the cognitive processes of selecting and organizing information [54]. Thus, designers should strive to remove or reduce irrelevant virtual content from the user's field of view. An exemplary example would be the instructional AR assembly system developed by Henderson and Feiner, which used a finite state machine to manage the visibility of content in the display [28]. Renderings were cleared and subsequently updated with new, relevant virtual guidance materials when a step-transition was cued.

As observed in 48 (96%) videos, designers agreed that step-by-step management of virtual content is crucial: *"As you're going through the steps, [...] you don't want to see all the parts because that's just overwhelming"* (P16), *"Because of the restricted field-of-view, you need to focus on the relevant parts only"* (P13).

RDG 9. Use salient attention-directing visuals to help users quickly detect and identify targets: AR-based assembly instructions should use highly salient visual indicators to direct users' attention to the right AR elements at the right time. Salient visual cues can help users rapidly and easily find relevant information, particularly when the environment is cluttered or when the task is complex. They can be used to reference specific objects, points of interest, and potential hazards.

Rationale and Designer Input: Using salient cues will produce faster and more accurate target identification, which will reduce search time and minimize identification errors [49]. Furthermore, actively guiding users' attention to relevant information improves problem solving [23, 24, 30] and can enhance spatial learning by improving memory recall of object properties such as size, shape, and even location [4, 64, 66, 67]. In our videos, we observed 31 instances of highly salient cues that we categorized into 7 types (colour highlights/overlay, bounding box/sphere, arrow, temporal modulation, superimposed text label, pointing, magnifier effect). Among all guidelines, this guideline received the highest average "usefulness" rating for enhancing user experience (4.56/5, SD: 0.61). Many designers had the same perspective as P12, who said *"contrasting colours is important and targeting is very important"*. However, several participants (P2, P9, P10) listed dynamic environmental conditions as a challenging design variable: *"you need to know what the environment looks like to figure out what the contrast is going to be"* (P9).

5 DISCUSSION

5.1 The generalizability and transferability of the proposed method pipeline

Our Identify-Rationalize pipeline has been structured in a systematic, step-by-step process to make it work as a general framework for guideline development. This structure is transferable to other technical domains for guideline development, but the specific implementation of each step should be adapted to the factors specific to the target domain.

The first factor to consider is resource availability in the target domain. For example, online videos were a great resource for AR-based assembly instructions as they were accessible, useful, and plentiful enough to provide insights into dynamic spatial experiences in such systems. However, in other domains, such as haptic interfaces, video data may not be sufficient and a more suitable resource, such as a user-curated wiki (e.g. Haptipedia [56]), may be used instead.

The second factor to consider is the domain's expectations for the guidelines. Different domains may value different strengths in their guidelines, such as rigor, comprehensiveness, and up-to-date information, due to differing community cultures or industry standards. The pipeline's execution must take these beneficial characteristics into account, which may result in a different emphasis on data sources and phases in each application. For instance, if the pipeline is applied to health technologies, the evaluation phase may require thorough examination by domain experts, despite the cost.

Guideline developers should also consider other factors, such as user behavior, policies and regulations, and technical limitations, when transferring the suggested method to a different domain.

5.2 Limitations of using online videos as a data source

While online public videos are a valuable resource, they are not without their limitations. When viewing and analyzing videos, one cannot extract information that lies "beneath the surface." For example, text in AR is a common pattern that contradicts recommendations from literature and designers. Is text common because it's easy to implement? Or are some creators unaware of common guidelines? The mere prevalence of a pattern does not guarantee its quality. Therefore, the triangulation and rationalization processes in Phases 2 and 3 of our pipeline were critical in converting the identified patterns into the respective guidelines.

Finding videos with specific properties and extracting desired information from them can be slow and taxing [42]. One way to improve this process is to leverage statistical or computational approaches [21]. When it comes to textual data analysis, discourse around a convergence of close reading (e.g. Grounded Theory) and distant reading (e.g. topic modelling) methods are already underway [5]. As video information retrieval and multimodal machine learning technologies advance, methods and tools for mixed-initiative video data analysis (i.e., distant *watching* of videos) will be able to accelerate this process.

To ensure ethical and legal compliance and build trust with domain experts who create online public videos, it is crucial to acknowledge their contributions and give credit where it is due. While using these videos offers advantages such as coverage and diversity of design patterns, it is important to disclose their video sources. This is not only ethical, but it also helps validate credibility and reliability of guidelines, and prevents domain experts from feeling undervalued. Without such disclosure, the guidelines could be seen as questionable.

5.3 Is our method pipeline dependent on expert consultation?

Here we discuss the feasibility of using the pipeline with minimal designer consultation. Although we involved 16 designers, where three of them are domain experts, in the evaluation phase (section 4.1), the analysis of the results from this study indicates that our method is capable of generating a robust set of guidelines *without* involving a large group of domain experts as elaborated below.

The videos served their role for capturing comprehensive design patterns successfully before incorporating designer input. A part of the evaluation protocol was to ask them if they can ideate aspects of AR instruction experience that our guildelines failed to capture, but the designer opinions were in agreement that the guidelines, generated without expert consultation, were addressing a wide range of interactions.

The guidelines received reasonably high ratings (see Fig. 5), indicating that the guidelines were appropriately rectified and validated through the literature-based rationalization phase (Phase 2), without the expert inputs. However, the designer inputs on DG 10 and 11 shows that the guideline can benefit from a small number of designer' input. Rather than acquiring many domain experts, which may be time-consuming and expensive, a few experienced designers can make a big difference. For instance, in the interview, the first participant (P1) could identified the feasibility problems in DG 10 and 11 as detailed in section 4.1.5.

When applying our methods in practice, it is advisable that the practitioners evaluate the cost-value trade-offs of involving domain experts in the guideline generation process. The evaluation results of our study indicates the minimum requirements for validating guidelines by involving a small group of expert in the last stage of the pipeline.

5.4 Novelty ratings of the guidelines

Guidelines implicitly emerge when best design practices gain traction within the field. We collected some of these practices and contribute a novel set of guidelines. Although, some of our guidelines, such as DG2, are relatively novel, our guideline is geared toward prioritizing comprehensiveness and validity checks rather than generating a set of individually novel guidelines. Because we examined and selected existing design patterns based on their relevance to theoretical or empirical literature, the majority of our DGs were grounded in previous work.

In Phase 3, the designers had significant challenges in properly evaluating the novelty of guidelines. The average novelty scores for each guideline were relatively low (M: 2.59, SD: 0.31, see Figure 5). Indeed, during interview sessions, comments such as "I've seen this before" or "it's common practice" were common amongst the designers. However, when asked *where* they had seen similar guidelines before, they had difficulty pinpointing the exact source: *"I couldn't point to somewhere specific, but I think it's commonly accepted"* (P6), *"It's pretty standard to do this type of thing"* (P13).

Although we do not claim that our guidelines are individually novel, we claim the novelty of the guidelines as a whole set. This is a non-obvious contribution because not all existing guidelines for generic AR design are applicable to assembly instructions; we rationalized which existing DGs are transferable and worth incorporating and provided thorough rationale for each guideline. Also, we make this data accessible to designers through the AR Gallery, a video-enriched presentation of our guidelines.

5.5 Characterizing information scarcity

The characterization of information scarcity concerning guideline development is *subjective* depending on where one is situated in the given information landscape; is it accessible and what are the barriers in doing so? For instance, domain experts may be a highly accessible for large tech companies who create relevant platforms. However, researchers in less-connected organizations may have difficulty obtaining a similar level of recruitment. Our study can be seen as a case study that addresses one particular instance of information scarcity. For us, AR designers experienced in assembly instructions were prized but limited resources. Thus, we began our inquiry of the design space by analyzing a publicly accessible dataset (i.e. online videos). When online public video are available for a given tech domain, leveraging them is a highly beneficial approach because they are accessible to most. Yet, we anticipate that the significant time cost of analyzing video data is a potential barrier in positioning videos as a viable data source for everyone. We suggested potential solutions in sections 5.1 and 5.2.

5.6 Reflecting on our design guidelines

5.6.1 Gaps between "ideal" and "real" design practices. In addition to the high 'agreement' and 'utility' ratings, our designer participants explicitly expressed their support of our guidelines: *"All guidelines are very reasonable"* (P1), *"You picked up a nice nuance that's particular to AR [...] It doesn't feel like general knowledge"* (P9), *"I wish some of these guidelines would be official and become a common practice"* (P11). Although the majority of our guidelines were positively received, they are, in a sense, "idealistic" recommendations. When it comes to design, there are various constraints and challenges that can make it difficult to implement guidelines into practice. One common constraint is feasibility. In our study, the average 'feasibility' ratings were lower than the ratings in the other dimensions, and tended to vary between guidelines. During interview sessions, designers reported that there are multiple layers of feasibility that determine the practical applicability of a given guideline. To specify, designers often prioritize the cost of design and development over creating the ideal user experience to find a reasonable trade-off between the two. An implementation of a particular design may require an infeasible level of technicality or designers may lack the proper support to bring the design suggestions to reality (e.g. for DG 7, creating high-quality 3D animations is costly and time-consuming).

5.6.2 Omission of safety issues. Physical assembly is a critical task with potential safety hazards, and a reliable AR system must provide measures to address these risks and protect workers. This can include visual and auditory cues, step-by-step instructions, and real-time feedback on the task being performed. AR systems may also incorporate sensors and monitoring tools to verify that workers are following proper procedures and wearing the appropriate protective gear. However, our guidelines do not sufficiently emphasize the importance of safety, as noted by the absence of a safety provision in the guidelines with only a limited exception in DG2. To ensure the robustness and reliability of the guidelines, it is recommended that practitioners implement additional verification measures. This can include incorporating design heuristics or relevant existing guidelines as part of the literature review during Phase 2.

5.6.3 Presenting design guidelines. During the interview sessions, the video examples were positively received by the designers. They found that visual examples were useful for understanding the guidelines and for anchoring their critiques. Indeed, rich discussions about specific guidelines were often centred around the visual examples. Initially, we contemplated including a "bad" example where a situation could have been improved by following the guideline. However, after discussion among the team members, we decided that including an example of the *absence* of a guideline would be too confusing and the "negative design space" too vast.

Our criteria for a “good” example was high relevance to the target guideline and clarity in communicating the task context. Each visual example was extracted from our video dataset. Although this method was convenient, some designers had difficulty comparing the examples due to their range in video quality and clarity. An alternative avenue for the presentation of guidelines is to standardize the examples so that they are visually consistent. Companies such as Apple already present their guidelines in such a way to ensure system-level presentation and brand uniformity. Video examples are a clear step above text-only descriptions for communicating guidelines for visual, dynamic experiences. However, we contend that some guidelines are only fully understood when experienced in AR. Future work in the area of AR guideline development could involve presenting a variety of guidelines for different contexts through a unified AR system, while considering the range of interaction primitives.

6 CONCLUSION AND FUTURE WORK

Taking AR-based assembly instruction as a case of technical domains characterized by limited access to domain experts, we established and executed a novel method pipeline to develop design guidelines for the information-scarce application areas. First, we analyzed online videos, an easily accessible and large dataset, to identify the central characteristics and design patterns of AR assembly instruction. Then, by surveying the literature, we rationalized the patterns, transformed them into prescriptive guidelines, refined our knowledge of the design space, and iteratively developed a set of domain-specific design guidelines. We validated our method by running designer interviews. The result supported that the resultant guidelines are comprehensive and practical.

Through presenting and evaluating our method pipeline, our study provides insightful guidance for interaction designers operating in an information-scarce domain. The use of online public videos as a data source offers a comprehensive, practical, and user-centered approach to analyzing design patterns, complementing academic literature. However, biases in the videos emphasize the importance of incorporating triangulation with other data sources, such as literature or design heuristics, to ensure robustness. Our work contributes to a broader movement aimed at overcoming challenges in the adoption and design of innovative systems.

One promising area of future work is to pursue ethnographic inquiry into how designers comprehend, rationalize, and adopt design guidelines into their practice. During our interview sessions with designers, we noticed that each participant anchored their evaluation upon a wide variety of individual factors such as their past experiences, design philosophy, and personal biases. For instance, some participants prioritized "accessibility," while others focused on "technical feasibility." Understanding how real designers evaluate guidelines will shed light on how to develop guidelines that dynamically address the specific concerns of each designer. Another avenue for future work is to characterize and counter new challenges that can emerge as the target technology becomes widely available. If so, abundance of information, not scarcity, can pose different kinds of challenges, such as data scalability or bias issues. Also, as standards and conventions emerge in a domain, promoting new or different design practices against the inertial practices can get more challenging.

ACKNOWLEDGEMENTS

This work was supported by a grant from the Korea Evaluation Institute of Industrial Technology (KEIT), funded by the Korean government (MOTIE, No. 20009940). Additional support came from the NSERC Discovery Grant and CREATE programs, as well as a generous gift from Adobe Research.

We acknowledge that this study was made possible by analyzing online public videos created by domain experts, whose efforts and expertise should be credited. Please refer to our supplementary materials for the full list of videos used in our study and their respective publishers.

Lastly, our heartfelt thanks go to the members of D-lab and MUX lab for their invaluable discussions and feedback.

REFERENCES

[1] Saleema Amershi, Dan Weld, Mihaela Vorvoreanu, Adam Fourney, Besmira Nushi, Penny Collisson, Jina Suh, Shamsi Iqbal, Paul N. Bennett, Kori Inkpen, Jaime Teevan, Ruth Kikin-Gil, and Eric Horvitz. 2019. Guidelines for Human-AI Interaction. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems* (Glasgow, Scotland Uk) *(CHI '19)*. ACM, New York, NY, USA, Article 3, 13 pages. https://doi.org/10.1145/3290605.3300233

[2] Lisa Anthony, YooJin Kim, and Leah Findlater. 2013. Analyzing User-generated Youtube Videos to Understand Touchscreen Use by People with Motor Impairments. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Paris, France) *(CHI '13)*. ACM, New York, NY, USA, 1223–1232. https://doi.org/10.1145/2470654.2466158

[3] Apple. 2019. *Apple Human Interface Guidelines.* Apple Inc. https://developer.apple.com/design/human-interface-guidelines/

[4] Reynold Bailey, Ann McNamara, Aaron Costello, Srinivas Sridharan, and Cindy Grimm. 2012. Impact of Subtle Gaze Direction on Short-term Spatial Information Recall. In *Proceedings of the Symposium on Eye Tracking Research and Applications* (Santa Barbara, California) *(ETRA '12)*. ACM, New York, NY, USA, 67–74. https://doi.org/10.1145/2168556.2168567

[5] Eric P. S. Baumer, David Mimno, Shion Guha, Emily Quan, and Geri K. Gay. 2017. Comparing grounded theory and topic modeling: Extreme divergence or unlikely convergence? *Journal of the Association for Information Science and Technology* 68, 6 (2017), 1397–1410. https://doi.org/10.1002/asi.23786

[6] Mark Billinghurst, Adrian Clark, and Gun Lee. 2015. A Survey of Augmented Reality. *Foundations and Trends in Human-Computer Interaction* 8, 2-3 (2015), 73–272. https://doi.org/10.1561/1100000049

[7] Mark Blythe and Paul Cairns. 2009. Critical Methods and User Generated Content: The iPhone on YouTube. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Boston, MA, USA) *(CHI '09)*. ACM, New York, NY, USA, 1467–1476. https://doi.org/10.1145/1518701.1518923

[8] T.P. Caudell and D.W. Mizell. 1992. Augmented reality: an application of heads-up display technology to manual manufacturing processes. In *Proceedings of the Twenty-Fifth Hawaii International Conference on System Sciences*, Vol. ii. IEEE Computer Society, 659–669 vol.2. https://doi.org/10.1109/HICSS.1992.183317

[9] Nikolas Chaconas and Tobias Hollerer. 2018. An Evaluation of Bimanual Gestures on the Microsoft HoloLens. In *2018 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*. 1–8. https://doi.org/10.1109/VR.2018.8446320

[10] Dempsey Chang and Keith V. Nesbitt. 2006. Developing Gestalt-based Design Guidelines for Multi-sensory Displays. In *Proceedings of the 2005 NICTA-HCSNet Multimodal User Interaction Workshop - Volume 57* (Sydney, Australia) *(MMUI '05)*. Australian Computer Society, Inc., Darlinghurst, Australia, Australia, 9–16. https://doi.org/10.5555/1151804.1151807

[11] Viviana Chimienti, Salvatore Iliano, Michele Dassisti, Gino Dini, and Franco Failli. 2010. Guidelines for implementing augmented reality procedures in assisting assembly operations. In *International Precision Assembly Seminar*. Springer, 174–179. https://doi.org/10.1007/978-3-642-11598-1_20

[12] Marie-Paule Daniel and Barbara Tversky. 2012. How to put things together. *Cognitive Processing* 13, 4 (Nov. 2012), 303–319. https://doi.org/10.1007/s10339-012-0521-5

[13] Abhiraj Deshpande and Inki Kim. 2018. The effects of augmented reality on improving spatial problem solving for object assembly. *Advanced Engineering Informatics* 38 (2018), 760–775. https://doi.org/10.1016/j.aei.2018.10.004

[14] Andreas Dünser, Raphaël Grasset, Hartmut Seichter, and Mark Billinghurst. 2007. Applying HCI principles to AR systems design. (2007). http://hdl.handle.net/10092/2340

[15] Stephen Elliott, Alexander Kurz, Peter Beddow, and Jennifer Frey. 2009. Cognitive Load Theory: Instruction-based Research with Applications for Designing Tests.

[16] P. Fite-Georgel. 2011. Is there a reality in Industrial Augmented Reality?. In *2011 10th IEEE International Symposium on Mixed and Augmented Reality*. 201–210. https://doi.org/10.1109/ISMAR.2011.6092387

[17] Nicholas R. J. Frick, Felix Brünker, Björn Ross, and Stefan Stieglitz. 2019. Towards Successful Collaboration: Design Guidelines for AI-based Services enriching Information Systems in Organisations. *CoRR* abs/1912.01077 (2019). arXiv:1912.01077 http://arxiv.org/abs/1912.01077

[18] Joseph L Gabbard and D Hix. 2001. Researching usability design and evaluation guidelines for augmented reality (AR) systems. *Laboratory for Scientific Visual Analysis, Virginia Tech, USA* (2001). Retrieved May 8, 2023 from https://www.rkriz.net/sv/classes/ESM4714/Student_Proj/class00/gabbard/index.html

[19] Joseph L. Gabbard, J. Edward Swan, and Deborah Hix. 2006. The Effects of Text Drawing Styles, Background Textures, and Natural Lighting on Text Legibility in Outdoor Augmented Reality. *Presence: Teleoperators and Virtual Environments* 15, 1 (2006), 16–32. https://doi.org/10.1162/pres.2006.15.1.16 arXiv:https://doi.org/10.1162/pres.2006.15.1.16

[20] Nirit Gavish, Teresa Gutierrez, Sabine Webel, Jorge Rodriguez, and Franco Tecchia. 2011. Design guidelines for the development of virtual reality and augmented reality training systems for maintenance and assembly tasks. In *BIO web of conferences*, Vol. 1. EDP Sciences, 00029. https://doi.org/10.1051/bioconf/20110100029

[21] Fabio Giglietto, Luca Rossi, and Davide Bennato. 2012. The open laboratory: Limits and possibilities of using Facebook, Twitter, and YouTube as a research data source. *Journal of technology in human services* 30, 3-4 (2012), 145–159. https://doi.org/10.1080/15228835.2012.743797

[22] Google. 2019. *Augmented Reality Design Guidelines*. https://designguidelines.withgoogle.com/ar-design/augmented-reality-design-guidelines/

[23] Elizabeth R Grant and Michael J Spivey. 2003. Eye movements and problem solving: Guiding attention guides thought. *Psychological Science* 14, 5 (2003), 462–466. https://doi.org/10.1111/1467-9280.02454

[24] Martin Groen and Jan Noyes. 2010. Solving problems: How can guidance concerning task-relevancy be provided? *Computers in Human Behavior* 26, 6 (2010), 1318 – 1326. https://doi.org/10.1016/j.chb.2010.04.004 Online Interactivity: Role of Technology in Behavior Change.

[25] Gregory M. Hallihan and L.H. Shu. 2013. Considering Confirmation Bias in Design and Design Research. *Journal of Integrated Design and Process Science* 17, 4 (2013), 19–35. https://doi.org/10.3233/jid-2013-0019 Publisher: IOS Press.

[26] Mary Hegarty. 1992. Mental animation: Inferring motion from static displays of mechanical systems. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 18, 5 (1992), 1084. https://doi.org/10.1037/0278-7393.18.5.1084

[27] Julie Heiser, Doantam Phan, Maneesh Agrawala, Barbara Tversky, and Pat Hanrahan. 2004. Identification and Validation of Cognitive Design Principles for Automated Generation of Assembly Instructions. In *Proceedings of the Working Conference on Advanced Visual Interfaces* (Gallipoli, Italy) *(AVI '04)*. ACM, New York, NY, USA, 311–319. https://doi.org/10.1145/989863.989917

[28] S. Henderson and S. Feiner. 2011. Exploring the Benefits of Augmented Reality Documentation for Maintenance and Repair. *IEEE Transactions on Visualization and Computer Graphics* 17, 10, 1355–1368. https://doi.org/10.1109/TVCG.2010.245

[29] CN2 Inc. 2015. *Printer Repaired with Augmented Reality*. Youtube. https://youtu.be/btW-MLz74ks

[30] Eric Jamet, Monica Gavota, and Christophe Quaireau. 2008. Attention guiding in multimedia learning. *Learning and Instruction* 18, 2 (2008), 135 – 145. https://doi.org/10.1016/j.learninstruc.2007.01.011

[31] D. Kalkofen, M. Tatzgern, and D. Schmalstieg. 2009. Explosion Diagrams in Augmented Reality. In *2009 IEEE Virtual Reality Conference*. 71–78. https://doi.org/10.1109/VR.2009.4811001

[32] kingsee青橙视界. 2016. *Intelligent Aircraft Maintenance-Kingsee AR Smartglasses Performance Support System (PSS) 航空智能维修*. Youtube. https://youtu.be/zP_5Z1A9Pcw
[33] Aida Komkaite, Liga Lavrinovica, Maria Vraka, and Mikael B. Skov. 2019. Underneath the Skin: An Analysis of YouTube Videos to Understand Insertable Device Interaction. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems* (Glasgow, Scotland Uk) *(CHI '19)*. ACM, New York, NY, USA, Article 214, 12 pages. https://doi.org/10.1145/3290605.3300444
[34] S. A. Konz and G. L. Dickey. 1969. Manufacturing Assembly Instructions; A Summary. *Ergonomics* 12, 3 (1969), 370–382. https://doi.org/10.1080/00140136908931061 arXiv:https://doi.org/10.1080/00140136908931061
[35] Sri Kurniawan and Panayiotis Zaphiris. 2005. Research-derived Web Design Guidelines for Older People. In *Proceedings of the 7th International ACM SIGACCESS Conference on Computers and Accessibility* (Baltimore, MD, USA) *(Assets '05)*. ACM, New York, NY, USA, 129–135. https://doi.org/10.1145/1090785.1090810
[36] F. Lamberti, F. Manuri, A. Sanna, G. Paravati, P. Pezzolla, and P. Montuschi. 2014. Challenges, Opportunities, and Future Trends of Emerging Techniques for Augmented Reality-Based Maintenance. *IEEE Transactions on Emerging Topics in Computing* 2, 4 (Dec 2014), 411–421. https://doi.org/10.1109/TETC.2014.2368833
[37] Joseph J LaViola Jr, Ernst Kruijff, Ryan P McMahan, Doug Bowman, and Ivan P Poupyrev. 2017. *3D user interfaces: theory and practice*. Addison-Wesley Professional.
[38] Karen K. Lee. 2012. Developing and implementing the Active Design Guidelines in New York City. *Health & Place* 18, 1 (2012), 5 – 7. https://doi.org/10.1016/j.healthplace.2011.09.009 Active Living Research.
[39] LlamaZoo. 2019. *Clarity 3D*. https://www.llamazoo.com/clarity-3d/
[40] Céline Mariage, Jean Vanderdonckt, and Costin Pribeanu. 2005. State of the art of web usability guidelines. *The handbook of human factors in web design* (2005), 688–700. http://hdl.handle.net/2078/17946
[41] Bernardo Marques, Samuel Silva, João Alves, António Rocha, Paulo Dias, and Beatriz Sousa Santos. 2022. Remote collaboration in maintenance contexts using augmented reality: Insights from a participatory process. *International Journal on Interactive Design and Manufacturing (IJIDeM)* (2022), 1–20. https://doi.org/10.1007/s12008-021-00798-6
[42] O. Marques. 2016. Visual Information Retrieval: The State of the Art. *IT Professional* 18, 4 (July 2016), 7–9. https://doi.org/10.1109/MITP.2016.70
[43] Davis L. Masten and Tim M.P. Plowman. 2003. Digital ethnography: The next wave in understanding the consumer experience. *Design Management Journal (Former Series)* 14, 2 (2003), 75–81. https://doi.org/10.1111/j.1948-7169.2003.tb00044.x arXiv:https://onlinelibrary.wiley.com/doi/pdf/10.1111/j.1948-7169.2003.tb00044.x
[44] Richard E Mayer. 2002. Multimedia learning. In *Psychology of learning and motivation*. Vol. 41. Elsevier, 85–139.
[45] Robert K. Merton. 1968. The Matthew Effect in Science. *Science* 159, 3810 (Jan. 1968), 56–63. https://doi.org/10.1126/science.159.3810.56 Publisher: American Association for the Advancement of Science.
[46] Microsoft. 2019. *Dynamics 365*. https://dynamics.microsoft.com/en-ca/mixed-reality/guides/
[47] Microsoft. 2019. *Mixed Reality Design Guidelines*. https://docs.microsoft.com/en-us/windows/mixed-reality/design
[48] Paul Milgram, Haruo Takemura, Akira Utsumi, and Fumio Kishino. 1995. Augmented reality: A class of displays on the reality-virtuality continuum. In *Telemanipulator and telepresence technologies*, Vol. 2351. International Society for Optics and Photonics, 282–292. https://doi.org/10.1117/12.197321
[49] Hans-Christoph Nothdurft. 2002. Attention shifts to salient targets. *Vision Research* 42, 10 (2002), 1287 – 1306. https://doi.org/10.1016/S0042-6989(02)00016-0
[50] Kenneth R Ohnemus. 1997. Web style guides: who, what, where. In *Proceedings of the 15th annual international conference on Computer documentation*. ACM, 189–197. https://doi.org/10.1145/263367.263392
[51] Susan Palmiter, Jay Elkerton, and Patricia Baggett. 1991. Animated demonstrations vs written instructions for learning procedural tasks: a preliminary investigation. *International Journal of Man-Machine Studies* 34, 5 (1991), 687 – 701. https://doi.org/10.1016/0020-7373(91)90019-4
[52] Cindy M. Robertson, Blair MacIntyre, and Bruce N. Walker. 2008. An Evaluation of Graphical Context when the Graphics Are Outside of the Task Area. In *Proceedings of the 7th IEEE/ACM International Symposium on Mixed and Augmented Reality (ISMAR '08)*. IEEE Computer Society, Washington, DC, USA, 73–76. https://doi.org/10.1109/ISMAR.2008.4637328
[53] Cledja Rolim, Dieter Schmalstieg, Denis Kalkofen, and Veronica Teichrieb. 2015. Design Guidelines for Generating Augmented Reality Instructions. In *2015 IEEE International Symposium on Mixed and Augmented Reality*. IEEE, 120–123. https://repositorio.ufpe.br/handle/123456789/25074
[54] Marc Ericson C. Santos, Arno in Wolde Lübke, Takafumi Taketomi, Goshiro Yamamoto, Ma. Mercedes T. Rodrigo, Christian Sandor, and Hirokazu Kato. 2016. Augmented reality as multimedia: the case for situated vocabulary learning. *Research and Practice in Technology Enhanced Learning* 11, 1 (20 Jan 2016), 4. https://doi.org/10.1186/s41039-016-0028-2
[55] Dominique Scapin, Corinne Leulier, Jean Vanderdonckt, Céline Mariage, Christian Bastien, Christelle Farenc, Philippe Palanque, and Rémi Bastide. 2000. A Framework for Organizing Web Usability Guidelines. In *6th Conference on Human Factors and the Web (HF Web 2000) (Proceedings. of 6th Conf. on Human Factors and the Web HFWeb'2000)*, Ph. Kortum and E. Kudzinger (Eds.). Austin, United States. https://hal.science/hal-01319628

[56] Hasti Seifi, Farimah Fazlollahi, Michael Oppermann, John Andrew Sastrillo, Jessica Ip, Ashutosh Agrawal, Gunhyuk Park, Katherine J. Kuchenbecker, and Karon E. MacLean. 2019. Haptipedia: Accelerating Haptic Device Discovery to Support Interaction & Engineering Design. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems* (Glasgow, Scotland Uk) *(CHI '19)*. Association for Computing Machinery, New York, NY, USA, Article 558, 12 pages. https://doi.org/10.1145/3290605.3300788

[57] Roger N Shepard and Jacqueline Metzler. 1971. Mental rotation of three-dimensional objects. *Science* 171, 3972 (1971), 701–703. https://doi.org/10.1126/science.171.3972.701

[58] Maarten Someren, Yvonne Barnard, and Jacobijn Sandberg. 1994. *The Think Aloud Method - A Practical Guide to Modelling CognitiveProcesses.* https://hdl.handle.net/11245/1.103289

[59] Anselm Strauss and Juliet Corbin. 1990. *Basics of qualitative research.* Sage publications. https://doi.org/10.4135/9781452230153

[60] Timo Strohmann, Laura Höper, and Susanne Robra-Bissantz. 2019. Augmented reality: an application of heads-up display technology to manual manufacturing processes. In *Proceedings of the Fifty-Second Hawaii International Conference on System Sciences.* https://doi.org/10.1109/HICSS.1992.183317

[61] J Edward Swan and Joseph L Gabbard. 2005. Survey of user-based experimentation in augmented reality. In *Proceedings of 1st International Conference on Virtual Reality*, Vol. 22. 1–9.

[62] Arthur Tang, Charles Owen, Frank Biocca, and Weimin Mou. 2002. Experimental evaluation of augmented reality in object assembly task. In *Proceedings. International Symposium on Mixed and Augmented Reality.* 265–266. https://doi.org/10.1109/ISMAR.2002.1115105

[63] Arthur Tang, Charles Owen, Frank Biocca, and Weimin Mou. 2003. Comparative Effectiveness of Augmented Reality in Object Assembly. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Ft. Lauderdale, Florida, USA) *(CHI '03)*. ACM, New York, NY, USA, 73–80. https://doi.org/10.1145/642611.642626

[64] Laura E Thomas and Alejandro Lleras. 2007. Moving eyes and moving thought: On the spatial compatibility between eye movements and cognition. *Psychonomic bulletin & review* 14, 4 (2007), 663–668. https://doi.org/10.3758/BF03196818

[65] Trimble. 2019. *Trimble Mixed Reality.* https://mixedreality.trimble.com

[66] Dirk Walther, Ueli Rutishauser, Christof Koch, and Pietro Perona. 2005. Selective visual attention enables learning and recognition of multiple objects in cluttered scenes. *Computer Vision and Image Understanding* 100, 1 (2005), 41 – 63. https://doi.org/10.1016/j.cviu.2004.09.004 Special Issue on Attention and Performance in Computer Vision.

[67] Ranxiao Frances Wang and Elizabeth S Spelke. 2002. Human spatial representation: insights from animals. *Trends in Cognitive Sciences* 6, 9 (2002), 376 – 382. https://doi.org/10.1016/S1364-6613(02)01961-7

[68] Michelle Q. Wang Baldonado, Allison Woodruff, and Allan Kuchinsky. 2000. Guidelines for Using Multiple Views in Information Visualization. In *Proceedings of the Working Conference on Advanced Visual Interfaces* (Palermo, Italy) *(AVI '00)*. ACM, New York, NY, USA, 110–119. https://doi.org/10.1145/345513.345271

[69] James R. Williams. 1998. Guidelines for the Use of Multimedia in Instruction. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* 42, 20, 1447–1451. https://doi.org/10.1177/154193129804202019

[70] Yu Xiao and Maria Watson. 2019. Guidance on conducting a systematic literature review. *Journal of planning education and research* 39, 1 (2019), 93–112. https://doi.org/10.1177/0739456x17723971

Received 2023-02-15; accepted 2023-05-30